

From Chinese Rooms to Irish Rooms: new words on visions for language^{*†}

Paul Mc Kevitt
Department of Computer Science
Regent Court, 211 Portobello Street
University of Sheffield
GB- S1 4DP, Sheffield
England, EU.
E-mail: p.mckevitt@dcs.shef.ac.uk

Chengming Guo
Computer Science Department
Tsinghua University
Beijing 100084, China.
Email: cmguo@sun.ihep.ac.cn

Abstract

Natural languages like English have been constrained in expressing perceptions like vision, sound and touch for years despite the efforts of Joyce (1922, 1939) and others. In situ, lexicons have been limited in their form and content. They have typically been structured in the form of sequences of natural language words with their content defined using flat symbolic descriptions in natural languages. In particular, we believe that today's dictionaries in general, and with respect to Artificial Intelligence (AI) systems in particular, are unnatural in the sense that they do not encode pictures for words just like we do in our heads. There is now a move towards integrated systems in Artificial Intelligence (AI) (see Dennett 1991, Mc Kevitt

*Paul Mc Kevitt is currently funded for five years on a British Engineering and Physical Sciences Research Council (EPSRC) Advanced Fellowship under grant B/94/AF/1833 for the Integration of Natural Language, Speech and Vision Processing.

†An abridged version of this paper has appeared in the Proceedings of the Post-COLING94 International Workshop on Directions of Lexical Research, Beijing, China, August, 1994 and will appear in book version at a later date.

1994a, 1994b) and that will cause a need for dire actions on lexical research in the form of integrated lexicons. We believe that lexicons must move towards a situation where natural language words are also defined in terms of spatial and visual structures. These spatial and visual structures will solve what have been two of the most prominent problems in the field of Natural Language Processing (NLP) for years: (1) Where are symbolic semantic primitive meanings in computer programs grounded? and (2) How come some words, typically in the defining vocabulary, in dictionaries have circular definitions so that words end up defining each other? We believe integrated lexicons will cause these two problems to go away and hence help solve Searle's Chinese Room Problem and move more towards Irish Rooms of people like James Joyce.

1 Introduction

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 others. 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, 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.

Although there has been much progress in developing theories, models and systems in the areas of Natural Language Processing (NLP) and Vision Processing (VP) there has been little progress on integrating these two subareas of Artificial Intelligence (AI). Although in the beginning the general aim of the field was to build integrated language and vision systems, few were done, and there quickly became two subfields. It is not clear why there has not already been much activity in integrating NLP and VP. Is it because of the long-time reductionist trend in science up until the recent emphasis on chaos theory, non-linear systems, and emergent behaviour? Or, is it because the people who have tended to work on NLP tend to be in other Departments, or of a different ilk, to those who have worked on VP? There has been a recent trend towards the integration of NLP and VP (see Denis and Carfantan 1993, Dennett 1991; Mc Kevitt 1992, 1994a, Pentland 1993, and Wilks and Okada (in press)).

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."

There has been fallout from the field of AI having become split into subfields. Many questions of ambiguity of words, sentences and paragraphs in NLP go away if one has some other perceptual source such as sounds or visual input feeding in to solve the ambiguity problem. People in NLP talked of solving ambiguities solely in terms of symbolic natural language frameworks and these debates sometimes wound round in circles as other perceptual inputs were forgotten. The dictionaries and knowledge bases for NLP suffered from

two limitations: (1) the split from vision processing, and (2) our history of having symbolic natural language dictionaries. In vision processing many systems were built which attempted to uncover objects solely on the basis of visual recognition whereas natural language and other perceptual inputs could have helped to solve many of these problems.

There have been two problems that have worried us in the field of NLP for years: (1) Where are symbolic semantic primitive meanings in NLP programs grounded? and (2) 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. So, it would seem that we have caused problems for ourselves by our reductionist tendencies. We show here how the problems can be solved in part by resorting to integrated representations for at least language and vision.

2 Integrated lexicons

There has been much work in AI and in NLP on defining dictionaries for use in large intelligent systems. Examples are the Machine Tractable Dictionaries (MTDs) produced from Machine Readable Dictionaries (MRDs) such as LDOCE (see Guo 1995, Guthrie et al. 1991) and COBUILD (see Sinclair 1987). In fact, the idea of processing MRDs to obtain lexicons for NLP systems has become one of the largest research areas in NLP. There has also been work on encoding large encyclopedias for AI systems (see Lenat and Guha 1989). However, we argue that such dictionaries and encyclopedias are limited as they have carried over bad habits from existing MRDs in that they only contain symbolic natural language definitions for words and do not contain the necessary knowledge for solving intelligent problems. We call here for spatial and visual representations to be added to the lexicons and knowledge bases for AI systems.

Partridge (1995) points out that language and vision have a single perceptual mechanism where he shows that independent work on a cognitive model of visual perception and of perception of lexical items reveals a common framework underlying the two sets of cognitive mechanisms. Marconi (1995) says that NLP systems have not been able to *understand* because they have no vision component in which symbols representing natural language concepts are grounded. Wilks (1995) discusses the relationship between language, vision and metaphor and argues that visual processing can embody structural ambiguity but not anything analogous to metaphor. He says that metaphor is connected with the extension of sense and only symbols can have senses.

In the recent moves towards constructing integrated AI systems integrated lexicons have been developed. Many of these use spatial representations for words. An interesting venture has been that of developing animations for describing primitive concepts in natural language. Narayanan et al. (1994) discuss the possibility of developing dynamic visual primitives for language primitives. Language primitives which themselves represent dynamic processes such as entering a building can be dynamically visualised. They map the Conceptual Dependency primitives of Schank (1972) such as PTRANS (which is the

physical act of transferring an object from one location to another) into pictures and show how simple stories can be mapped into picture sequences. Beardon (1995) has developed a visual language for iconic communication which he argues will be easier to use than language in a computer environment. He has developed a prototype system called CD-Icon which is also based on Schank's Conceptual Dependency representation.

Another area which has seen the rise of integrated lexicons is that of document processing. Rajagopalan (1994) defines a picture semantics for specifying the connection between patterns, colours, and brush types used in diagrams and a domain such as urban scenes. Srihari (1994) and Srihari and Burchans (1994) define a lexicon with spatial semantics for the problem of document understanding where textual captions are used as collateral information in the interpretation of corresponding photographs. Their system called PICTION uses a syntactic/semantic lexical database which is constructed from LDOCE, OALD (Oxford Advanced Learner's Dictionary), WordNet (see Beckwith et al. 1991) and manual augmentation with ontological information and visual semantics. The entire lexicon is represented as a LOOM¹ knowledge base. The lexicon captures fixed, visual properties of objects, such as size and colour as well as procedural information such as the scalar and shape properties of the object *hat*, and location *above head*, for sentences like "the person wearing the hat". Nakatani and Itoh (1994) have developed semantic representations with primitives for colours and words that qualify colours and use this for an image retrieval system that accepts natural language.

Many of the integrated representations for language and vision are spatial in nature. Olivier and Tsujii (1994) describe a quantitative object model and qualitative spatial and perceptual model for prepositions such as *in front of*. The meanings of prepositions in his system called WIP are automatically tuned in accordance with reference to objects. Gapp and Maa β (1994) describe spatial models for route descriptions given through verbal utterances. Reyero-Sans and Tsujii (1994) describe a visual interlingua of spatial descriptions which is used to represent linguistic meaning and is motivated by the problem of machine translation for English and Spanish. Chakravarthy (1994) describes a perceptually-based semantics of action verbs such as *hit* and *walk* where lexical entries for verbs can be built up from perceptual predicates by inheritance from existing more primitive lexical entries. The perceptual semantics contains state descriptions for verbs and would give, for example, for *cut*, the object, instrument, trajectory and duration of cutting. Dolan (1994) describes a framework for vision processing which makes use of a large lexical database which has been automatically derived from MRDs. He suggests that MRDs contain much of the information needed about the physical and common-sense properties of objects in vision processing. Meini and Paternoster (1995) discuss a lexical structure where each lexical entry has two pointers, one to a knowledge base and the other to a catalogue of shapes containing structural descriptions of objects denoted by the lexical items.

There are also full blown multi-modal lexicons under development. Young (1983) and Wright and Young (1990) describe a knowledge representation called Cognitive Modalities

¹LOOM (see ISX 1991) is a high level object-oriented programming language and environment for constructing knowledge based systems.

(CM) for neural networks which is a cognitive, non-verbal representation of information. The CM system uses a cognitive lexicon of some 8,000 modality specific elements which are grouped according to the sub-modal categories of perception and non-sensorimotor events. It is hoped the representation will be used for machine translation and information retrieval. Modalities such as touch and time are encoded in CM. Phonological information is currently being encoded into CM and of course CM includes the traditional pragmatic, semantic and syntactic information encoded in traditional lexicons.

Hence, we can see that there are now moves afoot to developing integrated lexicons and a quick review of them shows that central themes are (1) animation and iconization of primitives, (2) spatial relations and (3) multi-modal forms.

3 Word problems

One of the suggested solutions to problems of NLP over the years has been to reduce word and sentence representations to primitives (see Wilks 1977, 1978). Schank defined 14 of such primitives for Conceptual Dependency (see Schank 1972, 1973, 1975) and Wilks had some 80 in his Preference Semantics system (see Wilks 1973, 1975a, 1975b, 1975c, 1977). However, all this reductionist work had difficulties because of at least three reasons: (1) general inference rules applied to primitives don't follow through when applied to specific situations or contexts, (2) *grounding*: how are the primitives grounded in the world? i.e. what gives them their meaning?, (3) *circularity*: some words are defined in terms of primitives but those primitives are defined in terms of the original words. We shall focus on the second and third problems here. First, the grounding problem.

3.1 Primitive grounding

The problem with dictionaries such as LDOCE and MTDs 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 systems like Schanks' and Wilks' where are the primitives grounded? Harnad (1990, 1993) 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 Room² problem 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). In some sense what Searle

²Searle asked us to imagine a Chinese Room where a person 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).

is arguing is that the computer behaves like a hypertext system does, encoding text and being able to manipulate and move it around but having no sense of its meaning. Jackson and Sharkey (1995) argue that connectionist architectures are necessary for grounding perceptual systems or at least that such grounding will be easier with such architectures. Wilks (1995) points out that primitives in natural language do not have any obvious visual analogues and that no definition of primitives is necessary because they are explained by the procedural role they play in language as a whole. Katz (1972) claims that linguistic primitives play the role that neutrinos play in science.

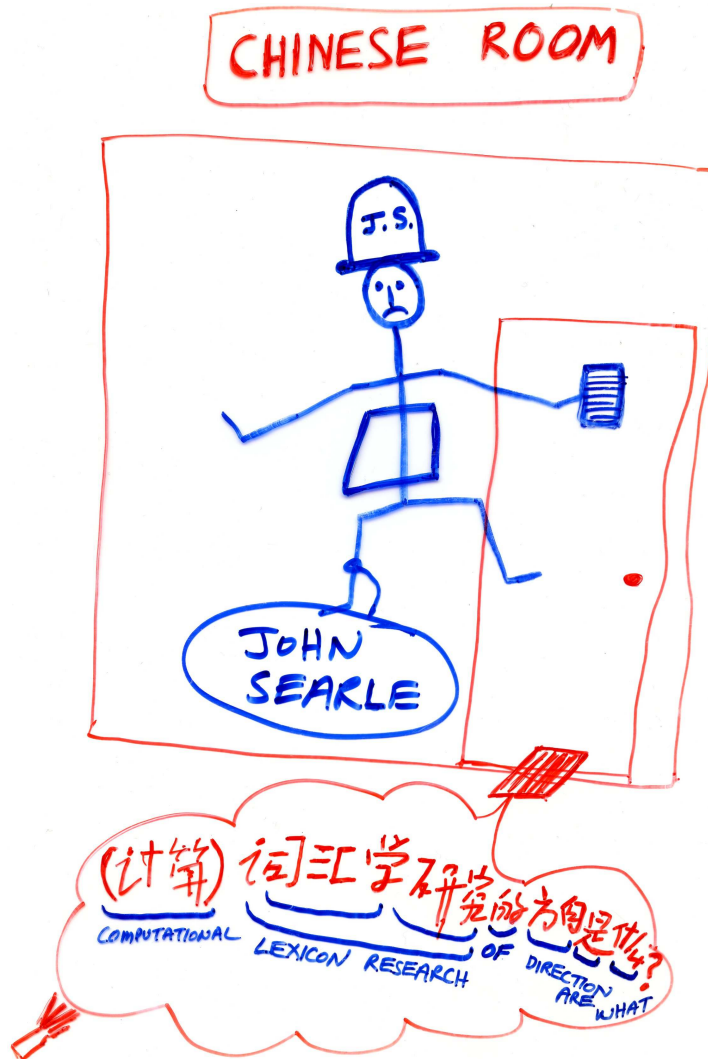


Figure 1: The Chinese Room

We argue here that defining vocabularies and primitives can be defined in terms of spatial or pictorial representations to obtain meanings. So, for example, taking a primitive for the concept of abstract transfer (called ATRANS by Schank) we can have a picture showing two agents with an object being transferred between the two. This picture could

be animated as demonstrated by Beardon (1995) and Narayanan et al. (1994) and could be shown to a user on demand. Furthermore, a definition of the changes in spatial relations with respect to ATRANS could be represented. For example, this would detail the object, instrument, trajectory and duration of transfer as defined in the perceptual semantics of Chakravarthy (1994). Hence, now we have an Irish Room, going even further than Joyce where he tried to bring perception into written symbols. The Irish Room is shown in Figure 2 where the sentence “What are the directions of lexical research?” in English is being passed to Séan 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 cannot understand English is locked in a room and has the 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, Séan 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 Séan begins to understand the words because he/she has reference to their meaning. Sounds, smells and even touch can be added in later! This solution to the Chinese Room problem has also been suggested by Harnad (1990, 1993), and discussed at length by Marconi (1995) and Meini and Paternoster (1995).

3.2 The Irish Room in a Chinese Room

In order to test whether the Irish Room idea works one of us (Paul Mc Kevitt) tried an experiment at Tsinghua University in Beijing, China on August 20th, 1994 at 9.00-10.00 PM. 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. The examples are shown in Figure 3 (note that I didn’t get the English translation which we have placed there for the reader). First I was given the sentence “You are very strong” in Chinese annotated with icons. I guessed the sentence meant “Some person SOMETHING big” which isn’t too bad! Then, it became interesting because I was given another sentence: “A cat is strong” in Chinese with one more icon (a cat) and recognised the ending characters were the same as those used at the end of the previous example. I guessed it right first time as the icon for cat was obvious! We only worked here with simple icons and the system was working quite well. We predict that with more fine tuned icons, videos, sounds and other sensory information the results would be much better. Next, the circularity problem.

3.3 Breaking the circle

How come in some dictionaries you look up a word like ‘gorse’ and 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.

Again we propose a solution here where one defines a word by using a definition that uses other words but also spatial and visual structures. These structures would give partial



Figure 2: The Irish Room

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 we have argued that 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.

AN EXPERIMENT

TSINGHUA UNIVERSITY
BEIJING / AUGUST 20th 1994
9.00-10.00 PM.

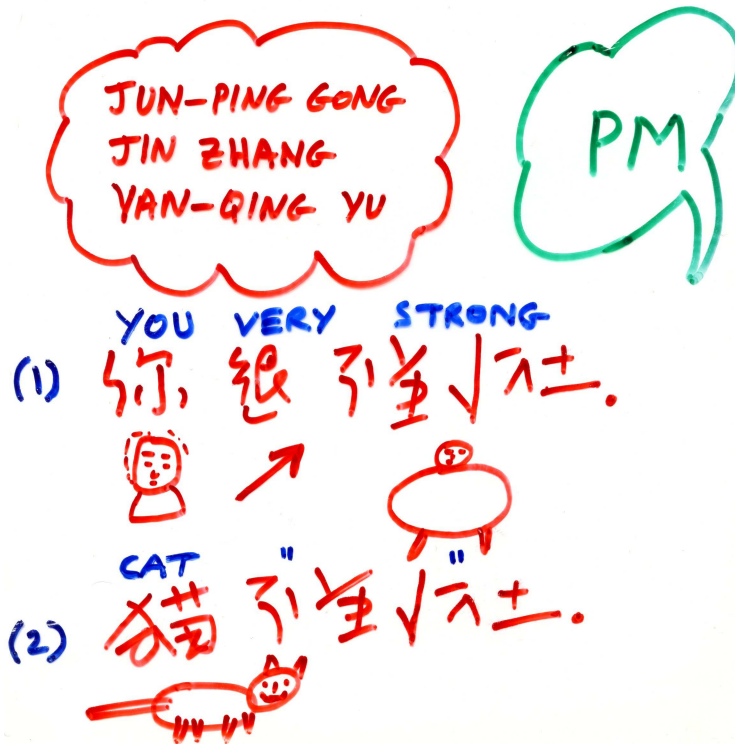


Figure 3: The Irish Room in China

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 MRDs for NLP (see Guthrie et al. 1991) and also this has been done by hand (see Guo 1995). 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 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, Mc Kevitt 1986, 1991a, 1991b, Mc Kevitt and Wilks 1987, and Mc Kevitt et al. 1992a, 1992b, 1992c, 1992d). 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*³ and *global* degree of expertise of a speaker in a dialogue.

The architecture of the OSCON system consists of six basic modules and two extension modules. There are at least two arguments for modularising any system: (1) it is much easier to update the system at any point, and (2) it is easier to map the system over to another domain. The six basic modules in OSCON are as follows: (1) ParseCon: natural-language syntactic grammar parser which detects query-type, (2) MeanCon: a natural-language semantic grammar (see Brown et al. 1975, and Burton 1976) which determines query meaning, (3) KnowCon: a knowledge representation, containing information on natural-language verbs, for understanding, (4) DataCon: a knowledge representation for containing information about operating system commands, (5) SolveCon: a solver for resolving query representations against knowledge base representations, and (6) GenCon: a natural-language generator for generating answers in English. These six modules are satisfactory if user queries are treated independently, or in a context-free manner. However, the following two extension modules are necessary for dialogue-modelling and user-modelling:

³By *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.

(1) DialCon: a dialogue modelling component which uses an intention matrix to track intention sequences in a dialogue, and (2) UCon: a user-modeller which computes levels of user-satisfaction from the intention matrix and provides information for both context-sensitive and user-sensitive natural-language generation.

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 1990, 1991 and Wilks and Ballim 1987) 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 Maaß 1994, Herzog and Wazinski 1994 and Maaß 1994).

6 Conclusion

We conclude here by pointing out that integrated lexicons 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 involve spatial and pictorial animated representations of objects and actions and that these representations will enable systems to conduct better processes such as analogical reasoning over lexicons. Integrated lexicons 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.

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, Intentions and Integration and we propose the following formula for future development:

$$CS = I \times I \times I = I^3$$

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. Such are our words on visions for lexicons.

7 References

- Ballim, Afzal and Yorick Wilks (1990) Stereotypical belief and dynamic agent modeling. In *Proceedings of the Second International Conference on User Modelling*, University of Hawaii at Manoa, Honolulu, Hawaii, USA.
- 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.
- Beardon, Colin (1995) Discourse structures in iconic communication. In *Artificial Intelligence Review, Special Volume on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.). Vol. 9, Nos. 2-3.
- Beckwith, R., C. Fellbaum, D. Gross, and G.A. Miller (1991) WordNet: a lexical database organized on psycholinguistic primitives. In *Lexicons: using online resources to build a lexicon*, . Hillsdale, N.J.: Lawrence Erlbaum.
- Brown, John Seely, Richard R. Burton and Alan G. Bell (1975) SOPHIE: a step towards creating a reactive learning environment. In *International Journal of Man-Machine Studies*, 7, 675-696.
- Burton, R. (1976) *Semantic grammar: An engineering technique for constructing natural-language understanding systems*. BBN Report No. 3453, Bolt, Beranek and Newman, Cambridge, MA, USA.
- Chakravarthy, Anil (1994) Towards a perceptually-based semantics of action verbs. In *Proceedings of the Workshop on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.), 161-164, Twelfth American National Conference on Artificial Intelligence (AAAI-94), Seattle, Washington, USA, August.
- Denis, M. and M. Carfantan (Eds.) (1993) *Images et langages: multimodalité et modalisation 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.
- Dolan, William B. (1994) Exploiting lexical information for visual processing. In *Proceedings of the Workshop on Integration of Natural Language and Vision Processing*,

- Mc Kevitt, Paul (Ed.), 185-188, Twelfth American National Conference on Artificial Intelligence (AAAI-94), Seattle, Washington, USA, August.
- Gapp, Klaus-Peter and Wolfgang Maaß (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.
- Guthrie, Joe, Louise Guthrie, Yorick Wilks, and H. Aidinejad (1991) Subject-Dependent Co-Occurrence and Word Sense Disambiguation. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, 146-152, Berkeley, California. Also Memoranda in Computer and Cognitive Science, MCCS-91-206, CRL, NMSU.
- Harnad S. (1990) The symbol grounding problem. In *Physica D.*, 335-46.
- Harnad S. (1993) Grounding symbols in the analog world with neural nets: a hybrid model. In *Think 2*, 12-20.
- 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.
- ISX 1991 (1991) *ISX Corporation*. LOOM Users Guide, version 4.1 edition.
- Jackson, Stuart and Noel Sharkey (1995) Grounding computational engines. In *Artificial Intelligence Review, Special Volume on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.). Vol. 10 (this volume).
- Joyce, James (1922) *Ulysses*. London: Faber and Faber.
- Joyce, James (1939) *Finnegans Wake*. London: Faber and Faber.
- Katz, J. (1972) *Semantic theory*. New York: Harper & Row.
- Lakoff, G. (1986) *Women, fire and dangerous things*. Chicago, Illinois: University of Chicago Press.
- Lenat, Doug and R.V. Guha (1989) *Building large knowledge-based systems: representation and inference in the Cyc project*. Reading, MA: Addison-Wesley.
- 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).
- Maaß, 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 (1986) *Building embedded representations of queries about UNIX*. Memorandum in Computer and Cognitive Science, M CCS-86-72, Computing Research Laboratory, Dept. 3CRL, Box 30001, New Mexico State University, Las Cruces, NM 88003-0001, USA.
- 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, Paul (1991b) *Analysing coherence of intention in natural-language dialogue*. Ph.D. Thesis, Department of Computer Science, University of Exeter, Exeter, England, UK, EU, September.
- Mc Kevitt, P. (Editorial) (1992) Natural language processing. In *Artificial Intelligence Review*, Vol. 6, No. 4, 327-332, Dec.
- 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, P. (Ed.) (1994b) Proceedings of the Workshop on Integration of Natural Language and Speech Processing. In *Twelfth American National Conference on Artificial Intelligence (AAAI-94)*, Seattle, Washington, USA, August.
- Mc Kevitt, Paul and Yorick Wilks (1987) Transfer Semantics in an Operating System Consultant: the formalization of actions involving object transfer. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence (IJCAI-87)*, Vol. 1, 569-575, Milan, Italy, EC, August.
- Mc Kevitt, Paul, Derek Partridge and Yorick Wilks (1992a) Approaches to natural language discourse processing. In *Artificial Intelligence Review*, Vol. 6, No. 4, 333-364, Dec..
- Mc Kevitt, Paul, Derek Partridge and Yorick Wilks (1992b) *Analysing coherence of intention in natural language dialogue*. Technical Report 227, Department of Computer Science, University of Exeter, GB- EX4 4PT, Exeter, England, EU.
- Mc Kevitt, Paul, Derek Partridge and Yorick Wilks (1992c) *Why machines should analyse intention in natural language dialogue*. Technical Report 233, Department of Computer Science, University of Exeter, GB- EX4 4PT, Exeter, England, EU.
- Mc Kevitt, Paul, Derek Partridge and Yorick Wilks (1992d) *Experimenting with intention in natural language dialogue*. Technical Report 234, Department of Computer Science, University of Exeter, GB- EX4 4PT, Exeter, England, EU.
- 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).
- Nakatani, Hiromasa and Yukihiro Itoh (1994) An image retrieval system that accepts natural language. In *Proceedings of Workshop on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.), 7-13, Twelfth American National Conference on Artificial Intelligence (AAAI-94), Seattle, Washington, USA, August.

- Narayanan, A., L. Ford, D. Manuel, D. Tallis and M. Yazdani (1994) Language animation. In *Proceedings of Workshop on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.), 58-65, Twelfth American National Conference on Artificial Intelligence (AAAI-94), Seattle, Washington, USA, August.
- Olivier, Patrick and Jun-ichi Tsujii (1994) Prepositional semantics in the WIP system. In *Proceedings of the Workshop on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.), 139-144, Twelfth American National Conference on Artificial Intelligence (AAAI-94), Seattle, Washington, USA, August.
- Partridge, Derek (1995) Language and vision: a single perceptual mechanism?. In *Artificial Intelligence Review, Special Volume on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.). Vol. 9, Nos. 4-5.
- 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.
- Rajagopalan, Raman (1994) Integrating text and graphical input to a knowledge base. In *Proceedings of the Workshop on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.), 14-21, Twelfth American National Conference on Artificial Intelligence (AAAI-94), Seattle, Washington, USA, August.
- Reyero-Sans, Irina and Jun-ichi Tsujii (1994) A cognitive approach to interlingua representation of spatial descriptions. In *Proceedings of Workshop on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.), 122-130, Twelfth American National Conference on Artificial Intelligence (AAAI-94), Seattle, Washington, USA, August.
- 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 C. (1973) Identification and conceptualizations underlying natural language. In *Computer Models of Thought and Language*, R. Schank and K. Kolby (Eds.). San Francisco, CA: Wh Freeman and Co..
- Schank, Roger C. (1975) *Conceptual information processing*. Fundamental Studies in Computer Science, 3. Amsterdam: North-Holland.
- 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.
- Srihari, Rohini (1994) Photo understanding using visual constraints generated from accompanying text. In *Proceedings of the Workshop on Integration of Natural Language and Vision Processing*, Mc Kevitt, Paul (Ed.), 22-29, Twelfth American National Conference on Artificial Intelligence (AAAI-94), Seattle, Washington, USA, August.
- Srihari, Rohini and Debra Burchans (1994) Visual semantics: extracting visual information from text accompanying pictures. In *Proceedings of the Twelfth American National Conference on Artificial Intelligence (AAAI-94)*, 793-798, Seattle, Washington, USA, August.
- 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 (1975a) Preference semantics. In *Formal semantics of natural language*, Keenan, Edward (Ed.). Cambridge, United Kingdom: Cambridge University Press. Also as Memo AIM-206, Artificial Intelligence Laboratory, Stanford University, Stanford, California, USA, July 1973.
- Wilks, Yorick (1975b) An intelligent analyzer and understander of English. In *Communications of the ACM*, 18(5), 264-274, May. Also in, *Readings in Natural Language Processing*, Barbara Grosz, Karen Sparck Jones and Bonnie Webber (Eds.), 193-204, Los Altos, California: Morgan Kaufmann, 1986.
- Wilks, Yorick (1975c) A preferential, pattern-seeking semantics for natural language inference. In *Artificial Intelligence*, 6, 53-74.
- Wilks, Yorick (1977) Good and bad arguments about semantic primitives. In *Communication and cognition*, Vol. 10, No. 3/4, 181-221.
- Wilks, Yorick (1978) 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.
- Wilks, Yorick and Afzal Ballim (1987) Multiple agents and the heuristic ascription of belief. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence (IJCAI-87)*, 119-124, Milan, Italy, August.
- Wilks, Yorick, and Okada, N., (Eds.) (in press) *Computer Language & Vision Across the Pacific*. Norwood, NJ: Ablex.
- Wittgenstein, Ludwig (1963) *Philosophical Investigations (translated by G.E. Anscombe)*. Oxford: Blackwell.
- Wright, Ron, E. and Young, D.A. (1990) The cognitive modalities (CM) system of knowledge representation and its potential application in C31 systems. In *Pro-*

ceedings of The 1990 Symposium on Command and Control Research SAIC Report SAIC-90/1508, pp. 373-381.

Young, D.A. (1983) Interactive modal meaning as the mental basis of syntax. In *Medical Hypotheses*, 10:5.