From Irish Rooms to Danish Rooms

Paul Mc Kevitt‡
Center for PersonKommunikation (CPK)
Fredrik Bajers Vej 7-A2
Institute of Electronic Systems (IES)
Aalborg University
DK- 9220, Aalborg
DENMARK, EU.
pmck@cpk.auc.dk

Abstract

There is a major motivating force which is driving the Humanities and Sciences/Engineering towards each other in the area of integration of language and vision processing by machines: SuperinformationhighwayS. This force is the ability now to have information in text, voice, sound, graphic and video forms available within minutes at local and remote sites through interfaces like Netscape[1] and search engines like AltaVista[2]. We have before proposed the Irish Room (see Mc Kevitt and Guo 1996) as a solution to the Searle's Chinese Room Problem and now also introduce it here in Denmark on a MultiLingual front. We also present IntelliMedia 2000+ a new project at Aalborg University, Denmark and these are a major part of the future of SuperinformationhighwayS.

1 Introduction

SuperinformationhighwayS are a major force driving the Humanities and Sciences/Engineering closer together as we find information in textual, voice, sound, graphical and video formats available over computer networks using interfaces like Netscape and search engines like AltaVista. People will be able to pose their queries for retrieving information about say stocks and shares, good restaurants in a city or their bank account by speaking that query to the machine. In turn, they will be able to direct the machine's graphical display of the information it is presenting in response.

Visual information comes in many formats from diagrams to videos as does language information both natural and formal. The Sciences/Engineering are more concerned with methods for transmitting, representing and retrieving information across networks while the Humanities are more concerned with the actual information itself. The area of Intelligent MultiMedia is concerned with computer understanding of information in multiple modalities such as language and vision and then the representation of it in terms of its semantics. This area involves technologies from the engineering side in terms of spoken language processing, natural language processing, Computer Science and Artificial Intelligence and from the Humanities side in terms of Linguistics, Cognitive Science, Psychology and studies of the mind.

We see SuperinformationhighwayS as bringing together the Humanities and Sciences/Engineering with a focus on language/linguistics and especially MultiLinguality and dictionaries/encyclopedias where the meaning and knowledge of words is defined. We believe that SuperinformationhighwayS will drive us towards the possibility of MultiModal dictionaries and encyclopedias which were difficult to construct beforehand on paper alone.

Today's dictionaries are sorely lacking in information that people have had in their heads for years. If one thinks of a *dog, a cat, parties, love, hate, sex, loud, bang, greasy, furry, running, jumping, swooning, ice cream*, etc. then one has a picture in one's head of these objects, emotions, sounds, feelings, and actions or some situation where they occurred in past personal history or in a film. Such pictures and sounds, and their manifestation in the symbols of written language itself were a major part of the emphasis of the writings of Joyce (1922, 1939) and oth-

‡Paul Mc Kevitt is also a British Engineering and Physical Sciences Research Council (EPSRC) Advanced Fellow at the University of Sheffield, England for five years under grant B/94/AF/1833 for the Integration of Natural Language, Speech and Vision Processing.

ers. For example, Joyce (1922) uses letters in English to produce the sounds of the waves as they come rushing towards the seashore on Dollymount Strand.

Today's dictionaries such as Longman's Dictionary of Contemporary English (LDOCE) (see Procter 1978), Collins' COBUILD (see Sinclair 1987) and Webster's (see Olney 1968), whether in text form or in electronic form do not contain much pictorial information; they typically encode words in symbolic natural language form with symbolic natural language descriptions. Encyclopedias do contain pictures but they do not contain definitions of words, rather knowledge about words and specifically objects in the world. It is not clear to us why dictionaries have had this bias towards symbolic natural language but it certainly seems very strange behaviour.

In the field of Artificial Intelligence (AI) where researchers are trying to build computer programs to do things people can do there has been little progress in the integration of Natural Language Processing (NLP) and Vision Processing (VP) although there has been some progress of late (see Denis and Carfantan 1993, Dennett 1991, Mc Kevitt 1994, 1996, and Pentland 1993). Dennett (1991, p. 57-58) says "Surely a major source of the widespread skepticism about "machine understanding" of natural language is that such systems almost never avail themselves of anything like a visual workspace in which to parse or analyze the input. If they did, the sense that they were actually understanding what they processed would be greatly heightened (whether or not it would still be, as some insist, an illusion). As it is, if a computer says, "I see what you mean" in response to input, there is a strong temptation to dismiss the assertion as an obvious fraud." Researchers in Natural Language Processing (NLP) have spent much of their time attempting to solve questions of ambiguities in sentences like "I saw the man on the hill with the telescope" without thinking that such ambiguities could already be solved if there was a picture together with the text. The dictionaries and knowledge bases for NLP suffered from two limitations: the split from vision processing, and simple symbolic natural language dictionaries.

There have been two problems that have worried us in the field of NLP for years: (1) grounding: where are symbolic semantic primitive meanings and even symbols themselves in NLP programs grounded? and (2) circularity: how come some words in dictionaries have circular definitions so that words end up defining each other? We believe that these two problems were caused in part by the fact that people were thinking of NLP without taking other perceptual sources into account. We argued in Mc Kevitt and Guo (1996) for the type of integrated representations which can be used to solve these problems for at least language and vision. Let's look at the problems in more detail.

2 Grounding primitives

The problem with Machine Readable Dictionaries (MRDs) such as LDOCE, and for that matter Machine Tractable Dictionaries (MTDs) produced from them, for years has been that words are defined in terms of a defining vocabulary of 2000 words but that there is no definition of the defining vocabulary itself. Also, in NLP systems like Schank's (see Schank 1972) and Wilks' (see Wilks 1973, 1977, 1978) where are the primitives grounded?

2.1 The Chinese Room

Harnad (1990) has brought the grounding problem further and said that symbolic processing systems have, in general a problem with grounding their symbols and that this problem can only be freed up by using other perceptual sources such as visual input. This is his answer to Searle's Chinese Room Problem[1] where Searle argues that a machine cannot "understand" the symbols it represents but just passes them around with no feeling for their meaning (see Searle 1980, 1984, 1990). The Chinese Room is shown in Figure 1 where the sentence "What are the directions of lexical research?" in Chinese is being passed to John Searle in the Chinese Room (note that John doesn't get the English translation which we have placed there for the reader). Searle is arguing that the machine behaves like a traditional MultiMedia system with little or no understanding of the meaning of its perceptual inputs.

2.2 The Irish Room

We have argued in Mc Kevitt and Guo (1996) that a computer should behave more like an Irish Room, going even further than Joyce (1922, 1939) where he tried to bring perception into written symbols. The Irish Room, shown in Figure 2, is where the sentence "What are the directions of lexical research?" in English is being passed to Seán the Leprechaun in the Irish Room (note that this time the input is annotated with icons). The Irish Room is one where a Leprechaun who understands Gaelic and who cannot understand English is locked in a room and has the

[1] Searle asked us to imagine a Chinese Room where a person who understands English and who cannot understand Chinese is locked in the room and has the task of using an English rule book for manipulating Chinese symbols. Then, to an outside observer, the person appears to be able to understand Chinese just as a computer program which manipulates symbols could appear to do so (see Searle 1984, 32-33).

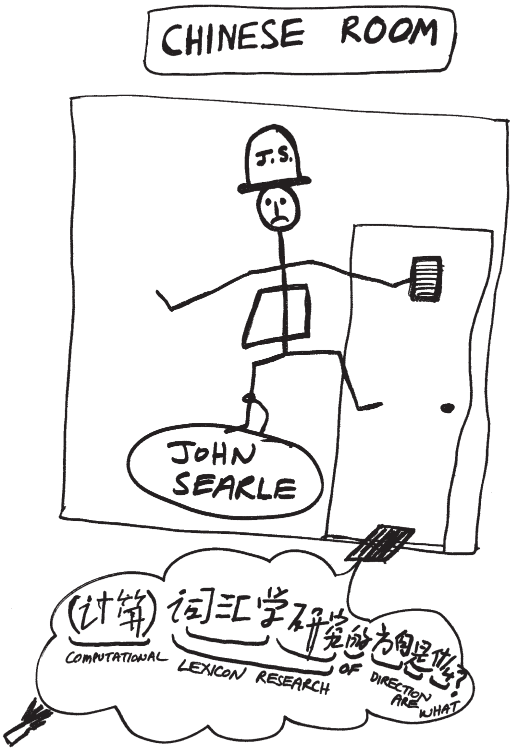

Figure 1: The Chinese Room

Figure 2: The Irish Room

task of using a Gaelic rule book for manipulating English words. Each English word has an icon or picture sequence attached to it. Then, to an outside observer, Seán the Leprechaun appears to be able to understand English just as a computer program which manipulates symbols could appear to do so. However, this time Seán begins to understand the English words because he/she has reference to their meaning. Sounds, smells and even touch can be added in later! This form of solution to the Chinese Room problem has also been suggested by Harnad (1990), and discussed at length by Marconi (1995) and Meini and Paternoster (1995).

2.3 Irish Room in a Chinese Room

In order to test whether the Irish Room idea works I tried an experiment at Tsinghua University in Beijing, China on August 20th, 1994 at 9.00-10.00 PM. I do not understand Chinese and I asked Jun-Ping Gong, Jin Zhang and Yan-Qing Yu to write a few sentences in Chinese and to annotate them with icons to see if I could guess what the sentences were. We worked with simple icons and I was able to guess much of the meaning of the simple examples used. We predict that with more fine tuned icons, videos, sounds and other sensory information the results would be much better.

2.4 Irish Room in a Danish Room

While I have been in Denmark it has come to my attention that there are not only a large number of Danish words which are similar to words in English but also to words in Gaelic. This introduces another phenomenon which I call the Irish Room in a Danish Room. The Irish Room in a Danish Room is where the sentence "I hvad retning gar leksikalsk forskning?" ("What are the directions of lexical research?" in Danish and literally, "In what direction goes lexical research?") is being passed to Seán the Leprechaun (who by now has also learned English) (or I) in the Irish Room (this time the input is also annotated with icons). Already, Seán the Leprechaun (and I) can see, and in particular hear, if there is spoken input, that 'Hvad' is like the English 'What' or the Gaelic 'Cad' (How) or a mixture of both and 'gar' like the English 'go' and 'leksikalsk' like the English 'lexical'! Hence, Seán and I can start processing the input before even looking at the icons.

The Irish Room in a Danish Room is where a Leprechaun who understands Gaelic and has now learned English but who cannot understand Danish is locked in a room and has the task of using a Gaelic rule book (and has also started constructing an English rule book) for manipulating Danish words. Each Danish word has an icon or picture sequence attached to it but also some of the words will be equivalent to Gaelic or English ones. Then, to an outside observer, Seán the Leprechaun appears to be able to understand Danish just as a computer program which manipulates symbols could appear to do so. However, this time Seán begins to understand the words because he/she has reference to their meaning from two other languages as well as pictorial icons. Also, if Seán learns German then he will be able to learn Danish even better since 'retning' is 'Richtung' and 'forskning' is 'Forschung' in German!

One has to be careful with using Irish Rooms in Danish Rooms, or the like, because it is not always the case that a word in one language will mean the same in another (called faux amis or false friends in French and 'lumske ord' in Danish).[2] Also, for example, in French 'glace' means 'ice' or 'icecream' and not 'glass' and 'land' in Danish (and German) means country and not the English 'land'. There is a joke that the Danish man who went to Norway said he had beer with his 'frokost' (lunch in Danish) and the Norwegian said that he knew Danes drank a lot but not that much! (frokost in Norwegian is breakfast!). Also, 'rolig' in Danish means quiet/calm and in Swedish means amusing/funny. 'Jeg' in Danish means 'I' but in Dutch means 'you' (by sound) and 'taske' in Danish means bag/case but in Swedish means 'hooker' (by sound). Even worse, 'ham' in Danish means 'him' in English and not the English 'ham'. 'og' in Danish translates to 'and' in English but means 'little' in Gaelic! 'fart' means velocity in Danish but means 'a bad smell' in English! In fact about 25 years ago when the Queen of England came to Copenhagen she visited the Magasin Department store and the store covered up the sign 'I FART' (translates to 'in velocity' or 'moving') on the elevator![3]

Hence, our argument here is that primitives can be grounded in computer systems by achieving integration of perceptual sources of especially linguistic and visual information. Next, we move to the circularity problem.

3 Breaking the circle

When you look up the word 'gorse' in some dictionaries you find that the definition of that word involves 'furze' and when you look up 'furze' its definition uses 'gorse'? In LDOCE, the primitive for 'disease' is defined to be 'disorder' or 'illness' and these in turn are defined as 'disease'. This has been a problem for dictionaries for years.

The solution is to define a word by using a definition that uses other words but also spatial and visual structures. These structures would give partial definitions of words so that there would only be at most partial circularity in definitions. The result is partial circularity or no circularity at all.

To sum up, spatial and pictorial or animated picture sequences can be used to ground symbolic natural language specifications of the meanings of words and reduce circularity in their definitions. Our solution has not happened by accident. It has been argued for years (see Barnden 1990, Lakoff 1986) that much of language use, such as metaphor, is based on spatial relationships and mental analogical mappings.

4 Learning words

Another problem that dictionaries have had for years has been how to acquire the knowledge in them. Techniques have been developed for automatically gleaning information from Machine Readable Dictionaries (MRDs) to create Machine Tractable Dictionaries (MTDs) for NLP (see Guthrie et al. 1991) and also this has been done by hand (see Guo 1996). A discussion of how new words can arise from an existing dictionary pool is given in Rowe and Mc Kevitt (1991). We believe that similar techniques will be needed for gleaning spatial and visual information automatically for integrated lexicons which encode both symbolic and spatial information. Just like people learn the word for a dog by looking at, hearing and even touching lots of prototype dogs we believe that eventually computers will need to be able to do this if they are to be considered intelligent. At least, computers should have visual as well as symbolic representations for dogs but should also be able to learn what zebras are by putting pictures of horses together with stripes to get them. Such pictorial information is missing from today's dictionaries. The ability to develop and learn new words such as metaphors is to a large extent based on spatial and pictorial mappings. Our systems of the future will need to be able to apply algorithms for such mappings to existing dictionaries to derive new ones. And, of course, Wittgenstein (1963, p. 42) pointed out already that "It is only in the normal cases that the use of a word is clearly prescribed."

5 Intentions

A theory of intention analysis (see Mc Kevitt 1991b) has been proposed as a model, in part, of the coherence of natural-language dialogue. A central principle

[2] An American colleague of mine gets upset about the fact that people here have translated the Danish "Institut for Electroniske Systemer" directly into the English "Institute of Electronic Systems" but then that makes it look like there is only one of these in Denmark or that it is special in some way! He would rather it was "Institute for Electronic Systems".

[3] Personal Communication, Fri. Nov. 29th, 1996 from Johnna E. Nonboe who was working in Magasin for the summer at that time.

of the theory is that coherence of natural-language dialogue can be modelled by analysing sequences of intention. The theory has been incorporated within a computational model in the form of a computer program called the Operating System CONsultant (OSCON) (see Guthrie et al. 1989 and Mc Kevitt 1991). OSCON, which is written in Quintus Prolog, understands, and answers in English, English queries about computer operating systems.

The computational model has the ability to analyse sequences of intention. The analysis of intention has at least two properties: (1) that it is possible to recognise intention, and (2) that it is possible to represent intention. The syntax, semantics and pragmatics of natural-language utterances can be used for intention recognition. Intention sequences in natural-language dialogue can be represented by what we call intention graphs. Intention graphs represent frequencies of occurrence of intention pairs in a given natural-language dialogue. An ordering of intentions based on satisfaction exists, and when used in conjunction with intention sequences, indicates the local (footnote: By local expertise we wish to stress the fact that sometimes experts can act as novices on areas of a domain which they do not know well.) and global degree of expertise of a speaker in a dialogue.

It has been pointed out recently by Schank and Fano (1995) that in order to perform tasks in the world understanding is a question of relating visual and linguistic input to the intentions (goals, plans and beliefs) derived from the task. They point out that expectations are a large part of understanding and say "We need to be able to reason about the things we can sense and the situations in which we will be able to detect them. Reminders to ourselves such as strings around fingers, notes on doors, alarm clocks, etc. all betray an underlying model of what we will notice (e.g. strings, sounds, notes) as well as the situations in which we will notice them (e.g. we are all likely to see our finger, pass through a door before leaving, and hear an alarm clock next to our beds."

We agree with Schank and Fano and believe that our own work in intention modelling can only be fulfilled by incorporating the analysis of visual scenes as well as symbolic natural language. In particular, our beliefs about people before they say anything at all are based on body language, clothes, looks, makeup, style and so on and work on modelling beliefs in language (see Ballim and Wilks 1991) will need to be augmented and integrated with work on determining beliefs from visual input. Indeed, work has already begun on determining intentions from vision and language (see Gapp and MaaB 1994, Herzog and Wazinski 1994 and MaaB 1994). Next, we move on to discuss an initiative which is focussing on the implementation of these ideas and others.

6 IntelliMedia 2000+

Aalborg University, Denmark has initiated a Multimodal and Multi-media User Interfaces (MMUI) initiative to focus on the area of Intelligent MultiMedia (IntelliMedia). The initiative involves the implementation of educational MMUI, the production of a number of real-time MMUI demonstrators, and the establishment of a strong technology-based group of MMUI experts.

Aalborg's strengths in traditional MultiMedia (Humanities and Computer Science) and in IntelliMedia (Electronic Systems) establish Denmark's position in international MultiMedia development.

The Institute of Electronic Systems (IES) has four research groups which focus on computer science (CS), medical informatics (MI), image/vision (LIA) and spoken language processing (CPK). The first two groups focus on providing a strong basis with demonstrator projects in medicine/distributed systems and hypermedia tools from computer science while the latter two focus on the two main components of IntelliMedia.

6.1 Computer Science (CS)

The research at the CS includes computer systems and the design/implementation of object-oriented programming languages and environments. The scientific approach covers the formally logical, the experimentally constructive, as well as the empirically descriptive.

Of particular interest for MultiMedia are the following subjects: principles for hypermedia construction, theories of synchronisation and cognition, distributed systems and networking, high volume databases, and the design and use of language mechanisms based on conceptual modelling. Furthermore, CS has a strong research tradition within the interplay between humans, organisations and information systems, and also within the subject of decision support systems and communicating agents, which is highly relevant for emerging research on models for user/system interaction.

CS contributions include experiments for performance evaluation of the available technology (e.g. high speed networking) and experiments on the methodology for design of MultiMedia systems. These contributions will be based on existing research activities, which includes networks, distributed models (Topsy), and prototype hypermedia environments.

6.2 Medical Informatics (MI)

The research in the Medical Decision Support System group is centered around medical knowledge-based systems and the development of general tools to support complex decision making.

The knowledge-based system technology based on Bayesian networks allowing for a proper handling of uncertain information has shown itself to be usable in creating intelligent coupling between interface components and the underlying knowledge structure. This technology may be integrated in IntelliMedia systems. The Bayes network paradigm, as developed in Aalborg, is already in practical use in user interfaces such as in Intelligence, a user and environment context sensitive help system in the major word processing and spreadsheet products from Microsoft.

It is foreseen that IntelliMedia systems will play a central role in the dissemination of information technology in the medical informatics sector. Systems representing complex knowledge, models and data structures e.g. advanced medical diagnostics system, virtual operation room, the telemedical praxis and so on, will require use of knowledge-based techniques for efficient interfacing.

6.3 Laboratory of Image Analysis (LIA)

The research at LIA is directed towards three areas: Systems for computer vision, computer vision for autonomous robots, and medical and industrial application of image analysis.

Research projects concern extraction of features for description of actions in an environment (i.e. the movement of people, fish, and blood cells) and utilising these descriptions for recognition, monitoring and control of actuators such as mobile robots (safe movements in a dynamically changing environment). This includes recognising and tracking dynamically changing objects, such as hands and human bodies, which has applications in IntelliMedia Systems.

So far the research has referred to sensory processing using single modalities, but it seems obvious that the available methods may be integrated into multimodal system, where a major objective is coordination and optimal use of available modalities. New IntelliMedia systems may also include much more flexible modes of interaction between computers, including both speech, body movements, gestures, facial expressions and sign language.

6.4 Center for PersonKommunikation (CPK)

Research at the CPK is focused within the following three areas: Spoken Language Dialogue Systems, Data Communications and Radio Communications. CPK is an engineering research center funded by the Danish Technical Research Council.

The research within Spoken Language Dialogue Systems has for a long time been focused on human-computer interfacing and interaction and to a large extent been developed in connection with ESPRIT and nationally funded projects. The results obtained so far are of high relevance to many foreseen practical MultiMedia applications and to Framework IV of the European Union (EU), and they may advantageously be utilised as partial basis for all activities of the MMUI initiative.

CPK has an already developed a Dialogue Specification, Design and Management tool called Generic Dialogue System (GDS) (see Baekgaard 1996 and Dalsgaard and Baekgaard 1994) which is an appropriate platform for IntelliMedia research, and which from the very beginning may be used in various specialisations and student projects.

6.5 Demonstrator CHAMELEON

The research demonstrator platform is an IntelliMedia Workbench which gives advice on the layout plans of buildings. It has a general architecture of communicating agent modules processing inputs and outputs from different modalities and each of which can be tailored to a number of application domains. The demonstrator is being developed in both a top-down and bottom-up manner making sure it is general enough for multiple application domains but at the same time keeping particular domains in mind. The demonstrator will be a single platform (called CHAMELEON) demonstrating that existing platforms for (1) image processing, (2) spoken dialogue processing, (3) expert systems and (4) microphone arrays can be interfaced to the single hub platform and act as communicating agent modules within it. The platform will be independent of any particular application domain and the intention is that it will be possible to run it over different server machines. We are using the programming languages Java, C++, and C. Our Generic Dialogue System (GDS) implemented in C and C++ is being considered as an initial platform tool.

The hub platform will demonstrate that (1) it is possible for agent modules to receive inputs particularly in the form of images, spoken language, expert system queries and present outputs particularly in the form of speech, images, and graphics; (2) individual agent modules within the platform can produce output in the form of semantic representations to show their internal workings; (3) the semantic representations can be used for effective communication of information between different agents for various applications; and (4) various means of synchronising the communication between agents can be tested to produce optimal results.

A reading group has been initiated under IntelliMedia 2000+ and this group meets frequently to have lively discussions on state-of-the-art papers in the area.

6.6 Teaching

Teaching is a large part of IntelliMedia 2000+ and three courses have been initiated: (1) Graphical User Interfaces, (2) Intelligent Multimedia Systems and (3) Readings in Advanced Intelligent MultiMedia. Graphical User Interfaces is a more traditional course involving teaching of methods for the development of optimal interfaces for Human Computer Interaction (HCI). The course brings students through methods for layout of buttons, menus, and form filling methods for interface screens and has hands on experience with the XV windows development tool.

IntelliMedia Systems involves the new and innovative topics of speech, language and vision processing. Here, minimodules are given on methods for recognising and interpreting spoken language in dialogue situations and speech and audio representation. The Dialogue Description Language (DDL) tool and Generic Dialogue System of CPK are explained and demonstrated. Vision minimodules are given on relationships between audio analysis and image analysis, 2D model based recognition of static gestures (hand signals), 3D model based tracking of human motion (limbs), and recognizing/'understanding' human motion patterns. There are minimodules on Natural Language Processing (NLP) and pragmatics. The course is augmented with videos and live demonstrations. Hence, this course is true IntelliMedia involving speech, vision and language processing. A guest lecture can be given as part of this module.

The course on Readings in Advanced Intelligent MultiMedia is innovative and new and involves active learning where student groups present research papers and then the whole class can have a general discussion of them. The presentations would include four aspects: (1) who the group is and what their project is, (2) a summary and critical analysis of the papers, (3) how the papers relate to their project and (4) how do the papers and their project relate to IntelliMedia 2000+. Then, the whole class discusses the readings and group presentation. I have asked for one group member to write up the minutes of our discussions for posterity. Everything is to be done in English. The idea here is that it will not only develop the students' presentation skills but also their ability to assimilate, analyse critically and use recent research in the field. I have chosen the papers from a selection of books which have just been published on the latest research.

A new international Master's Degree (M.Sc.) has been established and incorporates the courses just mentioned as core modules of a 1 and 1/2 year degree taught in English on IntelliMedia. More details can be found on WWW: http://www.kom.auc.dk/ESN/ A Lifelong Learning course is given in August for returning students of Aalborg University who wish to continue their education. This course is a compression of the core Intelligent MultiMedia courses.

The emphasis on group oriented and project oriented education at Aalborg University is an excellent framework in which IntelliMedia, an inherently interdisciplinary subject, can be taught. Groups can even design and implement a smaller part of a system which has been agreed upon between the groups.

To sum up, IntelliMedia 2000+ is a research and teaching initiative where we focus on the integration of at least language and vision processing.

7 Conclusion

SuperinformationhighwayS are forcing the merging of the Humanities and Sciences/Engineering in terms of representing and accessing information in multiple modalities including at least text, voice, sounds and images/videos (Intelligent Multimedia). Language will be input in multiple formats including multiple natural languages as well as formal languages and images in the form of simple diagrams right up to videos. The Humanities will be concerned more with the content of the information being passed while the Sciences/Engineering will be more concerned with representation and transmission.

We believe that integrated lexicons incorporating language and visual information are needed for language and vision processing where such lexicons provide extra structures which describe objects and actions rather than just having the flat symbolic representation which we have been used to. We believe that these extra structures will cause many of the problems of symbol grounding and semantic primitives to disappear and the lexicons will become grounded in many forms of perceptual input. Searle's Chinese Room Problem will go away as machines will have more of a feel for the meanings of the words they know in the form of an Irish Room.

We believe that the analysis of intentions is not only important for interpreting the actions of agents

in visual environments but also for determining what agents mean when they use words. That is, words have meanings which people intend them to have. Cognitive Science (CS) and Computer Science (CS) are converging on Information (I), Intentions (I) and Integration (I) and we propose the following formula for future development of theories of mind:
CS = I x I x I = I^3

As Horgan (1996) points out much of the future of science for 2000 and beyond will be in the integration of existing theories, models and systems.

Lexicons of the future will have in conjunction with flat semantic representations for word senses, spatial representations, pictures and sounds and these will all be used in computer systems for AI and for MultiMedia Interfaces. Metaphors and new uses will easily be derived from analogical mappings between spatial relations and pictures. We believe the computer will change the whole meaning of lexicon and lexicons of the future will not be constrained by the symbolic language descriptions of the past. IntelliMedia 2000+ is a new research and teaching programme at Aalborg University, Denmark which is implementing these ideas and others in the form of Intelligent MultiMedia. Such are our words on visions for the future of SuperinformationhighwayS.

References

Baekgaard, Anders (1996) Dialogue Management in a Generic Dialogue System. In *Proceedings of the Eleventh Twente Workshop on Language Technology (TWLT) Dialogue Management in Natural Language Systems,123-132*, Twente, The Netherlands.

Ballim, Afzal and Yorick Wilks (1991) *Artificial Believers.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Barnden, John A. (1990) *Naive metaphysics: a metaphor-based approach to propositional attitude representation (unabridged version).* Memorandum in Computer and Cognitive Science, MCCS-90-174, Computing Research Laboratory, Dept. 3CRL, Box 30001, New Mexico State University, Las Cruces, NM 88003-0001, USA.

Dalsgaard, Paul and A. Baekgaard (1994) Spoken Language Dialogue Systems. In *Prospects and Perspectives in Speech Technology: Proceedings in Artificial Intelligence*, Chr. Freksa (Ed.), 178-191, September. München, Germany: Infix.

Denis, M. and M. Carfantan (Eds.) (1993) *Images et langages: multimodalité et modelisation cognitive.* Actes du Colloque Interdisciplinaire du Comité National de la Recherche Scientifique, Salle des Conférences, Siège du CNRS, Paris, April.

Dennett, Daniel (1991) *Consciousness explained.* Harmondsworth: Penguin.

Gapp, Klaus-Peter and Wolfgang Maaβ (1994) Spatial layout identification and incremental descriptions. In *Proceedings of the Workshop on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.), 145-152, Twelfth American National Conference on Artificial Intelligence (AAAI-94), Seattle, Washington, USA, August.

Guo, Chengming (1995) *Machine tractable dictionaries.* Norwood, N.J.: Ablex.

Guthrie, Louise, Paul Mc Kevitt and Yorick Wilks (1989) OSCON: An operating system consultant. In *Proceedings of the Fourth Annual Rocky Mountain Conference on Artificial Intelligence (RMCAI-89), Subtitled, 'Augmenting Human Intellect By Computer'*, 103-113, Registry Hotel, Denver, Colorado, USA, June.

Harnad S. (1990) The symbol grounding problem. In *Physica D.*, 335-46.

Herzog, Gerd and Peter Wazinski (1994) VIsual TRAnslator: Linking perceptions and natural language descriptions. In *Artificial Intelligence Review, Special Volume on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.). Vol. 8, Nos. 2-3, 175-187.

Horgan, John (1996) *The end of science: facing the limits of knowledge in the twilight of the scientific age.* Reading, Mass.: Addison-Wesley (Helix Books).

Joyce, James (1922) *Ulysses.* London: Faber and Faber.

Joyce, James (1939) *Finnegans Wake.* London: Faber and Faber.

Lakoff, G. (1986) *Women, fire and dangerous things.* Chicago, Illinois: University of Chicago Press.

Marconi, Diego (1995) On the referential competence of some machines. In *Artificial Intelligence Review, Special Volume on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.). Vol. 10 (this volume).

Maaβ, Wolfgang (1994) From vision to multimodal communication: incremental route descriptions. In *Artificial Intelligence Review, Special Volume on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.). Vol. 8, Nos. 2-3, 159-174.

Mc Kevitt, Paul (1991a) Principles and practice in an operating system consultant. In *Artificial Intelligence and Software Engineering, Vol. 1 Chapter on 'AI Mechanisms and techniques in practical software'*, Derek Partridge (Ed.). New York: Ablex Publishing Corporation.

Mc Kevitt, P. (Ed.) (1994a) Proceedings of the Workshop on Integration of Natural Language and Vision processing. In *Twelfth American National Conference on Artificial Intelligence (AAAI-94)*, Seattle, Washington, USA, August.

Mc Kevitt, Paul (Ed.) (1995/96) *Integration of Natural Language and Vision Processing (Vols. I-IV).* Dordrecht, The Netherlands: Kluwer-Academic Publishers.

Mc Kevitt, Paul and Chengming Guo (1996) From Chinese Rooms to Irish Rooms: new words on visions for language. In *Artificial Intelligence Review Journal*, Vol. 10(1-2), 49-63 and In Paul Mc Kevitt, Paul (Ed.), Integration of Natural Language and Vision Processing: theory and grounding representations (Vol. III), 151-165. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Meini, Cristina and Alfredo Paternoster (1995) Understanding language through vision. In *Artificial Intelligence Review, Special Volume on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.). Vol. 10 (this volume).

Olney, J. (1968) *To: All interested in the Merriam-Webster transcripts and data derived from them.* Technical Report L-13579 Santa Monica, CA: System Development Corporation.

Pentland, Alex (Ed.) (1993) *Looking at people: recognition and interpretation of human action.* IJCAI-93 Workshop (W28) at The 13th International Conference on Artificial Intelligence (IJCAI-93), Chambéry, France, EU, August.

Procter, P. (1978) *Longman Dictionary of Contemporary English.* London: Longman.

Rowe, Jon and Paul Mc Kevitt (1991) An emergent computation approach to natural language processing. In *Proceedings of the Fourth Irish Conference on Artificial Intelligence and Cognitive Science*, University College Cork, IRL- Cork, Ireland, European Union (EU), September.

Schank, Roger C. (1972) Conceptual dependency: a theory of natural language understanding. In *Cognitive Psychology*, 3(4): 552-631.

Schank, Roger and Andrew Fano (1995) Memory and expectations in learning, language and visual understanding. In *Artificial Intelligence Review, Special Volume on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.). Vol. 9, Nos. 4-5.

Searle, J.R. (1980) Minds, brains and programs. In *Behaviour and Brain Sciences*, 3: 417-424.

Searle, J.R. (1984) *Minds, brains and science.* London: Penguin Books.

Searle, J.R. (1990) Is the brain's mind a computer program?. In *Scientific American*, 262: 26-31.

Sinclair, John (Ed.) (1987) *Looking Up: an account of the COBUILD project in lexical computing.* London: Collins.

Wilks, Yorick (1973) An artificial intelligence approach to machine translation. In *Computer Models of Thought and Language*, R. Schank and K. Kolby (Eds.). San Francisco, CA: Wh Freeman and Co..

Wilks, Yorick (1977) Good and bad arguments about semantic primitives. In *Communication and cognition*, Vol. 10, No. 3/4, 181-221.

Wilks, Yorick (l978) Semantic Primitives in language and vision. In *Proceedings of the Second Conference on Theoretical Issues in Natural Language Processing*, Champaign-Urbana, IL.

Wilks, Yorick (1995) Language, vision and metaphor. In *Artificial Intelligence Review, Special Volume on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.). Vol. 9, Nos. 4-5.

Wittgenstein, Ludwig (1963) *Philosophical Investigations (translated by G.E. Anscombe).* Oxford: Blackwell.