



Artificial Companions as a new kind of interface to the future Internet

Yorick Wilks

Oxford Internet Institute
1 St Giles
Oxford OX1 3JS
United Kingdom

www.dcs.shef.ac.uk/~yorick

This research report seeks to connect the future of the Internet to a new, even though relatively developed, technology; that of computer speech and language and its embodiment in a concept I shall call an Artificial Companion. I will argue here that such entities are coming into being and that they will change fairly rapidly from the rather primitive entities now available, largely in Japan, to being entities that will lead to profound social and psychological consequences; they will also constitute a new and much more personal way of dealing with information and its so-called 'overload' upon all of us.

This is not a paper in social science, but rather in speculative technology: however, the underlying technologies exist already and I will briefly describe them, along with some account of the current debates over their consequences. The crucial move in the paper will be when, after describing Artificial Companions, real and possible, I go on to argue that they can be seen as links to the Internet, at least for vulnerable classes of people (the old, the young) but perhaps for all of us when faced with the coming torrent of information on the Internet, particularly information about ourselves. Most citizens will not be able to deal with this at all, let alone control or 'own' it, without some assistance, and one form that assistance will take will be the persistent, personal, Internet agent, or Companion.

An assumption of the paper is that the jury is still out on whether the Internet leads to greater social inclusivity or not. The modern optimistic tradition (Castells, 1996) is that it does, and one can point to the almost total penetration of modern societies by the phone, and then the mobile phone, and assume that the Internet will be the same within some relatively short period. There is increasingly little distinction between mobile phone interfaces and the Internet, as they become screen interfaces to the World Wide Web, and as the Internet begins to take over more and more of phone call transmission using the VOIP protocol and forms of the interface pioneered by Skype.

On the other hand, there is still a more pessimistic tradition of technological determinism going back to Bell (1973) who cited the easy lives of the Eloi in H.G. Wells 'The Time Machine', as contrasted with the dark, subterranean, post-proletarian, Morlocks excluded from that life. Those who point to the plateauing of Internet penetration, and even its explicit rejection in some quarters¹ help keep the darker possibility alive.

¹ There are studies showing some proportion of the populations of developed countries are becoming disillusioned with the Internet and what it can provide, but nothing at the level of the spectacularly wrong (Investor's Business Daily 10 Jan 96): A senior VP at International Data Corp. predicts widespread consumer disillusion with the Internet: *'We also believe the Internet and World Wide Web phenomenon will go from an intoxication stage to a hangover stage during the next two years. About 20% of the large firms will jump off the Web because the consumer demographics are not developing quickly enough. But just as those 30 to 35 companies jump off, another 175 will hop on. The growth in online subscribers will continue to be very strong, but there will be underlying high turnover rates as users who are underwhelmed by content on the Web cancel and turn on their TVs'*.

What is already clear is that the Internet serves very well the academic and professional groups who invented and developed it, but it is far from clear that it serves the majority of the population well much beyond cheap holidays and pornography, and certainly not the elderly who still find video and even tape recorders hard to manage. The argument of this paper concerns the introduction of a relatively well-developed technology (Language and Speech Technology) as a remedy for this exclusion of targeted sections of the population. Here, for illustration, we shall concentrate on the elderly, who are not only relatively excluded from the Internet but also, in most developed societies, relatively better off than they were ten or twenty years ago, and so might be expected to purchase any technology that might re-enfranchise them into modern technological life, if it were made sufficiently attractive to them. The aim of the paper is to show that the Companion could be that right sort of concept for that purpose.

Language and speech technologies have come a long way in the last twenty years (see Cole et al., 1998), and most people encounter them in a form such as the dictation typewriters from IBM, or the provision of automatic translation of webpages by Babelfish, or perhaps when ordering train tickets from an automated call handler over the phone. Taking the Internet in a broad sense, one that now does not distinguish much between screen interfaces and phones, we shall see these technologies first in our mobiles, and they will pop up in odd and interesting contexts, like the Hewlett Packard initiative (Shalakova and Tucker, 2004) in Africa to put out basic webpage information, such as weather and farm commodity prices in nearby cities, but delivered over a mobile phone by an automatic speech generation system driven by a machine translation system—thus putting out, say, the weather forecast automatically in Zulu and Xhosa over a phone but both voices being artificial and translated from the same information on a webpage in some quite different language such as English or French. One can imagine a cut-rate starter Internet like that, requiring nothing but a mobile phone and no literacy, but being of huge value to poor African farmers.

However, we shall concentrate here on the classic Internet/Web combination, and not view our proposals in terms of other functionalities, such as doing more with mobile phone technology, even though, as is now clear with the recent deployment of fully functional screens on mobiles, the phone/terminal distinction is ceasing to have meaning.

Although what we discuss will be the Companion as a personal access agent to the classic Internet, it is clear that introductory, trial, forms of Companions are already appearing on the market. A Hong Kong firm has marketed an avatar for phones called Genivieve, which has a synthesised voice and an attractive female avatar but who must be typed to, and is said to answer questions and to reply in a number of languages. In April 2005 the BBC showed a documentary (Akino) about an elderly Japanese woman, Akino, who had been given a new furry computer Companion **Primo Puel**: she kissed it while it talked to her, and she said how safe she felt with it there, even when chattering away to itself in the next room. It was so much better against loneliness, she said, than praying in front of her dead husband's

shrine, which had been no help, she admitted. Puel is indeed a primitive Companion, not a robot but, even without the capability to move itself, it can record how Akino moves about the house and phone the health authorities if sensors in the rooms show that her routine changed. Our argument in this paper is that such Companions are already entering our society gently and by stealth, as it were, and we must think now about their technical basis, their limits, what role we want them to have, and how to protect ourselves from them and the effects of their arrival, should it become necessary.

But what will an artificial Companion be like? Who will need them and how much good or harm will they do? Will they change our lives and social habits in the radical way technologies have in the past: as trains, phones and television certainly have? Will they force changes in the law so that things that are not human could be liable for damages—until now, it is the case that if a machine goes wrong, it is always the maker or the programmer, or their company, which is at fault. Above all, how many people, with no knowledge of technology at all, such as the old and very young, will want to go about, or sit at home, with a companion that may look like a furry handbag on the sofa, or a rucksack on the back, but which will keep track of their lives by conversation, and be their interface to the rather elusive mysteries we now think of as the Internet or web.

One thing we can be quite sure of is that artificial Companions are coming. In a trivial way they have already arrived and millions of people have already met them. The Japanese toys Tamagochi (literally ‘little eggs’) were a brief craze in the West that saw sensible people worrying that they had not played with their Tamagochi for a few hours and that it might have begun to pine, where pining meant sad eyes and icons on a tiny screen, and playing with it meant pushing a feed button several times. The extraordinary thing about the Tamagochi (and later the Furby) phenomenon was that people who knew better began to feel guilt about their behaviour towards a small cheap and simple toy that could not even speak.

The brief history of those toys said a great deal about people and their ability to create and transfer their affections, despite their roughly correct knowledge of what is really going on inside their object. This phenomenon is almost certainly a sign of what is to come and of how easily people will find it to identify with and care for automata that can talk and appear to remember who they are talking to. This paper is about what it will be like when those objects are available: since most of the basic technologies such as speech recognition and simple machine reasoning and memory are already in place, this will not be long. Simple robot home helps are already available in Japan, but we only have to look at them and their hard plastic exteriors and their few tinny phrases about starting the dishwasher to realise that this is almost certainly not how a Companion should be.

Three technologies

Before moving to describe the integration that constitutes the Companion, we must first mention three technologies, not only in their own right but because, in each case, there have been misunderstandings about their achievements and goals. They are:

- Memories for Life and identity
- Language and speech technology
- Agents and The Semantic Web

The first of these is the major component technology that will make Companions possible. The second is Berners-Lee's (Berners-Lee et al., 2001) vision of how the Internet will change, and it is to that new Internet we intend the Companion as the human interface, on the ground that without it, the Internet may get harder not easier to use, and we shall return to the Semantic Web at the end of this paper. But as the Guardian reported a complainant who said that:

‘... the internet is killing their trade because customers ... seem to prefer an electronic serf with limitless memory and no conversation.’ (Guardian 8.11.03)

It is above all this lack of conversation with the Internet that the Companion seeks to remedy.

The third is a particular area of application that I argue will be crucial for the Companion, namely the owner's own life and its content as a principal focus of use of the future Internet, one where the Companion may well be a crucial tool for managing the content of the owner's life, which will soon be vast in information content terms.

Language and speech technologies are, for our purposes, two closely related methods for interfacing to the Internet; the first by typing to it to ask a question or to ask it to do something, and the second by speaking and listening, for the same purposes. The two are related, in that speech technology normally decodes speech waves—i.e. what is said into a microphone—into some form like written text inside a computer, which is then analysed so as to be understood, with the effect that both spoken and written input end up being analysed in similar ways by what we are calling ‘language technology’, which we can think of, loosely, as going from text to what it means.

The notion of a Companion

The paper introduces the notion of an Artificial Companion as a socially important paradigm for language and speech research in the next ten years: an intelligent and helpful cognitive agent which appears to know its owner and their habits, chats to them and diverts them, assists them with simple tasks but makes no technical demands on them at all, and might be most suitable for vulnerable social groups like the young and the old. The paper also discusses current aspects of the overall speech and language research program that a Companion will need.

The technologies needed for a Companion are very near to a real trial model; some people think that Artificial Intelligence (AI) is a failed project after nearly fifty years, but that is not true at all: it is simply everywhere. It is in the computers on 200-ton planes that land automatically in dark and fog and which we trust with our lives; it is in chess programs like IBM's Big Blue that have beaten the world's champion, and it is in the machine translation programs that offer to translate for you any page of an Italian or Japanese newspaper on the web.

And where AI certainly is present, is in the computer technologies of speech and language: in those machine translation programs and in the typewriters that type from your dictation, and in the programs on the phone that recognise where you want to buy a train ticket to, from among the four hundred or so British station names. But this is not a paper about computer technology any more than it is about robots; nor is it about philosophy.

Companions are not at all about fooling us as to their true natures, as in the Turing test scenario, because they will not pretend to be human at all: imagine the following scenario, which will become the principal one, running through this paper. An old person sits on a sofa, and beside them is a large furry handbag, which we shall call a Senior Companion; it is easy to carry about, but much of the day it just sits there and chats. Given the experience of Tamagochi, and the easily ascertained fact that old people with pets survive far better than those without, we will expect the Companion to be an essential lifespan- and health-improving object to own.

Other Companions are just as plausible as the Senior one, in particular the Junior Companion for children, that would probably take the form of a backpack, a small and hard to remove backpack that always knew where the child was. But the Senior Companion will remain our focus, not because of its obvious social relevance and benefit, possibly even at a low level of function that could be easily built with what is now available in laboratories, but because of the particular fit between what a Companion is and old people's needs.

Common sense tells us that no matter what we read by way of official encouragement, a large proportion of today's old people are effectively excluded from information technology, the web, the Internet and advanced

mobile phones because ‘they cannot learn to cope with the buttons’. This can be because of their generation or because of losses of skill with age: there are talking books in abundance now but many, otherwise intelligent, old people cannot manipulate a tape recorder, which has too many small controls for them with unwanted functionalities. All this is obvious and well known and yet there is little thought as to how our growing body of old people can have access to at least some of the benefits of information technology without the ability to operate a PC or even a mobile phone.

After all, the needs of the elderly are real—not just to have someone to talk to, but to deal with correspondence from public bodies, such as councils and utility companies demanding payment, with the need to set up times by phone to be visited by nurses or relatives, how to be sure they have taken the pills, when keeping any kind of diary may have become difficult, as well as deciding what foods to order, even when a delivery service is available via the net but difficult in practice for them to make use of.

In all these situations, one can see how a Companion that could talk and understand on the phone, and also gain access to the web, as well as to process written text in email could become an essential mental prosthesis for an old person, one that any responsible society would have to support. But there are also aspects of this which go beyond getting information, such as having the newspapers blown up on the TV screen till the print was big enough to be read, and dealing with affairs requiring some degree of reasoning, like paying bills from a bank account.

We have talked of Companions as specialised computer agents for tasks as simple as using the web to find a supermarket’s home delivery service for groceries. More interestingly, it may involve using the web to find out what has happened to their old school friends and workmates, something millions already use the web for. But we shall need some abstract notion of time lines and the coherence of life events on the web to sort friends and schoolmates from the thousands of other people with the same names.

The reasoning technologies we shall need to organise the life of a Companion’s owner may turn out to be the very same technologies needed to locate other individuals on the web and select them out from all the personal information about the world’s population that fills up the WWW, given that the web is now not just for describing the famous but covers potentially everyone. Two of my friends and colleagues who are professors of computer science have some difficulty distinguishing, and maintaining a difference, between themselves on the web and, in one case, a famous pornography supplier in Dallas, and in another case a reasonably well known disc-jockey in Houston, all of whom are highly ranked by the Google algorithm (Page et al., 1998).

These problems—of sorting out who exactly web information is about—will soon become not just quirky but the norm for everyone, and what I shall want to argue later is that the kind of computer agency we shall need in a Companion, one that deals with the web for us if we are old or maybe just lazy, is in fact closely related to the kind of agency we shall need to deal with

the web in any case as it becomes more complex. To put this very simply: the web will become unusable for non-experts unless we have human-like agents to manage its complexity for us. The Internet/web itself must develop more human-like characteristics at its peripheries if it is to survive as a usable resource and technology: just locating a particular individual on the web, when a majority of the EU and US populations have a web presence, will become far more difficult and time consuming than it is now. If this argument is right, Companions will be needed by everyone, not simply the old, the young and the otherwise handicapped. It is going to be impossible to use the web without its having some kind of a human face.

The notion of a Companion developed so far is anything but superhuman; it is vital to stress this because some of the public rhetoric about what companionable computers will be like has come from films such as *2001*, whose computer HAL is superhuman in knowledge and reasoning. He is a very dangerous Companion, and prepared to be deceptive to get what he wants, which may be not at all what we want. Seymour Papert at MIT always argued that it was a total misconception that AI would ever try to model the superhuman, and that its mission was to model the normal, which was much the same as AI-pioneer John McCarthy's emphasis on the importance of common sense reasoning was on capturing the shorthand of reasoning, the tricks that people actually use to cope with everyday life. Only then would we understand the machines we have built and trained and avoid them becoming too clever or too dangerous. This same impetus was very much behind Asimov's Laws of Robotics, which set out high-level principles that no robot should ever break if it is to bring no harm to humans.

The difficulty with such principles is fairly obvious: if a machine were clever enough it would find a way of justifying (to itself) an unpleasant outcome for someone, perfectly consistently with acceptable overall principles. Doing that has been a distinctively human characteristic throughout history: one thinks of all those burned for the good of their own souls and all those sacrificed so that others might live. In the latter case, we are probably grateful for those lost in what were really medical experiments—such as the early heart transplants—even though they were never called that.

It will not be possible to ignore these questions when presenting Companions in more detail, and in particular the issue of where responsibility and blame may lie when a Companion acts as a person's agent and something goes wrong. At the moment, Anglo-American law has no real notion of any responsible entity except a human, if we exclude Acts of God in insurance policies. The only possible exception here is dogs, which occupy a special place in English law, at least, and seem to have certain rights and attributions of character separate from their owners. If one keeps a tiger, one is totally responsible for whatever damage it does, because it is *ferae naturae*, a wild beast. Dogs, however, seem to occupy a middle ground as responsible agents, and an owner may not be responsible unless the dog is known to be of 'bad character'. We shall return to this later and argue that we may have here a narrow window through which we may begin to introduce notions of

responsible machine agency, different from that of the owners and manufacturers of machines.

It is easy to see the need for something like this: suppose a Companion told one's grandmother that it was warm outside and, when she went out into the freezing garden believing this, she caught a chill and became ill. One might well want to blame someone or something in these circumstances and would not be happy to be told that Companions could not accept blame and that, if one read the small print on the Companion's box, one would see that the company had declined all responsibility and had even got one to sign a document accepting this. All this may seem fanciful, and even acceptable if one's grandmother recovered and the company gave the Companion a small tweak so it never happened again.

This story makes no sense at the moment, and indeed the Companion might point out with reason, when the maintenance doctor came round, that it had read the outside temperature electronically and could show that it was a moderate reading and the blame should fall on the building maintenance staff, if anywhere. These issues will return later but what is obvious already is that Companions must be prepared to show exactly why they said the things they said and offered the advice they did.

A Companion's memory of what it has said and done may be important, but will be used only rarely one hopes; though it may be necessary for it to repeat its advice at intervals with a recalcitrant user: 'You still haven't taken your pills. Come on, take them now and I'll tell you a joke you haven't heard before'. James Allen in Florida is already said to have modelled a talking companionable pill for the elderly!

Life memories as a data problem on the Internet

It is reliably reported that many old people spend much of their day looking and sorting over photographs of themselves and their families, along with places they have lived and visited (Munro et al., 1999). This will obviously increase in absolute numbers as time goes on and everyone begins to have access to digitised photos and videos throughout their lives. One can see this as an attempt to create a coherent narrative of one's life: what drives the most literate segment of the population to write autobiographies even when, objectively speaking, they may have lived lives with little to report. But think of what will be needed if such a huge volume of material is to be sorted and overseen: it will not be long before the average person will be the owner of literally millions of digital images. How could the average person with no classification or editing skills even begin to make a coherent shape of such a mass of data? That problem is independent of whether one thinks of that data as being essentially on a private device (e.g. a laptop hard-drive) or on an Internet server; and of course this is also a distinction that commercial pressures will increasingly make obsolescent, as corporations offer

Internet/phone storage space to individuals in return for access to it (see Ferguson, 2005)—Google and Vodafone are both pioneering such schemes. A closely related notion of managing the totality of ‘memories for life’ has been developed, and has recently been declared a major challenge for future UK computing (see Memories for Life).

One could see what we are discussing here as no more than democratising the art of autobiography which was about Chaps, as Benchley readers will remember (whereas the art of Geography was about Maps). And this will mean far more, technically speaking, than simply providing ways in which people with limited manipulative skills can organise photos and videos into some kind of order on a screen: automated assistance will also require some guiding intelligence to provide and amplify a narrative that imposes a time order on this personal data. Lives have a natural time order, but this is sometimes very difficult to impose and to recover; even those with no noticeable problems from ageing find it very hard to be sure in what order two major life events actually happened: ‘I know I married Lily before Susan but during which marriage did my father die?’ would be a perfectly reasonable question that a life-organising Companion might assist with.

The frivolous example is to illustrate—and we shall see something of how it is actually done later on—how an artificial agent might assist in bringing events, whether in text or pictures, into some single coherent order so that one knew some things were before others, even if there were some events (e.g. Teddy’s and Joan’s marriages) that we could not order with certainty. This is the kind of thing today’s computers can be surprisingly good at, but it is a very complex and abstract notion, that of the time ordering of events, which can be simple (I know James was born before Ronnie) or only partial in some situations, such as whether I might know my brother’s children were born after my marriage and before my wife died, but not in what order they came.

These may seem odd or contrived questions but they do represent real problems at the border of memory and reasoning for many people, especially with age.

Speculations about the storage and organisation of all the information in a human life have already taken concrete form. DARPAUS launched the Lifelog program and then cancelled it after adverse publicity. Microsoft has begun such a program (see MyLifeBits) and the UK Science and Engineering Council launched an expertise network Memories for Life (see above) to explore the consequences of the possibility that all the experiences of an 80-year life could be stored in 28 terabytes (Alan Dix’s estimate). Meanwhile a range of websites and applications associated with photo libraries have been set up to annotate and organise photographic memories, of which the best known is FLIKR (www.flickr.com).

Another reason why this notion of ordering life narratives is so important, for a Senior Companion and elsewhere, is that it is also a way of separating different but similar lives from each other on the web, and these two notions are closely related as we shall show. This is jumping ahead a little, but many

people know the experience of searching for, say, 'George Bush' in Texas with Google on the web and finding there are about 25 of them who merit some attention from Google. Since two of them have been US Presidents, they cannot be distinguished from each other by name and job alone, and one must then use life events, dates of birth and so on to separate them. To put the whole thing simply, distinguishing closely related or confusingly named people on the web requires something like a coherent life line of some of the events in their lives, which is the same notion we have been discussing for imposing coherence on the, possibly muddled, whole life memories of old people. We shall return to this issue in the next section.

There is another aspect to the possibility of storing a life on the web, one which is not so much for the user as for their relatives later on. One can reasonably suppose that years of talking with an elderly user, and helping them organise their memories, will mean that the Companion is essentially the 'software manager' of the life of its user, the life-story that has been built up from those hours of conversation. But what should now happen to this biographical account that the Companion has, and which could exist in several forms: e.g. a document the Companion could print, or a collage of things said by the user and Companion put together coherently from recorded pieces, or even a single autobiographical account in the Companion's version of the user's own voice. This is not at all far-fetched: there are speech generation packages now available that can be trained to imitate a particular person's voice very accurately with plenty of training time, and time together is exactly what the Companion and user would have.

I would suggest that memoirs like these, produced over a long period by the Companion, are just what the user's relatives will want after the user's death, something more coherent and less gruesome than the tiny video clips one can now find and run at the touch of a button on some Italian tombstones. The need for these is now greater than it was, as people live farther from their parents when old and see them less. Many people have wished they had spent more time discussing a parent's earlier memories before their deaths: how one's parents had met and fallen in love, for example, but then suddenly it is too late to ask, or unobtainable because of shyness on the part of parent or child. This kind of limited memoir is an important role a Companion might come to play in society; experience may well show that old people will reveal memories and anecdotes to a Companion they would not feel able to tell even close relatives. There is a long tradition in AI of arguing that people may sometimes prefer machines to people in certain roles: AI-pioneer Donald Michie always claimed that drivers always preferred traffic lights (or 'robots' as they are still, significantly, called in some parts of the world, like South Africa) to policemen on traffic duty.

Memory and identity

The possibility of a Companion constructing and storing a biography for its owner raises in a novel form the issue of identity: is the Companion to be distinguished from the owner whose life it knows so well and whose voice it will almost certainly be able to imitate? This may be just another version of the old joke about pets getting like their owners, since we have chosen to describe the Senior Companion as a kind of pet. Other forms of identity will also be touched on later, in particular the identity of the Companion as an authorised agent for the owner, and its similarity and distinctness from the owner, while on the other hand functioning as a web agent in the world of electronic information. We have not yet properly introduced the notion of a web agent: it is, very roughly, an intelligent software package that one can encounter on the Internet and which will give expert advice. Web agents now make very limited deals and transactions of all sorts with each other, and learn to trust each other, as they already do in a rudimentary sense in the world of banking, where automatic agents in their thousands clinch money transactions between financial institutions (see Ramchurn et al., 2004). The kind of agent now visible to a user, but still rudimentary, are those which will search the whole web to find the cheapest source of, say, a particular type of camera.

Two other issues of identity in an artificial Companion world will also draw upon other ways in which identity has become an issue for Internet users and which are relatively well known. The first, known to all newspaper readers, is that of chat room users who pretend to be what they are not; normally this is quite innocent pretence, and little more than hiding behind a pseudonym during conversations and sometimes pretending to be a different kind of person, often of the opposite sex. In that sense, the Victorian game of sex-pretending, on which Turing based his famous imitation game for computers, has come back as a widely played reality. The problems only arise, and they are very real, when impressionable people, usually children, are lured into meetings with people who have been encountered under false pretences.

The other issue of identity, which is the core problem for web searchers, is finding too many people under the same name with a Google search and trying to find the right one, or even how many there are, a topic we already touched on above. It is a much researched problem at the moment, namely how to sort out exactly how many people there are in the pages retrieved by Google for a given name. One can see the pressing interest here, in that many scientists now rate how famous they are by how many Google hits their name gets, compared to their competitors. But how can they be sure all those retrieved are really themselves? A well known British computing professor has a very long list indeed, which is boosted by a mid-Western disc jockey with the same name!

The George Bush example, mentioned earlier, suggests that the best way must require something like a rule that looks at dates and times, on the assumption that two people with the same name are very unlikely to have the

same date of birth. And other rules will almost certainly deal with aspects of people, such as their occupations. This method would not help, of course, with the unusual but very real cases like Vanbrugh, the 18th Century playwright, and his separation from the 18th Century architect of Blenheim Palace of the same name who, amazingly enough, were one and the same person, hard though it would be for most rules (and people) to accept.

There is a further kind of complication in that, even if we could sort out muddled identities of these kinds, given enough information, in some cases people do not agree on how many objects or people there are under discussion, so that it becomes a matter of conflict, or individual belief, how many people are being talked about. In the case of Vanbrugh we can imagine a strained conversation between a person sure they were two similarly named individuals and someone else who knew they were not. It is, as one could put it, very difficult but just possible for people to communicate who do not agree on what things there are in the world, on the ontology, to use a word in its original sense that is now normally used to mean something quite different. Later we shall touch on methods of analysis and descriptions of knowledge that might allow people to reason with others, or their Companions, when they disagreed in this way about what there is out there. How should a Companion discuss relatives when it was sure Jack was Ethel's auntie, and its owner even said that the day before, but is now convinced they are quite different people? This is a deep matter to which we shall return later, and is discussed in detail in Ballim and Wilks (1991).

Problems of identity will arise both in the context of representing individuals in a Companion's world, which is very much based on that of its user, one the Companion seeks to learn, and also in the wider world of information, which for convenience we will often identify with what can be found on the web or Internet. It is quite normal now to hear people speak of 'virtual worlds' in connection with the web and Internet, although it is usually unclear exactly what they have in mind. The only obvious place where virtual worlds belong naturally is in computer games, whose interaction with the web will be crucial in the case of Companions for the young, who spend more time in games worlds than the rest of the population.

The general issue here is in the interaction of these considerations with identity; having a verifiable identity is part of what it means to be a human being, or at least a modern human being. If a Companion is to have human-like characteristics, one will want to explore how its identity can have human-like features, as opposed to machine-like features where identity is usually a trivial matter: a car is identified uniquely simply by the sum of the numbers stamped on its chassis and its engine and there is no interesting issue about that outside the world of car fraud.

If a Companion is to be an interface to the Internet, say, for a user who is technologically incompetent yet who must conform to the standards of identity that society requires and imposes, then the Companion will have to understand identity to some degree and possibly be able to manipulate slightly different forms of it. In the US and UK, identity is currently established

by a range of items with numbers, from passports through credit cards to health, driving licence and tax numbers (some with associated passwords or PINs), with the Social Security number having a definite primacy in the US. In most EU countries there is a single ID number, of which the simplest is on the life-long single Personnummer in Sweden. States prefer a citizen to be identified by a single number and in the UK there is currently strong pressure for something closer to the Swedish model, although UK law has, at the moment, no clear definition of identity (with legally registered unique names and addresses, as in most of the EU) and there is no legal problem in the UK with having a number of identities simultaneously and bank accounts for each (as a famous case brought, and lost, by the Post Office showed some years ago) so long as there is no intention to defraud.

All this is important since identity checks are the basis of all web transactions and if a Companion is to deal with a person's affairs it might need something approaching a power of attorney or at least an understanding of how identity is established in web transactions, as well as a method for establishing that its owner approves of what it is doing in individual transactions, in case of later disputes (e.g. by angry relatives after an old person's money has been spent). Dave Birch (2005) has recently described a scenario in which security of identity can be achieved without the imposition of unique identity, which security-minded authorities would certainly resist but which may be very important to someone who feels that they have a right to buy something on the Internet while revealing only, say, their age but not who they actually are.

All these issues are the subjects of active research programs, but what I am introducing here is the possibility that a Companion, as a new kind of artifact among us, may focus our minds on a set of intellectual and practical issues in a new way, even though some of them are very traditional issues indeed.

The state of language and speech technology

How does this rather airy vision connect to the general state of R&D in speech recognition and natural language processing at the moment? My own belief is that most of the components needed for a minimally interesting Companion are already available; certainly the Companion is not particularly vulnerable to one major current technical weakness, namely the imperfect recognition rate of available Automatic Speech Recognition (ASR) systems. This, of course, is because a Companion is by definition dedicated to a user and so the issue of user-independent ASR does not initially arise, except when the Companion needs to make its own phone calls and understand what is said to it.

However, the Companion is not merely an application wholly neutral between current disputes about how best to advance speech and language systems, in part because it will surely need a great deal of representation of human knowledge and belief and therefore the Companion's development would seem to need overall approaches and software architectures that allow such

representations and, ultimately, their derivation from data by machine learning. This last clause is very important because there has been a profound methodological shift in speech and language research in the last two decades. Before that, it was generally assumed that the knowledge of the world and of language that a machine intelligence required could be programmed in directly, the content being provided by the researcher's intuition. In the case of language, this assumption followed directly from Chomsky's (1972) approach to linguistics: that intuitions about the nature of language can be computed by rules written by experts who have intuitive knowledge of their (native) language.

All this has now turned out to be false: no effective systems have ever been built on such principles, nor (outside machine translation, perhaps) are they ever likely to be. The revolution that has replaced those doctrines holds that such knowledge, world or linguistic, must be gained from data by defensible (i.e. non-intuitionistic) procedures like machine learning.

In the late 1980s when symbolic natural language processing (NLP) was invaded by an empirical and statistical methodology driven by recent successes in speech processing. The shock troops of that invasion were the IBM team under Jelinek which developed a wholly novel statistical approach to machine translation (MT), one that was not ultimately successful (see Wilks 1994 for a discussion) but did better than anyone in conventional MT initially expected, and set in train a revolution in methodology in NLP as a whole.

Although the IBM team began without any attention to the symbolic content of linguistic MT, they were forced, by their inability to beat conventional MT systems in DARPA competitions, to take on board traditional linguistic notions such as lexicons, morphology and grammar, but they imported them not from intuitions but in forms such that they could be learned in their turn and that fact was the ultimate triumph of their revolution.

The present situation in dialogue modeling—such as will be needed for a Companion—is in some ways a replay, at a lower level, of that titanic struggle. The introduction into ASR of so called 'language models'—which are usually no more than corpus bi-gram statistics to aid recognition of words by their likely neighbours—have caused some, like Young (2002) to suggest that simple extensions to current speech (ASR) methods could solve all the problems of language dialogue modeling.

Young describes a complete dialogue system seen as what he calls a Partially Observable Markov process, of which subcomponents can be observed in turn with intermediate variables and named (in order):

- Speech understanding
- Semantic decoding
- Dialogue act detection
- Dialogue management and control
- Speech generation

Such titles are close to conventional for an NLP researcher, e.g. when he intends the third module as something that can also recognise what we may call the *function* of an utterance, such as that it is a command to do something and not a pleasantry. Such terms have been the basis of NLP dialogue pragmatics for some thirty years, and the interesting issue here is whether Young's Partially Observable Markov Decision Processes, are a good level at which to describe such phenomena, implying as they do that the classic ASR machine learning methodology can capture the full functionality of a dialogue system, when its internal structures cannot be fully observed, even in the sense that the waves, the phones and written English words can be. The analogy with Jelinek's MT project holds only at its later, revised stage, when (as we noted earlier) it was proposed to take over the classic structures of NLP, but recapitulate them by statistical induction. This is, in a sense exactly Young's proposal for the classic linguistic structures associated with dialogue parsing and control with the additional assumption, not made earlier by Jelinek, that such modular structures can be learned even when there are no distinctive and observable input-output pairs for the module that would count as data by any classic definition, since they cannot be word strings but symbolic formalisms like those that classic dialogue managers manipulate.

The intellectual question of whether the methodology of speech research, tried, tested and successful as it is, can move in and take over the methodologies of language research may seem to many a completely arcane issue, like ancient trade union disputes in shipbuilding, say, as to who bored the holes and who held the drills. But, as with those earlier labour struggles, they seem quite important to the people involved in them and here, unlike shipbuilding, we have a clash of expertise but no external common-sense referee to come in and give a sensible decision.

Jelinek's original MT strategy was non/anti-linguistic with no intermediate representations hypothesised between speech input and speech output, whereas Young assumes roughly the same intermediate objects as linguists but in very simplified forms. So, for example, he suggests methods for learning to attach Dialogue Acts to utterances but by methods that make no reference to linguistic methods for this (known since Samuel et al., 1998) and, paradoxically, Young's equations do not make the Dialogue Acts depend on the words in the utterance, as all linguistic methods do. His overall aim is to obtain training data for all of them so the whole process becomes a single throughput Markov model, and Young concedes that this model may only be for simple domains, such as, in his example, a pizza ordering system.

All parties in this dispute, if it is one, concede the key role of machine learning, and all are equally aware that structures and formalisms designed at one level can ultimately be represented in virtual machines of less power but more efficiency. In that sense, the primal (Chomsky, 1959) dispute between Chomsky and Skinner about the nature of the human language machine was quite pointless, since Chomsky's transformational grammars could be represented, in any concrete and finite case, such as a human being, as a finite state machine, of the sort espoused by Skinner.

All that being so, researchers nonetheless have firm predelictions as to the kinds of design within which they believe functions and capacities can best be represented, and, in the present case, it is hard to see how the natural clusterings of states that form a topic (such as, for example, how to build a jet plane, piece by piece) can be represented in finite state systems. It is equally difficult to see how the human ability to return in conversation to a previously suspended topic can be represented plausibly in such a way. But these are all matters that can be represented and processed naturally in well understood virtual machines above the level of finite state matrices (see Wilks et al., 2004).

There is no suggestion that a proper or adequate discussion of Young's views has been given here, only a plea that machine learning must be possible over more linguistically adequate structures than finite state matrices if we are to be able to represent, in a perspicuous manner, the sorts of belief, intention and control structures that complex dialogue modelling will need; it cannot be enough to always limit ourselves to the simplest applications on the grounds, as Young puts it, that « the typical system S will typically be intractably large and must be approximated ». In the end, the case put here may be no more than that the structures we use to represent our language, including to machines, must be comprehensible to us as humans.

The Semantic Web

Mention has been made earlier of the new form (Berners-Lee et al., 2001) of the WWW as envisaged by Berners-Lee and colleagues to follow his original conception. This is a large topic and suitable for a separate paper (e.g. Wilks, 2006) and can be seen in two quite different ways: first, as the existing WWW but augmented by annotations on the items of all the texts it contains, so as to give more direct access to the meaning content of the texts.

On this view, the Semantic Web (SW) is an outgrowth of both language technologies, as described above, and their notion of augmentations, which is partly inherited from initiatives in the Humanities (e.g. the Text Encoding Initiative; see TEI). These annotations could be seen as imposing a 'point of view' on the SW, so that, for example, it might be possible to use the annotations to prevent me seeing any web pages incompatible with The Koran, and that might be an Internet-for-me that I could choose to have. But there is no reason why such an annotated web should necessarily, as some have argued (e.g. Nelson, 2005), impose a unique point of view. The technology of annotations is quite able to record two quite separate annotation data (as meta-data) for the same texts, and no uniformity of point of view is either necessary or desirable.

The second view of the SW, and one that Berners-Lee prefers, is that of an Internet whose content is accessible to Agents, partly through annotations and partly through data-bases whose semantics are well-known and

understood. These agents (see e.g. Walton, 2006) operate on the Internet and provide services to customers, such as updating their diaries, finding cheap gas supplies etc. Such agents are therefore rather different from the concept of Companions, for they are transitory, and not designed for a permanent relationship with an owner based on extensive knowledge about the owner. One should note here, however, that contemporary work on the SW (e.g. Bontcheva et al., 2003; Ciravegna et al., 2003) has no need to choose between these two sources and functionalities I have distinguished, but rather seeks to combine both.

A third strand in the genesis of the SW is that of traditional AI itself and its long and honourable tradition of modelling reasoning, planning and knowledge representation. Some would argue the SW is no more than a weaker form of AI which has sacrificed representational power to gain a system that works on a large scale.

Companions will draw on all these strands in the SW as well as that of the ECAs, or Embodied Conversational Agents (see e.g. Ruttkay and Pelachaud, 2004), although these have conventionally been conceived of not in language terms but in graphical, avatar, glance, expression and presence terms—i.e. with the emphasis on the visual, whereas the Companion is fundamentally an agent that establishes a relationship through talking, with all that entails in terms of politeness, emotion, personality and how those slippery but real concepts can be modelled in automata. But again, none of these borderlines are firm: ECA: Companion, SW Agent: ECA, and most of the questions and technologies touched on in this paper apply not only to possibly permanent Companions but to a whole range of interactions with the Internet, from pseudo-boyfriends and girlfriends, to recent results on determining and simulating author personalities in weblog texts (see Oberlander and Nowson, 2006).

In the coming decade the European Commission is planning huge investments in all these technologies under its Information Society Technologies (IST) program, and the edges of this research and the barriers to its advance should be much clearer during the coming Seventh Framework Programme (The COMPANIONS project will be supported 2006–2010 as the Integrated Project **IST-34434: Intelligent, Persistent, Personalised Multimodal Interfaces to the Internet**).

References

- Akino: Japan's High Tech Carers on BBC Website (2004)
http://news.bbc.co.uk/1/hi/programmes/this_world/golden_years/4436633
- Ballim, A. and Wilks, Y. (1991) *Artificial Believers: the ascription of belief* (Lawrence Erlbaum: Hillsdale NJ).
- Bell, D. (1973) *The Coming of Post-Industrial Society: A Venture in Social Forecasting* (Basic Books: New York).
- Berners-Lee, T., Hendler, J. and Lasilla, O. (2001) *The Semantic Web*. Scientific American.
- Birch, D. (2005) *The Age of Sousveillance*. Guardian on Line:
<http://technology.guardian.co.uk/online/comment/story/0,,1527735,00.html>
- Bontcheva, K. and Cunningham, H. (2003) *Information Extraction as a Semantic Web Technology: Requirements and Promises*. Adaptive Text Extraction and Mining workshop.
- Castells, M. (1996) *The Rise of the Network Society* (Blackwells: Oxford).
- Chomsky, N. (1959) Review of Skinner's Verbal Behaviour. *Language* 35:26-58.
- Chomsky, N. (1972) *Language and Mind* (Harcourt Brace: New York).
- Ciravegna, F. (2003) Designing adaptive information extraction for the Semantic Web in Amilcare. In S.Handschuh and S.Staab (eds) *Annotation for the Semantic Web, Frontiers in Artificial Intelligence and Applications* (IOS Press).
- Cole, R., Mariani, J., Uszkoreit, H., Varile, N., Zaenen, A., Zampolli, A. and Zue, V. (1998) *Survey of the State-of-the-Art in Human Language Technology* (Cambridge University Press: Cambridge).
- Ferguson, C.H. (2005) What's Next for Google. MIT Technology Review
<http://www.technologyreview.com/articles/05/01/issue/ferguson0105.asp?trk=nl>
- FLIKR. <http://www.flickr.com/>
- Memories for Life and Photocopains. <http://www.memoriesforlife.org/>
- MyLifeBits (Microsoft)
<http://research.microsoft.com/barc/mediapresence/MyLifeBits.asp>
- Munro, A., Höök, K. and Benyon, D.R. (eds) (1999) *Social Navigation of Information Space* (Springer: London).
- Nelson, T. (2005) Personal communication.

- Oberlander, J. and Nowson, S. (2006) Whose thumb is it anyway?: Classifying author personality from weblog text. <http://www.hcrc.ed.ac.uk/~jon/papers/drafts/pc8.pdf>
- Page, L., Brin, S., Motawani, T. and Winograd, T. (1998) The PageRank Citation Ranking: Bringing Order to the Web. Stanford Digital Library Technologies Project.
- Ramchurn, S.D., Huynh, D. and Jennings, N.R. (2004) Trust in multiagent systems. *The Knowledge Engineering Review* 19(1).
- Ruttkay, Z. and Pelachaud, C. (2004) Evaluating Embodied Conversational Agents (Kluwer: Berlin).
- Samuel, K., Carberry, S. and Vijay-Shankar, R. (1998) Dialogue Act Tagging with Transformation-Based Learning. *Proc. COLING98*, Montreal.
- Shalanova, K. and Tucker, R. (2004) Issues in porting TTS to minority languages. *Outside Echo* at Hewlett Packard Laboratories, Bristol.
- TEI (Text Encoding Initiative). <http://www.tei-c.org/>
- Walton, C. (2006) Agents and the Semantic Web (Oxford: Oxford University Press).
- Wilks, Y. (1994) Stone Soup and the French Room: the empiricist–rationalist debate about machine translation. Reprinted in Zampolli, Calzolari and Palmer (eds) *Current Issues in Computational Linguistics: in honor of Don Walker* (Kluwer: Berlin).
- Wilks, Y., Webb, N., Setzer, A., Hepple, M. and Catizone, R. (2004) Machine Learning approaches to human dialogue modelling. In: Kuppervelt and Smith (eds) *Current and New Directions in Discourse and Dialogue* (Kluwer: Berlin).
- Wilks, Y. (2006) The Semantic Web and the Apotheosis of annotation. *Journal of Web Semantics* (submitted).
- Young, S. (2002) Talking to machines—statistically speaking. *Proc. ICSOS02*.