

## ARTIFICIAL COMPANIONS AS A NEW KIND OF DIALOGUE INTERFACE TO THE FUTURE INTERNET

*Yorick Wilks*

*University of Sheffield, UK.*

This paper seeks to connect the future of the Internet to a new, even though relatively under-developed, technology, that of computer speech and language and its embodiment in a concept I shall call an Artificial Companion. 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:

Language and speech technology

Agents and the Semantic Web

The first of these is the major component technology that will make Companions possible: it is the assumption that the whole of our life information over which we have control, texts, images, video etc., will be stored on and accessed through, the Internet. 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. The third notion above is that agents will change from transitory software entities that e.g. locate a cheap camera on the internet, to more permanent social Companion entities that deal with a user through dialogue over a long period, learn his or her needs and preferences and elicits large quantities of life data through conversation.

### *Introduction*

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.

### *Two component technologies*

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

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 it 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 pro-

---

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

---

grams 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 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 of 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 modeled a talking companionable pill for the elderly!

### *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

---

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

---

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 1980's 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 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 hypothesized 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 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 predilections 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 modeling 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 modeling reasoning, planning and knowledge representation. Some would argue the SW is no more than 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 of 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 modeled 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].

---

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

---

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*

1. Ballim, A., and Wilks, Y. (1991) *Artificial Believers: the ascription of belief*. Lawrence Erlbaum, Hillsdale NJ.
2. Berners-Lee, T., Hendler, J., and Lasilla, O. (2001). *The Semantic Web*. Scientific American.
3. Bontcheva, K., and Cunningham, H. (2003) *Information Extraction as a Semantic Web Technology: Requirements and Promises*. Adaptive Text Extraction and Mining workshop.
4. Chomsky, N. (1959) Review of Skinner's *Verbal Behaviour*, *Language* 35: 26-58.
5. Chomsky, N. (1972) *Language and Mind*, Harcourt Brace, New York.
6. 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.
7. Cole, R., Mariani, J., Uszkoreit, H., Varile, N., Zaenen, A., Zampolli, A., and V. Zue, (1998) *Survey of the State-of-the-Art in Human Language Technology*, Cambridge University Press.
8. Ferguson, C. H. (2005) *What's Next for Google*, In *MIT Technology Review*: <http://www.technologyreview.com/articles/05/01/issue/ferguson0105.asp?trk=nl>
9. FLIKR: <http://www.flickr.com/>
10. *Memories for Life and Photocopains*: <http://www.memoriesforlife.org/>
11. Oberlander, J. and S. Nowson, (2006) *Whose thumb is it anyway?: Classifying author personality from weblog text*.
12. <http://www.hcrc.ed.ac.uk/~jon/papers/drafts/pc8.pdf>
13. Page, L., Brin, S., Motawani, T., and T. Winograd (1998), *The PageRank Citation Ranking: Bringing Order to the Web*, Stanford Digital Library Technologies Project
14. Ramchurn, S. D., Huynh, D. and Jennings, N. R. (2004) *Trust in multiagent systems*. *The Knowledge Engineering Review* 19(1).
15. Ruttkay, Z., and C. Pelachaud, (2004) *Evaluating Embodied Conversational Agents*, Kluwer, Berlin.
16. Samuel, K., Carberry, S., and Vijay-Shankar, R. (1998). *Dialogue Act Tagging with Transformation-Based Learning*. In *Proc. COLING98*, Montreal.
17. TEI: <http://www.tei-c.org/>
18. Walton, C. (2006). *Agents and the Semantic Web*. Oxford, Oxford University Press.
19. 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
20. Wilks, Y., Webb, N., Setzer, A., Hepple, M., and Catizone, R. (2004) *Machine Learning approaches to human dialogue modelling*. In Kuppervelt, Smith (eds.) *Current and New Directions in Discourse and Dialogue*, Kluwer, Berlin.
21. Wilks, Y. (submitted 2006) *The Semantic Web and the Apotheosis of annotation*. *Journal of Web Semantics*.
22. Young, S. 2002. *Talking to machines—statistically speaking*, *Proc. ICSOS02*.