



ACTOR-NETWORK INSPIRED DESIGN RESEARCH: METHODOLOGY AND REFLECTIONS

Ben Kraal¹

¹School of Design, Faculty of Built Environment and Engineering, Queensland University of Technology
b.kraal@qut.edu.au

ABSTRACT:

This paper describes how Actor-network theory, an approach to sociological analysis which comes from the field of Science and Technology Studies, can be used in the analysis of field work for design research. A case study of the investigation of the use of speech recognition dictation software is used to present an actor-network approach to the analysis of field work for design.

Using an actor-network approach is useful in analyzing qualitative data, however it does not replace, or attempt to replace, the creativity involved in designing. Instead it is used to guide design, beginning a process of ensuring that a final product is useful and useable. The Actor-Network approach used here showed that users of speech recognition dictation software faced unique challenges in integrating the software into their work practices. An understanding of these challenges points the way to improved speech recognition software in the future. Finally, an actor-network approach is shown to be of potential further interest to design research.

I. INTRODUCTION

The research reported in this paper arose after the Magistrates Court approached our research team because of our expertise in speech recognition systems. The Court wanted to know if it would be possible to use automatic speech recognition¹ (ASR) to streamline parts of their work practice. In addition to investigating the Court's work processes, which are not reported in this paper, we performed field studies of two different groups of existing users of ASR in order to understand users made seemingly difficult to use software useful enough to work with everyday.

Analysing the fieldwork from the existing ASR users revealed that using ASR was a complex process that involved the interrelation of many heterogeneous elements. Actor-network theory (ANT) is one methodology that is geared to the analysis of complex socio-technical systems. ANT is particularly concerned with how social order is established and maintained. This proved particularly useful in our investigations of users of ASR.

I.1 AUTOMATIC SPEECH RECOGNITION

Automatic speech recognition (ASR) software, particularly dictation software, is marketed as an easy-to-use system that allows a person to control a computer by voice. People who use ASR software in the workplace often do so as the result of an occupational overuse injury that makes typing painful or impossible.

The usability and utility of ASR applications is assumed to be closely related to the software's accuracy at recognising words (Huang *et al*, 1999; Halverson *et al*, 1999). Most current commercial ASR software applications are advertised as having a high accuracy, perhaps 98% or greater. This accuracy rate can be achieved; however it takes a great deal of commitment from the user to sustain accurate recognition.

Talking to an ASR application is often compared with taking dictation with a secretary. When speaking with a secretary, dictation can proceed in fits and starts and the speaker can revise at any time. If the speaker stumbles over a word, or starts a sentence badly, they are able to revise as they speak, using commonly accepted conversational methods. The speaker and the secretary cooperate to make dictation work. Dictating to a computer is quite a different task. Dictation can only begin once the software is trained to the user's voice, a process which can be time-consuming. If the user misspeaks, stumbles or

¹ In the ASR field, speech recognition is what people do. Automatic speech recognition is what computers do.

changes their mind in the middle of a sentence, the computer does not know but treats all sounds as intended and attempts to recognise each sound as a speak act. All punctuation, except spaces, must be spoken aloud. The resulting text does not look as if it was written by a person. Even if the recognition was perfect, the dictation would contain every stumble and misspoken word. At the current state of the art the software application cannot cooperate with the user in the same way as the secretary does with the speaker.

Every person who uses an ASR application productively overcomes the deficiencies in the application by changing how they work, changing their work environment and coming to understand, at least a little, how the ASR application works. The user must make up for the lack of cooperation from the software. The idea that using a software application in a particular environment involves some degree of modification to the environment, the existing work practices and even to users themselves is not a new one, indeed it is almost a cliché to suggest that use is contextual. It is equally trite to say that ASR software is hard to use. This paper is concerned with the ways in which users of ASR applications overcome the software's deficiencies and the analysis of observations of those users using a method derived from Actor-Network Theory.

1.2 MOTIVATION FOR USING ACTOR-NETWORK THEORY

After the interviews had been completed, transcripts were produced and then the data was “worked” using a Grounded Theory (Glasser, B. and Strauss, A., 1967) approach, reading and re-reading the transcripts, comparing and contrasting them and allowing the broad themes to emerge inductively. Grounded theory aids in understanding data but does not need to stand alone. In this analysis, two different approaches were used to “work the data” in order to obtain deeper understanding and insight into the problems of using ASR in an organisation. The Locales Framework (Fitzpatrick, G., 2003), particularly the interaction trajectory aspect (Graham, C. et al., 2005) which address the “co-evolution of action, locale and social world as the trajectory unfolds” (Fitzpatrick, G., 2003 pp. 120) was valuable in establishing the broad themes of the analysis. However the mechanics of how co-evolution of action, locale and social world occurs is not specifically addressed in the Locales Framework. Actor-Network Theory addresses the co-evolution of interactants in a locale, or actors in a network, in a way that was useful in the analysis of this research.

Actor-Network Theory (ANT) is an approach to doing sociological analysis. ANT comes from the area of sociology called Science and Technology Studies which has tended to be interested in how science is done

and how the social influences science (Latour, 1987). ANT has been used in computer science and information systems as a way to analyse how computer systems are used in the workplace (Tatnall and Gilding, 1999; Abdelnour Nocera and Hall, 2004). Using ANT for systems analysis of computer systems is well within its capabilities.

ANT is useful for explaining how technologies and people interact over time. ANT has been used for analysing the interplay between many different social systems and technologies from scallops and fishermen (Callon, M., 1986b), electric vehicles (Callon, M., 1986a), the Portuguese expeditions to India for the spice trade (Law, J., 1986, Law, J., 1987), seatbelts (Latour, B., 1992) and alternative public transport systems (Latour, B., 1996). ANT has more recently come to the attention of computer systems researchers (Hanseth, O. and Monteiro, E., 1998, Tatnall, A. and Gilding, A., 1999). What all of these disparate analyses have in common is the language of ANT, which is particularly geared towards describing “the small, concrete technical and non-technical mechanisms which go into the building and use” (Hanseth, O. and Monteiro, E., 1998) of socio-technical systems.

The ANT research frame (after Callon, M., 1986b) directs the researcher’s gaze towards particular aspects of a socio-technical system that are useful during analysis. It comprises four inter-related overlapping steps that describe how stable actor-networks come to be established. Stable actor-networks are of research interest because they represent the status quo. The research frame can be used to ask how stable networks became stabilised. The failure of an actor-network to become stabilised can equally be examined by the research frame. In the (1) interestment step, actors are made interested in joining an actor-network. The way in which individual actors are interested is unique to the particular actor-network. In the (2) enrollment step, actors agree to play a role in the network, they are translated (Law, J., 2003) into the network and are inscribed (Akrich, M., 1992) with a program of action (Abdelnour Nocera, J. L. and Hall, P., 2004). Put another way, actors who join a network are given a script to follow. In any network, one or more actors attempt to establish themselves as a (3) point of passage. The point of passage is the actor who assigns roles to, or acts as a spokesperson for, the other actors in the network. Conflicts arise in an actor-network when more than one actor attempts to establish themselves as a point of passage. Finally, in the (4) trial of strength it is seen whether the actors adopt the roles assigned to them. These steps are not necessarily linear and all four steps are constantly in play. The most important of these concepts used in this paper are inscription and translation which are intimately tied to design (Akrich, 1992).

The analysis that this ANT-derived process allows is to show how users translate the software and how they strengthen their inscription in order to make the software useful. The ease with which the new

actor-network is established and maintained is related to the flexibility of the existing actor-network in which the users are working and the flexibility of the software, which is itself part of the actor-network. The analysis of the interviews shows that using ASR dictation systems in the workplace is difficult, for issues of integration rather than accuracy.

2. FIELD STUDIES

Finding experienced users of ASR systems was difficult. Six of the interviewees (see Table 1) were found through fortuitous contact with a professional speech recognition trainer. All six had previously suffered from a “occupational overuse injury” and worked at the time of the interviews in different agencies within the Australian Federal Public Service. Two ASR application users worked in the Parliamentary Reporting Service (which is often called the Hansard department after the name of the main document it produces) of Australian Federal Parliament House. That the Parliamentary Reporting Service used desktop speech recognition was discovered through an article in the *Canberra Times*, the newspaper of Australia’s federal capital.

The interviews were arranged at a time and location suitable to the interviewees. Depending on their status within their organization and security requirements, some interviews were conducted at interviewee’s desks. Where access to an interviewee’s work space was not possible, interviews took place in locations near the interviewee’s work place. Where possible, the interviewees demonstrated some of their work practices during the interview.

Table 1: List of interview participants and their affiliation

Interviewee	Employer	Injury
Margaret	Public Service Department	Occupational Overuse
Jane	Public Service Department	Occupational Overuse
Yvonne	Public Service Department	Occupational Overuse
	Public Service Department	Occupational Overuse
	Public Service Department	Occupational Overuse
Robyn	Hansard	None
Robyn	Hansard	None

Names of interviewees have been changed. Because the Parliamentary Reporting Service (Hansard) is so small, relevant quotes from the two interviewees from that location are attributed to “Robyn” to further preserve their anonymity.

3. METHODOLOGY: ANALYSIS OF INTERVIEWS

The ANT-derived methodology used in this research relates specifically to the analysis of data acquired during field work, mainly interviews and observations of workplaces. It is in the nature of ANT that it is easier to describe through a demonstration of its use (for example, Law, 2003). This section is a demonstration of how an ANT-derived methodology can be used to reveal the complexity involved in using a product, in this case commercial off-the-shelf ASR dictation software.

There are two major themes that arose in the analysis of the interviews and observations of the two sets of ASR users. The first theme is translation, following Law (2003). The second theme is enrollment and the related idea of inscription (Akrich, 1992). Using examples from the interviews, these themes are elaborated upon in the following subsections.

3.1 TRANSLATION

Actor-Network theory tells stories about how things, objects, actors, come to be how they are. In ANT, actors come to be how they are through a process of interaction with other actors. Importantly, actors can be human or non-human, material or immaterial. Interaction changes actors. It *translates* actors.

Margaret, speaking about the new frame of mind that using ASR in the workplace required said:

“When dentists started using rubber gloves at the beginning of the HIV crisis—you know I remember a time when dentists didn’t bother using gloves—and then they went through, ‘oh, I’ve got to put the gloves on, it feels different’. But the impetus was there.

There are disciplines that you’ve got to follow when you’re using [ASR]. There are inconveniences, there are logistical issues and unless you have the impetus there, you are not going to use it.”

The idea of translation through interaction resonates with the experience of using dictation software and making it useful in the workplace. In using dictation software in the workplace, the user, the software and

their work are translated. The user becomes a person who knows about the nuances of dictation software—they become an expert—and can reason about the use of the software in sophisticated ways. The software is changed (usually) in order to become ever-so-slightly more closely aligned with the user's work. Many of the injured interviewees used "macros" to make their use of ASR faster and more closely aligned with their work practices. Macros are ways to combine many commands within software into one command. Often, the injured interviewees created their own macros to work around limitations in the ASR software.

In some cases the user's work is changed, sometimes dramatically, in order to better fit with what is possible to achieve in using the software. Yvonne changed jobs to have work that was a better fit with the capabilities and limitations of ASR:

"So, why did I start using it? I got quite a serious overuse injury and my work used to involve writing about statistics and that was very keyboard intensive work and very mouse intensive too and I actually left that work to move to a policy area because it's less keyboard intensive work and the work is mostly words and not the mouse. I moved over from [one department] to [another department] because it was better for my injury."

The user, the software and their work are no longer the same after an interaction. They are translated.

Law (2003), described Akrich's work on machines that used recycled materials to make briquettes for use in fireplaces. The machines are originally made in Sweden and are then sent to be used in Nicaragua. The way the machines are used in Sweden and Nicaragua are so different as to be almost unrecognisable. The way the machines came to be so different, to be translated so significantly, was through a series of "negotiations". In this paper, the translation of ASR software is not so significant, yet in comparing how dictation software is used at Hansard and by injured users elsewhere in the Public Service it is also possible to show that they are different due to a series of negotiations.

The ASR users at Hansard have a completely different experience with ASR to the injured users. They use ASR, not because they are injured, but because it allows them to quickly turn Parliamentarians speech into text which can then be edited and published as the day's Hansard. An audio recording of what is said in the House of Representatives and the Senate and various committees is made available to the Hansard editors. The Hansard editors who use ASR then re-speak what they hear on the recording. Some Hansard editors do not use ASR but use other technologies including CAT, or Computer Aided Transcription which is uses a chording keyboard. Learning CAT can take several years but learning ASR, for Hansard's purposes, is a much shorter process, perhaps as little as four to six months. While the injured users must

use ASR to control their whole interaction with their computer, the Hansard users use ASR (or CAT) to transform speech into text which can then be edited using more typical means. Additionally, the Hansard organisation explicitly supports ASR, providing the editors with high-end computers, ASR-suitable microphones and work spaces where talking to a computer for part of the day does not disturb other workers. “The two networks are different in every respect” (Law, 2003). This is true in Akrich’s work as it is in my ASR users study. The injured users’ and the Hansard users’ networks look very different, despite using the same software for ostensibly the same purpose. As the technology is remade and re-negotiated for each situation it is translated by the users, the organisation and even the work situation. It even plays different roles in each situation; for the Hansard users it plays the role of a tool in an array of tools, to be used when necessary, but for the injured users it plays the role of a limb, to be constantly acted through and with.

3.2 ENROLLMENT AND INSCRIPTION

In ANT terms, enrollment is the process by which actors become part of a network. As stated above, an actor in an actor-network is human or non-human, material or immaterial. Consequently, a network in an actor-network is a heterogeneous assemblage of actors. The actors enrolled in the actor-networks described in this work are users, software, hardware and any other person, thing or idea that becomes necessary for ASR to work *at work*.

The power of ANT, in this instance, is to show how necessarily entangled the different elements of software, hardware and work environment are in making ASR useful in the workplace. This interrelatedness was not always anticipated by those who were trying to help the injured interviewees. It was often assumed that installing ASR on an injured users computer would “fix” their inability to type or use a mouse:

“People assume because they have ticked off we’ve given you this drug, you’ve had physio, you’ve tried this, you’ve done that, here’s [ASR].”

Off-the-shelf ASR, as all of the interviewees used, does not work well with some software, particularly that which is highly specialised. Even after changing jobs to find work more suited to using ASR, more than one interviewee found themselves unable to work because of a change in some non-ASR software that they were required to use. Margaret, who was quite senior in her department, was “effectively disenfranchised from the [new] department document creation, sharing and storage system” because it did not work with the ASR system she used.

In contrast, the Hansard editors were always able to work with ASR effectively because their organisation was committed to compatibility between ASR and the applications that the software that they needed to use with ASR. This was made easier because none of the Hansard editors were injured so there were many fewer applications that needed to work with ASR for the editors to be effective.

The Hansard editors who used ASR were also provided with extremely powerful desktop computers, far in excess of what would typically be provided to a non-user of ASR, and very high quality microphones to use with ASR. Powerful computers are essential because ASR is extremely resource intensive, particularly in terms of processor speed and computer memory required. A high quality microphone is also essential because the quality and accuracy of the recognition provided by ASR is highly dependant on the quality of the audio provided to the software.

All of the injured interviewees said that they were not provided with adequate, appropriate, hardware in their initial experiences with ASR. Significantly, all of the interviewees who were successful users of ASR in the workplace had negotiated for powerful computers and high-end microphones. This was impressive because with the exception of Margaret, all of the injured interviewees were not particularly senior within their respective organisations and could not have ordinarily expected high-end hardware.

In order to be effective users of ASR the interviewees had to convince a person in power of the benefit of providing them with advanced hardware. In ANT terms they had to *interest* and then *enroll* that person in their actor-network and then interest and enroll the advanced hardware in their actor-network as well.

Enrollment is precarious and requires constant maintenance so that the links between actors in the network are sustained. Wendy, who was not working at the time of the interviews, was the victim of the links in her network failing:

“So, if you don’t have the expertise and as you may have gathered I have no technical genes at all, or at least no skills, and I don’t know it’s not working properly; I only know I’m nearly going mad.”

The integration between ASR and the rest of the applications on Wendy’s computer was very poor. However, she was not expert enough, and did not initially have access to experts, to diagnose the problem and fix it. Instead, she was subject to inexpert advice and felt that the failure of the ASR system was her fault. Later, when she encountered a different ASR trainer, she found that her system was improperly configured. Wendy also had political and environmental problems that prevented her from using ASR as effectively as some of the other interviewees.

Wendy said that she was, for various reasons, the target of animosity because she was an ASR user. She said that other people she worked with resented the “special treatment” that she received, including having her cubicle relocated so that she was further away from a noisy photocopier that interfered with the accuracy of her ASR.

None of the other interviewees reported similar experiences of ill-will, though many had special workplace arrangements that allowed them to use ASR more effectively. For example, some of the injured users sat further away from their colleagues to lessen the impact of ambient workplace noise on recognition accuracy and also to minimize the distraction to others of constantly speaking to a computer. Other injured users had higher cubicle walls than was normal. Some of the Hansard editors had small single-person “booths” where they worked and others worked in four-person offices that were specially configured with sound-absorbing panels.

3.3 SUMMARY OF ANALYSIS

It is a cliché to say that ASR is difficult to use or that using software is a situated experience. What the analysis of the interviews with ASR users showed is that using ASR in the workplace is highly dependant on the quality of the relationships between many heterogeneous elements including ASR, other software, hardware that makes using ASR possible and the social world in which the use of ASR takes place. The ease with which ASR is used depends on the stability of the actor-network that is composed of those heterogeneous elements. ANT says that a user/actor must actively maintain its network or the network will fail. The analysis, above, demonstrates some of the difficulty that the injured users have in establishing and maintaining the actor-networks that make using ASR possible in their work. By nurturing their networks most of the users were able to establish and maintain their practice of using ASR. Conducting all the heterogeneous elements in a network requires power that is often out of reach of single users in an organisation. The Hansard department is a good example of a powerful actor sustaining the use of dictation software. The Hansard department is powerful enough that it can compel the heterogeneous actors to follow the scripts assigned to them. The injured interviewees have to work much harder to maintain their networks and their practice of using ASR. Once practice is established, it becomes easier to sustain the existing networks. Practice stabilises networks.

4 REFLECTIONS: IMPLICATIONS FOR DESIGN RESEARCH

In the research reported in this paper, an ANT-derived methodology is used to analyse fieldwork and interviews with users of ASR software. The results of this analysis are highly specific though there are overriding themes that emerge regarding the actual experiences of ASR users. The problem for design research is how to translate the specificity of these findings into results that can be generalized and applied elsewhere, both in the design of future ASR systems and more broadly in other areas.

The implication for future ASR systems is that the usefulness of ASR seems not to be a function solely of recognition accuracy but instead seems to be highly dependant on the alignment of the work process to the capabilities of ASR. This was illustrated most clearly in the interviewees who left jobs where they were unable to use ASR for jobs that were a closer match to what ASR made possible. The second implication for ASR is that using ASR productively in a workplace, even if the work is a good match for what ASR can do, seems to be influenced by social, political and indeed architectural factors. While it can be argued that issues of this sort are outside of a designers' direct influence they can not be disregarded as they may affect the eventual successful use of a designed product.

These insights were used to propose a design for a future speech recognition system (Kraal, 2006; Kraal Dugdale and Collings, 2006;).

It is possible to extend these recommendations more widely than ASR, though confirming their application more widely obviously requires more work. It can be said, however, that the analysis provided above contributes to an appreciation among design researchers of the (increasing) complexity of the environments in which designed products will be used.

Finally, this work has shown that an ANT-derived analysis can produce results that are both specific and generalisable.

REFERENCES

- Abdelnour Nocera, J. L. and Hall, P. (2004). Global software, local voices. In F. Sudweeks and C. Ess (editors), *Cultural attitudes towards communication and technology 2004*, pages 29–42. Karlstad, Sweden.
- Akrich, M. (1992). The de-scription of technical objects. In W. Bijker and J. Law (editors), *Shaping technology/building society*, pages 205 –224. MIT Press.

- Callon, M. (1986a). The sociology of an actor-network: The case of the electric vehicle. In M. Callon, J. Law, and A. Rip (editors), *Mapping the dynamics of science and technology: sociology of science in the real world*, chapter 2, pages 19–34. Macmillian, Houndsmills.
- Callon, M. (1986b). Some elements of a sociology of translation: domestication of the scallopes and the fishermen of St Brieuc Bay. In J. Law (editor), *Power, Action and Belief*, chapter 10, pages 196–233. Routledge & Kegan Paul, London, Boston and Henley.
- Fitzpatrick, G. (2003). *The Locales Framework: Understanding and designing for wicked problems*. Kluwer Academic Publishers.
- Glasser, B. and Strauss, A. (1967). *The Discovery of Grounded Theory*. Aldine Publishing Co., Chicago.
- Graham, C., Cheverst, K., and Rouncefield, M. (2005). Technology for the humdrum: trajectories, interactional needs and a care setting. In *OZCHI '05: Proceedings of the 19th conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction*, pages 1–10. Computer-Human Interaction Special Interest Group (CHISIG) of Australia, Narrabundah, Australia.
- Hanseth, O. and Monteiro, E. (1998). Understanding information infrastructure. <http://heim.ifi.uio.no/~oleha/Publications/bok.html>. Retrieved on 28 January, 2006.
- Huang, X., Acero, A., Alleva, F., Hwang, M., Jiang, L., and Mahajan, M. (1999). From Sphinx-II to whisper—making speech recognition usable. In C.-H. Lee, F. K. Soong, and K. K. Paliwal (editors), *Automatic Speech and Speaker Recognition: Advanced Topics*, chapter 20, pages 481–508. Kluwer Academic Publishers, Norwell, MA, USA.
- Karat, C.-M., Halverson, C., Horn, D., and Karat, J. (1999). Patterns of entry and correction in large vocabulary continuous speech recognition systems. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 568–575. ACM Press, New York, NY, USA.
- Kraal, B. (2006) *Considering Design for Automatic Speech Recognition in Use*. Unpublished PhD Thesis. The University of Canberra, Australia.
- Kraal, Ben J. and Dugdale, Anni and Collings, Penny (2006) *Scenarios for Embracing Errorful Automatic Speech Recognition*. In *Proceedings OZCHI 2006* 206, pages pp. 341-344, Sydney.
- Latour, B. (1987). *Science in action : how to follow scientists and engineers through society*. Harvard University Press, Cambridge, Mass.
- Latour, B. (1992). Where are the missing masses? In W. Bijker and J. Law (editors), *Shaping technology/ building society*, pages 205 –224. MIT Press.
- Latour, B. (1996). *Aramis, or, The love of technology*. Harvard University Press, Cambridge, Mass.
- Law, J. (1986). On methods of long-distance control: vessels, navigation and the Portuguese route to India. In J. Law (editor), *Power, Action and Belief*, chapter 10, pages 196–233. Routledge & Kegan Paul, London, Boston and Henley.
- Law, J. (1987). Technology and heterogeneous engineering: The case of the Portuguese expansion. In W. Bijker, T. Hughes, and T. Pinch (editors), *The Social Construction of Technological Systems*. MIT Press, Cambridge, Ma.
- Law, J. (2003). *Traduction/trahison: Notes on ant*. <http://www.lancs.ac.uk/fss/sociology/papers/law-traduction-trahison.pdf>. Retrieved from theWorldWide Web on Monday July 11, 2005.
- Tatnall, A. and Gilding, A. (1999). Actor-network theory and information systems research. In *Proceedings of the 10th Australasian Conference on Information Systems*. Wellington, New Zealand.