



Title	Meaning in philosophy and meaning in information retrieval (IR)
Authors(s)	Thornley, Clare V., Gibb, Forbes
Publication date	2009
Publication information	Thornley, Clare V., and Forbes Gibb. "Meaning in Philosophy and Meaning in Information Retrieval (IR)." Emerald, 2009. https://doi.org/10.1108/00220410910926158 .
Publisher	Emerald
Item record/more information	http://hdl.handle.net/10197/3579
Publisher's version (DOI)	10.1108/00220410910926158

Downloaded 2026-05-01 23:49:09

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)



© Some rights reserved. For more information

Meaning in philosophy and meaning in information retrieval (IR)

Clare Thornley

School of Information and Library Studies, University College Dublin, Belfield,

Dublin 4, Ireland, and

Forbes Gibb

Department of Computer and Information Sciences, University of Strathclyde, 26

Richmond Street, Glasgow, G1 1XH, UK

Abstract

Purpose -The paper explores the question of whether the differences between meaning in philosophy and meaning in information retrieval (IR) have implications for the use of philosophy in supporting research in IR.

Design/methodology/approach - Conceptual analysis and literature review.

Findings - There are some differences in the role of meaning in terms of purpose, content and use which should be clarified in order to assist a productive relationship between the philosophy of language and IR.

Research Implications -This provides some new theoretical insights into the philosophical context of IR. It suggests that further productive work on the central concepts within IR could be achieved through the use of a methodology which analyses how exactly these concepts are discussed in other disciplines and the implications of any differences in the way in which they may operate in IR.

Originality/value - This paper suggests a new perspective on the relationship between philosophy and IR by exploring the role of meaning in these respective disciplines and highlighting differences, as well as similarities, with particular reference to the role of information as well as meaning in IR. This contributes to an understanding of two of the central concepts in IR, meaning and information, and the ways in which they are related. There is a history of work in IR and information science (IS) examining dilemmas (Neill, 1987; Ellis, 1992, 1996) and this paper builds on this work by relating it to some similar dilemmas in philosophy. Thus it develops the theory and conceptual understanding of IR by suggesting that philosophy could be used as a way of exploring intractable dilemmas in IR.

Keywords Philosophy, Theory, Meaning, Information retrieval

Paper Type Conceptual paper

Introduction

This paper looks at one question: do the differences between meaning in philosophy and meaning in IR matter when IR uses philosophy to inform its theory and practice? This is an important question for IR because despite the existence of an interesting and thought provoking tradition of using philosophy to provide a theoretical basis to the subject and to support system design (see, for instance Blair (1990, 2006) and Hjørland (1998)) there has not been much critical analysis of the exact relationship between meaning in philosophy and meaning in IR. It is often assumed that the discussion on meaning within philosophy must bear a close relationship to the way meaning works in IR. Blair (1990, 2006) argues that Wittgenstein's philosophy is a very productive source of insights into the nature of meaning in IR. In particular the perceived shift from Wittgenstein's early work on logic and meaning (Wittgenstein,

1922) to his work on meaning as a social practice (Wittgenstein, 1953) is seen as a framework for a similar shift in focus within IR from detailed content analysis to an emphasis on context and use. This is discussed by Blair (2006, p.8) in terms of the limitations of the logical analysis of content in refining the precision of meaning definition in documents. This paper explores the hypothesis that work on meaning within the philosophy of language is always pertinent to IR and argues that, in some areas, there are important differences between meaning in IR and meaning in philosophy which should be clarified. A critical discussion of the use of philosophy in IR is provided by examining the nature of the role of meaning in philosophy *vis-à-vis* the role of meaning in IR. This makes a contribution to our understanding of interdisciplinary work in IR and suggests that examining the differences, rather than just the similarities, between meaning in philosophy and meaning in IR could provide some new theoretical insights. In particular the nature of these differences can reveal some of the conflicts and tensions inherent to the problem of meaning and show how these are manifested in IR.

Scope of paper and definitions of terms

Philosophers used in this paper

The problem of defining meaning has a long history in philosophy. This history can be simplified as a divide between the reference theory of meaning (Frege, 1892) and the social practice theory of meaning (Wittgenstein, 1953; Putnam, 1999). The former examines what meaning 'is' by looking for qualities, both within objects and in terms of our mental experience when we understand a word, that somehow contain meaning. This presents the problem of how these two different types of entity, the subjective experience and the objective object, can have a relationship which

accurately produces meaning. Wittgenstein (1953) shifted the nature of this debate by arguing that, rather than looking at a relationship between different entities, we should look at the ways in which we use meaning. He argued that context, which is the social, physical and temporal situation in which symbols that represent meaning are generated and used, is at least as important as content. In his view our ability to understand content is influenced by how much we know about its context. This paper mainly focuses on the social practice theory of meaning within philosophy, as discussed in the work of Wittgenstein (1953), which emphasises the importance of context over content in terms of understanding meaning. It is argued that this approach can be understood as a possible solution to some of the distances and conflicts which exist within meaning. There are many problematic relationships within meaning such as our perception of the objective world and the actual objective world, and our perception of meaning and the perceptions of other people; social practice theory argues that it is our shared social and physical context which makes these relationships work effectively. The potential conflicts between these different aspects of meaning are significantly reduced by the existence of context.

Definition of meaning

In this paper meaning is understood as a complex and often conflicting relationship between the subjective and the objective (Thornley, 2005; Thornley and Gibb, 2007). The relationship between content and context exists within the general subjective/objective relationship. Content is normally what meaning is about in the sense of that to which it refers. This content can be understood as both an external object and also the subjective experience of understanding the meaning of the word that describes the object.

Thus it is argued that in some cases meaning is best understood by looking for its reference or content but that this process can only be understood within the broader context of its use and purpose. The issues under consideration are then whether philosophy and IR are referring to meaning in the same way and whether they are using meaning for the same purpose. The hypothesis is that there are some differences in terms of both content and context and that these differences have implications for developing theory and possibly practice in IR.

Definition of information

The concept of information used in this paper is a change in a knowledge state. This is based on Brooke's (1980) fundamental equation of IS which defines information as that which modifies knowledge structure. More recently this has been described as 'knowledge communicated' (Capurro and Hjørland, 2003). This implies that information must be different from a current knowledge state in order to change it. It also, however, requires some similarity to a current knowledge state as it has to be recognisable. Information does not necessarily have to be totally new, it can just change our level of certainty about something we believed to be true as, for example, when it confirms a hypothesis (Ingwersen and Jarvelin, 2005). Both these theories of information go beyond earlier work on information by, for example, Shannon and Weaver (1949) which focuses on the efficiency and reliability of the transmission of messages regardless of the comprehension of the message by the recipient. Information in this paper is taken to be textual information as most previous work on philosophy and IR deals with this type of information.

Definition of information retrieval

In one of the foundational text books of IR, Salton and McGill provide the following definition of IR.

“Information retrieval (IR) is concerned with the representation, storage, organisation and accessing of information items.” (Salton and McGill, 1983, p.1)

Within IR there are two broad research traditions (Ellis, 1996) which interpret the nature of the problem defined above in different ways. There is the traditional objective approach which focuses on the design of IR algorithms and evaluates systems through large-scale tests (Sparck Jones, 2000) with pre-determined relevance judgements, and a more subjective or cognitive tradition (Ingwersen, 1992; Ingwersen and Jarvelin, 2005) which favours research on the user experience and the role of context and work task on IR. The issues discussed in this paper are pertinent to both traditions although, generally, it is only the latter research tradition which explicitly discusses meaning in its work.

Definition of theory

Theory is used in the philosophical sense of an exploration of conceptual problems which cannot necessarily be proved or disproved by an appeal to data. In information science (IS) and IR this is normally referred to as meta-theory. This paper does not discuss mathematical or logical theory as used either in philosophy or IR, or scientific theory in the sense of hypotheses which can be falsified by data. It should be noted that the later work of Wittgenstein generally discounts the importance of theory in improving our understanding of meaning and regards an accurate description of meaning in use as the best way to promote understanding. This paper does not attempt to provide a generalised theory of meaning in IR; rather the objective is to outline and clarify some new issues for consideration.

Theoretical understanding of the relationship between meaning and IR

It is argued that Wittgenstein's philosophy emphasises the importance of the similarity between aspects of meaning that may appear to be different; i.e. our shared environment and activities are more important than any differences in the 'content' of our understanding. When we are actually using meaning it normally works fairly well; it is only in abstract contemplation that meaning becomes obscure and problematic. Wittgenstein's work reveals the importance of context in minimising and mitigating differences of understanding within meaning.

Much of the work in IR which uses philosophy (Blair, 1990, 2003, 2006) recognises that within IR these differences can regain importance (mainly because of the loss of an immediate shared context) and suggests as a possible solution some reintroduction of this lost context into meaning representation.

“If the information retrieval systems cannot be physically near the activities and practices they support, then it may be useful to bring some of this context into the descriptions of the documents themselves. This enhancement could be done by linking documents to the respective activities and practices in which they might be used, and weeding out information germane to activities that have concluded. The first step in information retrieval design, then is to develop a detailed taxonomy of the various activities and practices that produce or use the information on the system.” (Blair, 2003, pp.37-38)

In IR there is a distance (in terms of time, space, social environment, etc.) between the context in which symbols are generated (in document production and then representation) and the context in which they are retrieved (or, quite often, not retrieved) and used. Blair argues that the usual indeterminacy of language is exaggerated in IR because IR takes place outside the context in which the meanings of the documents were generated.

“We can now see that the ambiguity of indeterminacy in language is unavoidable when it functions outside of the activities or institutions in which it has a role.” (Blair, 1990, p. 323)

Buckland (1991, p.61) claims that “the lack of a direct link between the source of the message and its recipient” is a central feature of IR and that this “delay and indirectness are liable to exacerbate difficulties caused by problems of definability”. Neill (1987) also highlights the importance of the distance between the relatively static nature of information stored in an IR system and the changeable and unpredictable ways in which different users may wish to look for it.

“The history of the development of intellectual access to the store of knowledge is a history of the tension between the fluid uniqueness of the individual enquirer and the essential stability and concreteness of the store of knowledge itself” (Neill, 1987, p.208)

This distance between the storage and analysis of information within the IR system and the actual use of information by a system user can result in a lack of a shared understanding of meaning between information source and information seeker. The indexer and the searcher are actually using different terms to describe the same thing (a document on a topic of relevance to the searcher). Thus in one sense they mean the same thing (in terms of reference to a document or set of documents). They are, however, using different terms to describe these documents, so the user is failing to retrieve relevant documents.

“One of the major obstacles which the enquirer faces is the fact that he may find it difficult to think of all the possible search terms which are likely to have been assigned to documents which he would find useful.” (Blair, 1990, p.53)

Blair (1990, 2006) argues that IR should be understood primarily as a communication process and that the goal of IR is to align the information provider’s (either a human

indexer or algorithm) understanding of the meaning of the document with the understanding of the person searching for the document.

“Inquirers are trying to describe the information they need in a way that indexers would understand and indexers (or automatic indexing processes) are trying to describe the content and context of documents in the collection in ways that would be understandable to the inquirers.” (Blair, 1990, p.188)

In practice, this alignment may be lacking at a number of levels (Gibb and Smart, 1990). In a worst case scenario there may be as many as five world views of the same issue: the language used by an author to express his or her ideas, the terms agreed by domain experts that constitute the controlled vocabulary in an indexing language, the indexer’s interpretation of the indexing language and the document being indexed, the end-user’s expression of their information need, and possibly an information intermediary who is trying to interpret that need. This paper examines, on a conceptual level, the extent to which this potential ‘mismatch of meaning’ really is one of the most important difficulties in IR. It is argued that the main way in which meaning is used in IR, to provide information, has some important differences from meaning as discussed in the social practice philosophy of language. These differences should be clarified in order to understand the ways in which a philosophy of meaning can and cannot contribute to IR.

Structure of paper

This issue is tackled by firstly examining some work in philosophy on meaning and then examining work in IR on meaning. In both cases this is approached by examining the nature of the questions which philosophers or those working in IR are trying to answer. Some of the differences between these two approaches to meaning are then identified and examined. Finally the implications and importance of these

differences are analysed in terms of both an historical understanding of IR and the implications and role of new Web 2.0 technologies.

Meaning in philosophy

In this paper it is argued the most of the philosophy of language can be understood as an attempt to answer the following questions. Firstly, how are the subjective and objective aspects of meaning related; i.e. how can a mental experience of meaning relate to an external object? Secondly, how does meaning manage to be normative i.e. how is it that we normally have a shared understanding of meaning? Finally, is it possible to have a theory of meaning which is complete and has explanatory power?

Wittgenstein (1953) and more recently Putnam (1973, 1999) argue that it is only possible to progress our understanding of these questions if we acknowledge the role of our shared physical and social context. Meaning is a rule based process: generally words relate to objects and we agree on this because there are conventions in place which keep our use of words consistent. Thus it is not really a mystery how the subjective and objective aspects of meaning are related; rather, it is just part of the way we relate to our environment and people around us. Unless there is a shared system of correcting and checking meaning it is impossible to adequately explain how meaning could be used to communicate successfully. Wittgenstein (1953, 35) reminds us that even the most simple actions of language learning, such as pointing to an object and giving its name (i.e. ostensive knowledge), only work because we have learnt what pointing is.

The social practice theory of meaning is interpreted in this paper as the view that some differences within meaning which appear to be important are in fact rarely significant; for example the difference between different individual's understandings

of meaning and the difference between an object and the meaning of the word which represents it. In most cases a shared physical and social context makes the relationship between these apparently disparate aspects of meaning fairly straightforward. There are, however, some differences within meaning that we tend to ignore but which are in fact very important. These include the great variety of different ways that we use language such as for naming, arguing, praying, begging etc., which Wittgenstein (1953, 23) calls the “multiplicity of language games”. Wittgenstein argues that we get the importance of these differences the wrong way round. Thus it appears difficult to answer the questions of meaning concerning the relationship between the subjective and objective and the ability of meaning to be normative when in fact these are relatively straightforward problems once the role of context is acknowledged. In contrast much energy in philosophy has been directed at developing a theory of meaning, including to some extent the earlier work of Wittgenstein (1922), on the assumption that this was a problem which was possible to solve through a general theory. Once one acknowledges the incredible variety and differences within meaning, however, this project becomes untenable and Wittgenstein argued that the only way to understand meaning is to describe its practice accurately rather than to develop generalised theories.

Wittgenstein’s analysis of language seems to cover both these issues of similarity and difference within meaning. He uses the expression ‘form of life’ to describe the general shared social and physical context in which meaning operates and in this sense the unity of language is emphasised.

“ “So you are saying that human agreement decides what is true or false?” - It is what humans *say* that is true or false; and they agree in the *language* they use. This is not agreement in opinions but in form of life.” (Wittgenstein, 1953, 241)

The term ‘language game’, however, is used to describe the great variety of different ways in which we use language and the particular nature of some contexts so, in this sense, the diversity of language is emphasised. Philosophical work on meaning within the social practice tradition aims to shift our approach to questions about how meaning works. Instead of explaining how different aspects of meaning may be related within a theory it suggests a reassessment of these differences through the use of context and an understanding of meaning based on observation and description rather than on abstract theories. This does not mean that content or reference are irrelevant to meaning, see for example commentary on Wittgenstein by Winch (1969), rather that they are only part of a complex picture.

Meaning in IR

This section explores the relationship between the problem of meaning within philosophy and the problem of meaning within IR. The shift in focus within philosophy from a reference or content based theory to an approach emphasising context and diversity of use has been described earlier in this paper and, to some extent, a similar shift in focus has occurred within IR. This development of meaning within IR is discussed by examining the history of IR’s approaches to the main meaning questions identified previously. Firstly, how are the subjective and objective aspects of meaning related; i.e. how can a mental experience of meaning relate to an external object? Secondly, how does meaning manage to be normative; i.e. how is it that we normally have a shared understanding of meaning? Finally, is it possible to have a theory of meaning which is complete and has explanatory power?

The subjective/objective relationship

The question of the relationship between the subjective experience of meaning and an external object is normally manifested in IR as the problem of how to relate the information need of the enquirer to the existence of a relevant document. This is made more complex because the document is normally represented by some kind of document surrogate and the information need is normally represented as a query. Thus IR is not just a problem of meaning but a problem of meaning representation. The question is then whether this should be seen as a problem of content, or one of context, or some combination of both. Content can be defined as the text within the document and the query. Context can be seen as part of content in terms of how the content of one document relates to other documents within the collection or as something external to the text in terms of the document's possible uses and the context from which the information need arises.

In his early work in IR, van Rijsbergen discusses the dilemma of representing content in terms of both an individual document and its relationship to other documents.

“There are two different ways of looking at the problem of characterising documents for retrieval. One is to characterise a document through a representation of its contents, regardless of the way in which other documents may be described, this might be called representation without discrimination. The other is to insist that in characterising a document one is discriminating it from all, or potentially all, other documents in the collection i.e. discrimination without representation.”(van Rijsbergen, 1979, p.29)

He concludes that “in practice one seeks some sort of optimal trade off between representation and discrimination” (1979, p.29). In this case it is the content of the document which is the primary concern and its relationship to the content of other documents. Context is an issue in the sense of other related documents but not in the sense of the use or social context of the documents concerned.

The difficulty with content based representation can be ambiguity. This ambiguity can take a number of different forms. Some words are homonyms i.e. have the same form but different senses as, for example, the term 'bank' in 'river bank' or 'high street bank'. Some words are metonyms i.e. the name of an attribute of an object is substituted for the actual object, as, for example, when 'crown' is used instead of 'king'. As well as these particular cases there is the general problem, as Blair observes (1990), of the many different ways in which any one object or event can be described. Thus a document representation and a query may share subject matter but not many terms because different terms have been used by the indexer and the user to describe similar subject matter. In this case the document would not be retrieved even though it probably would be relevant; i.e. a failure of recall. Alternatively a document representation and a query may share the same terms but they may not be about the same thing because their different contexts may give them different meanings. In this case the document would be retrieved but would probably not be relevant; i.e. a failure of precision.

The normative nature of meaning

This then raises the question of how meaning manages to be normative or successfully shared; how do we know that other people mean the same thing by using words as we do? Within IR this is normally seen as the problem of how the perception of meaning held by the enquirer and the perception of meaning held by either the human indexer or automatic system can be optimally shared. This can be difficult because, as Blair (1990) emphasises, there is room for interpretation and ambiguity in IR; i.e. there are many different ways to describe the same document and it is hard to

ensure that the enquirer will use similar terms to the indexing system when describing the same document.

“One of the major obstacles which the enquirer faces is the fact that he may find it difficult to think of all the possible search terms which are likely to have been assigned to documents which he would find useful.” (Blair, 1990, p.53)

“It is impossibly difficult for users to predict the exact words, word combinations and phrases that are used by all (or most) relevant documents and only (or primarily) by those documents.” (Blair, 1990, p.101)

Blair argues that the most effective way to represent the content of a document is to analyse its meaning which is primarily a product of its context and use rather than just a product of its textual content.

“The goal of any document indexing strategy should be to build as much of the missing activity or institutional context back into the language of representation.” (Blair, 1990, p.323)

“To represent the content of information, we must understand what it means, and to understand what it means we have to understand how it is used.” (Blair, 2006, p.337)

Thus in IR the debate about the subjective/objective relationship in meaning and the problem of how meaning manages to be normative are to some extent a reflection of the same debate in philosophy in terms of the respective roles of content and context.

Developing a theory of meaning

In terms of developing a theory of meaning which could contribute to IR there is a very broad range of views on what might constitute a theory in this case and the usefulness or otherwise of pursuing this ambition. One of the most fundamental questions is whether trying to understand meaning, either in terms of more complex

linguistic analysis of content or in terms of the wider context and use of documents, is actually relevant in terms of developing workable IR systems. This debate has some similarity to the issues discussed in Searle's (1984) Chinese Room analogy. This parable, as Searle describes it, is concerned with the distinction between syntax and semantics. Syntax is the set of formal procedures or rules of language; computers operate at this level. Semantics is what these symbols stand for and can be understood as their content or what they are about; i.e. they are the subjective experience of meaning. Symbols by themselves have no meaning because they have no semantic content; i.e. they are the objective aspect of meaning. Thus a computer can check through its interpreter that its code is syntactically valid but it cannot know what the code means or to what it refers. The Chinese room analogy involves an English speaking man in a room with a rulebook in English for the manipulation of Chinese symbols. Symbols are passed to him; he manipulates them using the rulebook and then passes them out. Thus, it might appear to an observer that the man in the room understands Chinese but he does not. He is merely using rules to manipulate symbols. He is running the program for manipulating Chinese symbols but this cannot be equated with an actual understanding of Chinese.

Searle argues that even if the rule following procedures might convince an observer that understanding is occurring, it is in fact not occurring. When the observer found out that the man was using a book, he would realise that he had not, in fact, been dealing with someone who could understand Chinese. He would not be convinced that it is no longer necessary to learn foreign languages because understanding them is just a question of getting a good rulebook. Understanding is something more than simply the ability to convince somebody else that you do have understanding. Interestingly Turing (1950) uses a very similar kind of thought experiment to argue the exact

opposite: that thinking is nothing over and above the ability to convince someone that one is thinking. In his case if a computer can convince an interrogator that it is thinking then that is counted as convincing evidence. This is contrasted to gender whereby if a (hidden) woman can persuade an interrogator that she is a man then this just proves that the interrogator is mistaken. Thus gender is irreducibly biological whilst thinking is not.

Within IR the rulebook could be said to be the many effective and well-tested statistical techniques that use term frequency and distribution to assess the extent of similarity between documents and queries. Ingwersen and Jarvelin (2005, p.35) compare automatic indexing to the process in Searle's Chinese room. These algorithms are not explicitly to do with meaning in that they clearly do not concern themselves with mental states or the use and social context of documents but they do manage to produce results which are often acceptable to the user. As Sparck Jones (2004) argues many of the statistical models from the early work in IR, are now used in web technology precisely because more complex natural language processing technology, which can be seen as having closer theoretical links to meaning, often does not work well enough. These statistical techniques offer theory in terms of mathematical models but remain frustrating for many working in IR (and also often in IS) as they cannot really offer any explanation as to why they work in terms of our ordinary understanding of language. This is not an argument against using these techniques but it is an important factor to consider when analysing the potential role of a theory of language in IR. Does it matter that these techniques produce reasonable results even though they appear to have no clear link to a particular theory of meaning or language? Does understanding IR within the context of a theory of language have value, for example possibly in the longer term, even if reasonable techniques which

don't use this already exist? Within most disciplines it could reasonably be assumed that an understanding of theory would provide some guidance for improving practice; i.e. if we can identify why something works it will be easier to exploit these factors to improve performance.

In her early work on IR testing Sparck Jones (1981, p.216) observes that the history of IR research at the time could be described as “a long and not altogether successful attempt to convert descriptive hypotheses into explanatory ones.” In van Rijsbergen's early work we also see some ambiguity about the practical importance of a detailed understanding of language and meaning in developing effective IR systems.

“Undoubtedly a theory of language will be of extreme importance to the development of intelligent IR systems. But, to date, no such theory has been sufficiently developed for it to be applied successfully to IR. In any case satisfactory, possibly even very good, document retrieval systems can be built without such a theory.”(van Rijsbergen, 1979, p.15)

Work in IR which uses the philosophy of meaning is, however, normally explicitly concerned with how an understanding of language could help design better IR systems. Blair (1990) argues that IR is a linguistic problem and thus if we want to understand IR we first have to understand language.

“We must have some idea how language works if we are able to use it effectively in representing documents for retrieval.” (Blair, 1990, p.321)

So the debate within philosophy on the possibility and efficacy of a theory of meaning is manifested within IR in a similar way. There is a tension between the pragmatic approach that an accurate description of techniques which work (notwithstanding the inherent difficulty of IR evaluation) is the most fruitful line of enquiry and those that argue that a deeper understanding of meaning must eventually, at least in the longer

term, have some impact on the effectiveness of IR systems. Interestingly it is those in the objective tradition in IR who in one sense could be said to be more aligned to the social practice philosophy of language at least in terms of their pragmatic approach to design improvements.

The differences between meaning in philosophy and meaning in IR

The previous section discussed some of the similarities between the way meaning is discussed in philosophy and the way it is discussed in IR in terms of how they approach the central questions of meaning. This section examines some of the differences and their possible implications for our understanding of the role of meaning in philosophy and IR. It is argued that meaning in philosophy has some important differences, both in terms of purpose, content, context and use, than meaning in IR.

Purpose of investigation

Firstly it is necessary to examine the general purpose of a philosophical analysis of meaning as opposed to an analysis of meaning for IR. Wittgenstein's work on meaning can be understood as an attempt to dismantle our ubiquitous misunderstanding about meaning and reveal its actual use in action. This has a philosophical purpose in so far as it clarifies our ways of thinking and stops us being muddled by our misunderstandings of language.

“Philosophy is a battle against the bewitchment of our intelligence by means of language.” (Wittgenstein, 1953, 109)

He is not, however, trying to solve a particular social or technical problem. Within IR meaning is normally analysed with the purpose of improving our ability to design and use IR systems.

Does understanding meaning have to lead to practical results for IR and, if so, should we insist that these seem technically possible in the short-term or work on the assumption that a detailed understanding of meaning will become increasingly useful as computers become more able to deal with this level of complexity? This paper does not present a particular solution to this question but the issue of whether the problems raised by IR about the nature of meaning are of scholarly interest regardless of their immediate technological implications is perhaps ripe for further discussion. Theoretically speaking it may matter that philosophical ideas are used in a way which links them to practical results as this has the potential to distort interpretations to fit in with what is practically possible. If, as Wittgenstein argues, the meaning of a word is its use then if meaning is used in IR in the sense of a practical and technological question rather than a philosophical one, whilst simultaneously using philosophy as a theoretical and conceptual framework, this may have the potential to create exactly the kind of conceptual misunderstandings which Wittgenstein warns us against. This is not an argument against using philosophy in IR, rather an observation that philosophy may improve our understanding of problems in IR but we cannot assume that this improved understanding will always lead us to, at least in the short term, proposed solutions. If information science is to be understood as an academic discipline as well as a technological and professional endeavour then we should not shy away from exploring problems that we may find hard to solve or oversimplify them to give the impression that we can solve them.

Purpose of IR

This is the question of whether IR needs to concern itself with meaning in order to produce documents relevant to the enquirer. This is not just a practical question of whether, as a matter of fact, sophisticated language models actually do improve performance, but a question of whether the question of meaning actually has any pertinence to the purpose of IR. This, of course, depends both on what is meant by meaning and what the purpose of IR is perceived to be. Both these questions are complex and the history of the debate on this issue raises some interesting questions about the role of meaning in IR.

Salton and McGill (1983), in one of the foundational textbooks for IR, argue that sophisticated analysis of meaning in terms of document content is not appropriate to the purpose of IR. This is not a problem with the language analysis technology *per se* but rather that language models which try and incorporate meaning are not relevant to IR. This kind of analysis is more suited to artificial intelligence (AI) systems which generally try and provide answers to questions in a tightly circumscribed subject area; IR's goal is to provide relevant documents.

“...a fundamental difference exists between information retrieval on one hand and certain other language processing tasks on the other. In retrieval one needs to render a document retrievable, rather than to convey the exact meaning of the text. Thus two items dealing with the same subject matter but coming to different conclusions are treated identically in retrieval, that is, either they are both retrieved or they are both rejected. In a question-answering system or language translation system, these documents would of course be treated differently.” (Salton and McGill, 1983, p.267)

Thus, from this perspective, the exact meaning of the text is not relevant to IR; it is sufficient to approximately match the subject matter. Ingwersen, coming from a very different research tradition than Salton and McGill, also argues against complex

natural language processing claiming that “the principle of translating request and texts into meaning goes against the entire issue of ‘aboutness’” (1992, p.191).

Ingwersen argues that IR should be concerned with information rather than meaning because “information goes beyond meaning” (1992, p.25); i.e. analysing the meaning of a document is no guarantee that it will successfully inform a user.

“IR, regardless at which level it is performed, is not and is *not* intended to be satisfied with a semantically correct ‘translation’ of any text or picture. Retrieving meaning is not sufficient, or indeed perhaps not necessary in IR. IR is pre-occupied with providing information, which may act as a supplement to a human conscious or unconscious mental condition in a given situation.” (Ingwersen, 1992, p.25)

In Ingwersen’s case this reveals his concern for all the factors outside the text such as work task, etc., which will affect the relevance of a document over and above a textual analysis of its meaning. Here he uses ‘meaning’ in the sense of the content of the document within the IR system and ‘information’ to describe the factors outside the document which can affect understanding and use of the document. In his view the main concern must be the change in knowledge state of the user and an exact analysis of the meaning of the document is not necessarily required for this to happen. Thus, in contrast to Blair, his view is that agreement in meaning is not in fact one of the most important problems in IR because its relationship towards becoming informed is not clear cut.

“Evidently, at a pragmatic level individual actors may share the conventional understanding or idea of a sign *and* the information interpreted from the sign; but we may also frequently expect that the actors obtain different information (and cognition) although their understanding of a sign is in agreement. This phenomenon gives rise to

the notion that *information goes beyond meaning*.”(Ingwersen and Jarvelin, 2005, pp. 37-8)

Here we see that, from this perspective, meaning, in terms of the content of document, may not always have a direct causal relationship to the process of becoming informed; i.e. how the user interprets and uses the document. For some researchers within the traditional objective approach in IR this issue is not within the remit or purpose of IR. In his early work Van Risjbergen (1979) cites Lancaster (1968) when discussing the definition of IR. Within this definition the purpose of IR is not to inform the user in terms of changing their knowledge state, rather, “it merely informs of the existence (or not) and whereabouts of documents relating to his request” (Lancaster, 1968 in van Risjbergen, 1979, p.4). Within this perspective if a document is retrieved which has some relationship to the request then the role of IR in the process has come to an end. The exact ways in which it may be used and the myriad of factors, such as previous knowledge, task situation etc., which may affect its pertinence for the user, are not the kind of issues with which IR system designers should concern themselves.

This discussion reveals a number of different perspectives on the role of meaning in fulfilling the objectives of IR. If meaning is understood as primarily a question of social context and use, as in Blair’s work, then he argues that the lack of this context in IR is its central problem. This divorce from context exacerbates ambiguity and leads to a lack of shared understanding of meaning between users and systems analysing textual content. Thus an understanding of the role of context and use in meaning could lead to a greater agreement on the meaning of terms between the user and the system and thus more effective retrieval. If IR is seen primarily as a communication process between user and system a shared understanding of meaning is central to improving its efficacy.

If meaning is interpreted as a detailed analysis of the linguistic content of the text, over and above statistical term frequencies, then some within the traditional school of IR (van Rijsbergen, 1979; Salton and McGill, 1983; Sparck Jones, 2004) argue that this is not normally justified in terms of results, as statistical techniques are so effective. Sparck-Jones (2004) observes that the statistical technique of *tf*idf* weights (term frequency-inverse document frequency), which is an old technique within the time span of IR, is still extensively used in modern web search-engines and has shown potential in newer forms of text analysis.

“At the same time there has been a striking new development in the last decade, stimulated by TREC and related language and information processing applications that have explored a range of tasks including topic tracking, question answering, and summarising. These tasks rely, in various ways, on identifying key texts or text segments for more detailed attention, and *tf*idf* has proved a very handy tool for this purpose.” (Sparck Jones, 2004, pp. 522-523)

Thus, from this research perspective, the establishing of a shared perspective on meaning between user and document analysis system is not the central problem of IR because it is not an essential component of refining and improving statistical techniques.

If meaning is understood as the content of documents and information as factors outside the document which influence its interpretation and use, as argued in the work of Ingwersen (1992), then the analysis of meaning is only one part of designing effective IR systems. Information, in terms of changing the knowledge state of the user, is to do with many factors which are not to do with meaning in Ingwersen's theory; i.e. which are outside the content of the text. The process of becoming informed is not the same thing as accurately communicating meaning and thus a shared perspective of meaning between the user and the system is not essential. In

Ingwersen's view a text can have different significance to different people and multiple interpretations and this is not a weakness of IR but something to be exploited in terms of providing multiple perspectives on documents.

Thus the relevance of meaning to the purpose of IR has different interpretations depending on how different research traditions interpret meaning and how broadly they interpret the scope of IR; i.e. should it concern itself with factors outside the immediate IR system. Wittgenstein's work primarily focuses on the use of context to reduce ambiguity in meaning, however, it would appear from this review of IR perspectives, that this should not necessarily be the prime concern of IR and that it may not always lead to more effective IR systems.

The role of information in IR

It can be seen from the previous discussion that one difference between meaning in philosophy and meaning in IR is the additional aspect of informing with which IR contends. This debate highlights one of the dilemmas and differences between meaning and information in IR. The central issue can be seen as whether one focuses on IR as communication as we see in Blair, in which case agreement on meaning is central, or whether IR is seen as primarily as information provision, in which case some agreement in meaning is certainly very important but it is not identical with successful informing. In this case information requires enough shared meaning to allow some communication to occur but this meaning does not have to be identical. In a previous paper (Thornley and Gibb, 2007) it is argued that IR in many ways operates within the tension between meaning and information. Thus it is both an act of communication and an act of informing and change and that these aspects of IR can

often be strongly related but are also in conflict, hence the development of the dialectical model of IR (Thornley, 2005).

There needs to be both some similarity between the query and the retrieved document (i.e. some shared meaning) for it to be relevant. However, there also need to be some differences (i.e. some information) between the query (and the associated knowledge state) and the document in order for the document to provide information. At the simplest level a document and a query must share some terms for the document to be relevant to the query but they must also not share identical terms or the document could not inform the user; i.e. it would be exactly the same as what they had input as a query. At the same time these differences between the document and the query have to be similar enough to the user's current knowledge state to be recognisable. Salton and McGill summarise the essential workings of an IR system as a way of assessing the similarity between queries and documents.

“Every information retrieval system can be described as consisting of a set of similar items (DOCS), a set of requests (REQ), and some mechanism (SIMILAR) for determining which, if any, of the information items meets the requirements of the requests.” (Salton and McGill, 1983, p.10)

The relationship between the level of this similarity, however, and whether or not “it meets the requirements of the requests” is a complex one. Bawden's (2006) analysis of convergence and divergence in IR discusses the often paradoxical relationship between the role of IR in matching the similarity of documents and queries, and the use of information to support creative thinking and ideas. He observes that “IR systems work essentially by means of convergent information processing in that their goal is to achieve as close a match as possible between the representation of queries

and retrieved documents”. He argues, however, that creative thinking is often divergent but that this in itself can rely upon the identification of similarities between apparently differing and conflicting entities or themes. The surprising and unexpected in IR can also be the most useful tool for creative thinking.

“There would appear to be a considerable mismatch between the element of serendipity that often characterises creative thinking and what IR systems are essentially designed to deliver.” (Bawden, 2006 p.537)

Thus there is often a tension between an IR system’s purpose in terms of delivering documents with similar meaning to the query provided by the user and an IR’s system’s purpose in terms of informing or changing the knowledge state of the user. Communicating through spoken language, as discussed by Wittgenstein and other social practice theory philosophers, is often about ensuring that enough meaning is shared so that a message can be communicated. This is related to, but also different from, searching for relevant documents in so far as ‘the message’ could be useful in ways which are unpredictable and tangentially related to its content, and thus it is not always appropriate to try and make IR more like everyday communication in language.

Role of technology

A clear difference between meaning as discussed in philosophy and meaning in IR is that IR uses technology to manipulate meaning. What implications does this have for the relevance of a philosophical understanding of meaning to IR? The base technology which IR exploits is the technology of writing. Wittgenstein and Putnam primarily focus on the immediate spoken word and do not discuss in any great detail the nature of meaning within text. In most case IR deals with recorded meaning, i.e.

meaning which has been separated from its original source and context, and this is, in many but not all cases, of a textual nature. The relative balance between communication and information may well be very different in the spoken word as opposed to documents. The lack of context in documents, in terms of its removal from the time and place in which they were originated, is also not only a weakness in terms of communication but often a strength, in terms of information, as it allows information to exist through time and in different places. Thus it is quite an intellectual leap to assume that work on the meaning of the immediate spoken word will always be appropriate when dealing with text.

The more recent technology of computers and, specifically, the new developments in terms of social computing in Web 2.0 (O'Reilly, 2005), raise some interesting questions about the role of meaning in IR. Firstly these technologies have improved the analysis of the relationship between documents as can be seen in Google's use of 'PageRank' (Brin and Page, 1998) which exploits information on links between documents when assessing relevance. In terms of meaning this can be interpreted as an increased role for context in terms of relationships between documents. In their paper on Google Brin and Page (1998) provide what they term an intuitive justification for 'PageRank' in so far as "it can be thought of as a model of user behaviour": 'PageRank' gives a high weight to pages which are frequently cited or have cited pages which themselves are frequently cited.

"The citation (link) graph of the web is an important resource that has largely gone unused in existing web search engines. We have created maps containing as many as 518 million of these hyperlinks, a significant sample of the total. These maps allow rapid calculation of a web page's "PageRank", an objective measure of its citation importance that corresponds well with people's subjective idea of importance." (Brin and Page, 1998. p.112)

Is an “intuitive justification” and corresponding “well with people’s subjective idea” as near to a theory as an IR system can get? It is interesting that no mention of philosophy is made in this paper even though the use of collective information on page use does have some relationship to the social practice philosophy of language. Does it matter that philosophy was not the inspiration even if it could be used later as some kind of theoretical framework?

Web 2.0 can also be seen to add context in so far as it allows users to generate and annotate content, such as for example wikipedia (<http://en.wikipedia.org>), which is then shared amongst all users. Does this make meaning in IR more like the social practice activity of meaning as discussed in philosophy? Does the ability to adapt and link users negate some of the qualities of writing technology which tend to fix meaning? In terms of information provision does it make access to documents more like communication and less like information? The collaborative nature of ‘wikipedia’ does seem to work in a very similar way to the community based normative nature of meaning as discussed by Wittgenstein, as do folksonomies, and the collaborative filtering used by sites such as ‘amazon’ (<http://www.amazon.com>) also exploits the potential shared interests of users by pooling the content of their searches. In analysing this question it is firstly important to acknowledge that these new technologies have not totally replaced many of the IR systems dealing with detailed academic documents, they provide a new way of accessing some knowledge. Thus they provide a new model of accessing and organising information which works well in some cases although the lack of peer or expert review does raise some concerns about its reliability and accuracy. In work on the nature of online communities (Komito, 2001) there is also an acknowledgement of the difference between

communities of locality, which generally have to incorporate differences, and conflict and virtual communities which can contain self-selected, like-minded people. It may be that collaborative information analysis also tends to reinforce views and affiliations rather than effectively impart new and complex knowledge. Thus it should be understood as one of the language games within IR and not as a total replacement for the other ways of analysing and accessing information.

As well as these developments in collaborative technology the more traditional and objective tradition within IR is becoming more interested in context in its various manifestations including tailoring IR systems to the needs of particular users (Ingwersen, Ruthven and Belkin, 2007). In philosophical terms this seems more in line with Wittgenstein's views on the variety of different ways of using language although, in a similar way to Web 2.0, the impetus for research does not appear to be philosophical theory.

Conclusions

In returning to the original question under consideration in this paper “do the differences between meaning in philosophy and meaning in IR matter?” what kind of conclusions can be drawn? It has been argued that although there are some similarities between the role of meaning in philosophy and the role of meaning in IR there are also some important differences in terms of content, context, use and purpose. These are often to do with the relationship between information and meaning in IR. The communication of shared meaning is related to but also different from the process of becoming informed. Indeed it is this complex and dilemma-ridden relationship which is at the heart of many of the conceptual and technical difficulties of IR. These differences between meaning and information are not always something that can or

should be overcome in order to 'improve' IR systems, rather, in some cases; they reveal the importance of fundamental and intractable differences and conflicts in IR. Thus we can use philosophy to illuminate the nature of these intractable problems in IR and thereby develop our theoretical understanding. One insight that we can certainly take from the social practice philosophy of language is that progress in understanding will come from balancing conflicting and diverse requirements in information provision rather than through the development of general solutions.

References

- Bawden, D. (2006), "Einstein in the office: is information really necessary?", *Journal of Documentation*, Vol. 62 No. 3, pp. 305-306.
- Blair, D.C. (1990), *Language and Representation in Information Retrieval*, Elsevier Science Publishers, Amsterdam.
- Blair, D.C. (2003), "Information retrieval and the philosophy of language", *Annual Review of Information Science and Technology*, Vol. 37, pp. 3-50.
- Blair, D.C. (2006), *Wittgenstein, Language and Information: 'Back to the Rough Ground!'*, Springer, Dordrecht.
- Brin, S. and Page, L. (1998), "The anatomy of a large-scale hypertextual web search engine", *Computer Networks*, Vol. 30 Nos. 1-7, pp. 107-117. Available at: <http://infolab.stanford.edu/pub/papers/google.pdf> (accessed 28 March 2008).
- Brookes, B.C. (1980), "The foundations of information science. Part I: philosophical aspects", *The Journal of Information Science*, Vol. 2, pp. 125-133.
- Buckland, M.K (1991), *Information and Information Systems*. Greenwood, New York.
- Capurro, R and Hjørland, B. (2003), "The concept of information", *Annual Review of Information Science and Technology*, Vol. 37, pp. 343-411.
- Ellis, D. (1992), "The physical and cognitive paradigms in IR research", *Journal of Documentation*, Vol. 48 No. 1, pp. 45-64.
- Ellis, D. (1996), *Progress and Problems in Information Retrieval*, Library Association Publishing, London.

Frege, G. (1892), "On sense and reference", *Zeitschrift für Philosophie und Philosophische Kritik*, 100, pp. 25-50, in Moore, A.W. (Ed), (1993), *Meaning and Reference*, Oxford University Press, Oxford.

Gibb, F. and Smart, G. (1990), "Structured information management processing and retrieval - potential benefits for the offshore industry", in Myers, A. (Ed). *Proceedings of the Offshore Information Conference, Glasgow, 1990*, Institute of Offshore Engineering, Edinburgh, pp. 189-201.

Hjørland, B. (1998), "Information retrieval, text composition, and semantics", *Knowledge Organisation*, Vol. 25 No. 1/2, pp. 16-31.

Ingwersen, P. (1992), *Information Retrieval Interaction*, Taylor Graham, London.

Ingwersen, P. and Jarvelin, K. (2005), *The Turn: Integration of Information Seeking and Retrieval in Context*, Springer, Dordrecht.

Ingwersen, P., Ruthven, I. and Belkin, N. (2007), "First International Symposium on Information Interaction in Context", *ACM SIGIR Forum*, Vol. 41 No. 1, pp. 117-119.

Available at: http://www.sigir.org/forum/2007J/2007j_sigirforum_ingwersen.pdf

(accessed 28 March 2008).

Komito, L. (2001), "Electronic communities in an information society: Paradise, mirage, or malaise?", *Journal of Documentation*, Vol. 57 No. 1, pp. 115-129.

Lancaster, F.W. (1968), *Information Retrieval Systems: Characteristics, Testing and Evaluation*, Wiley, New York.

Neill, S.D. (1987), "The dilemma of the subjective in information organisation and retrieval", *Journal of Documentation*, Vol. 43 No. 3 pp. 193-211.

O'Reilly, (2005), *What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software*. Available at:

www.oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

(Accessed 28 March 2008)

Putnam, H. (1973), "Meaning and reference", *The Journal of Philosophy*, Vol. 70 No. 19. in Moore, A.W. (Ed), (1993), *Meaning and Reference*, Oxford University Press, Oxford.

Putnam, H. (1999), *The Threefold Cord: Mind, Body, and the World*, Columbia University Press, New York.

Salton, G. and McGill, M.J. (1983), *Introduction to Modern Information Retrieval*, McGraw-Hill International, Tokyo.

Searle, J. (1984), *Minds, Brains and Science*, Harvard University Press, Cambridge.

Shannon, C.E. and Weaver, W. (1949), *The Mathematical Theory of Communication*, University of Illinois Press, Urbana.

Sparck Jones, K. (1981), "Retrieval system tests 1958-1978", in Sparck Jones, K. (Ed), *Information Retrieval Experiment*, Butterworths, London, pp. 213-255.

Sparck Jones, K. (2000), "Further reflections on TREC", *Information Processing and Management*, Vol. 36 No. 1, pp. 37-85.

Sparck Jones, K. (2004), "IDF term weighting and IR research lessons", *Journal of Documentation*, Vol. 60 No. 5, pp. 521-523.

Thornley, C. and Gibb, F. (2007), "A dialectical approach to information retrieval", *Journal of Documentation*, Vol. 63 No. 5, pp. 755-764.

Thornley, C. (2005), *A dialectical model of information retrieval: exploring a contradiction in terms*, PhD thesis, University of Strathclyde, Glasgow. Available at:

[http://www.cis.strath.ac.uk/~ir/research_students/digital%20library/clarethornleyphd.](http://www.cis.strath.ac.uk/~ir/research_students/digital%20library/clarethornleyphd.pdf)

pdf (accessed 28 March 2008)

Turing, A.M. (1950), "Computing machinery and intelligence", *Mind*, LIX No. 236, pp. 433-460.

Van Rijsbergen, C.J. (1979), *Information Retrieval*, Butterworths, London.

Winch, P. (1969), "The unity of Wittgenstein's philosophy", in Winch, P. (Ed), *Studies in the Philosophy of Wittgenstein*, Routledge, Kegan and Paul, London, pp. 1-20.

Wittgenstein, L. (1922), *Tractatus Logico Philosophicus*. Ogden, C.K. (Trans), Routledge, London.

Wittgenstein, L. (1953), *Philosophical Investigations*, Blackwells, Oxford.