IR and the Dialectic of Meaning

Clare Thornley

Dept of Information Science, University of Strathclyde
Glasgow, Scotland
clare@dis.strath.ac.uk

Ruben Leon

Dept of Information Science, University of Strathclyde
Glasgow, Scotland

Abstract

The aim of this paper is to provide some elucidation of the tensions inherent in the problem of meaning and use this to inform both theoretical and pragmatic issues in IR. The difficulties of theory in IR are discussed and a new model of meaning as a dialectic is introduced. This model consists of two dialectics: objective versus subjective and individual versus shared. It is shown that this model assists IR in revealing some of the difficulties in its theoretical assumptions and thus provides a clearer focus for improvements in practice.

1 Theories of Meaning

The general aim of any theory is to provide a way of understanding the nature of a particular phenomenon and as a result to have a tool for explaining and predicting its behaviour. In developing a theory of meaning then, the first step must be to examine what kind of thing meaning is. This process raises many questions including: what is the relationship between a word and its meaning; what happens when we understand the meaning of a word; how can we tell that other people use words in the same way; why is meaning sometimes ambiguous. It is a complex problem and even the most superficial study of meaning reveals that it involves a wide range of different factors. Many philosophers and linguists have developed ideas as to how language works but there is no clear consensus on the nature of the problem [1-5].

I argue that one way to gain insight into the problem of meaning is to understand it as a series of relationships which exist under tension. These relationships consist of the following sets of oppositions: the individual and the world; the individual and other individuals. Theories about language, or disciplines such as IR which often work in language with no explicit theory, will have assumptions both about the nature of the constituents of these relationships and the form these relationships take. It is not possible to work without theory and claims that IR is just a question of engineering are just as dependent on a particular theoretical perspective as the most advanced logical models.

Given the difficulty of the problem of meaning, I argue, following Wittgenstein, that the best approach is to work towards a series of elucidations about how language actually works, rather than to try and build a monolithic theory and then force language to inhabit it. IR requires an understanding of how meaning happens in the particular process of information seeking and in the next section I expand on why I think this process can be made clearer by using a model of meaning as a dialectic.

2 Meaning as Dialectic

A dialectic is a relationship of opposites in tension, the elements of the opposition often require each other but they cannot be reconciled together. I argue that meaning is a product of such tensions which cannot be resolved, as we move towards one aspect of meaning (e.g. objectivity) we must necessarily move away from another aspect (e.g. subjectivity). This means that every gain in one aspect results in a corresponding loss in another. In terms of IR, this provides new insight into the inverse relationship between recall and precision and the earliest experiments by Cleverdon [6] which saw indexing systems as tools to either increase specificity or exhaustivity. The problem is balancing the gain against the loss and the appropriate decision here cannot be a general one but must depend on the context of a particular information need. A 'holy grail' solution would mean a resolution of a dialectic which is not possible, rather progress can be made by articulating some of the tensions and showing how a particular location on these dialectics will affect IR. I will now examine two of these tensions explaining the general theoretical problem, its relation to IR and possible implications for practice.

3 Objective versus Subjective

This tension in language arises from the fact that, in order to describe the world, language must have some relationship to objects and this relationship must also make sense to the people that use language. Language is both about the world (objective) and about us (subjective). What is the mechanism behind the fact that words refer to things in the world and to what extent is this process of reference at the centre of meaning.

3.1 Theoretical Problem

Objective normally means the perspective that reveals things as they really are and is not influenced by the interests or needs of the subject (the person observing). It is the view that is most about the object and least about the subject. It can also be seen as the perspective which is least influenced by the position of the subject, the philosopher Thomas Nagel, defines objectivity as the view from nowhere [7]. The problem of objectivity also raises the question of how our perceptions unavoidably influence what we observe, this question was raised by Kant in the 18th century and more recently by modern physics. If we could perceive things exactly as they are (the Kantian term is *dinge an sich*, thing in itself) then the subject would need to disappear as any perception can only be from a particular perspective [8]. If you wish to lose the influence of the perspective then you have to lose the person who perceives.

In terms of the problem of meaning this tension between the objective and subjective is central in that language is the medium through which the subject perceives and interacts with the object (things in the world). The quality of this dual nature is complex in that a word's relation to an object can either be seen as some kind of objective or transcendental fact or it can be seen as a product of human usage. The former view is taken by the early work of Wittgenstein [3] and the imperative then is to improve language by tightening up this relationship (devising a language in which each word unambiguously refers to an object). His later work [4] takes the second perspective and suggests that language is best understood by examining how it is used in particular situations, attempts to improve language only result in an inability to see how it actually operates.

Whatever perspective is taken on this question it appears to be an empirical fact that a word is both about something in the world and a tool that we use. If a word doesn't relate to the world in some way and is never used by any subject then it can have no meaning. Its meaning is both subjective (some subject somewhere has to use it) and objective (if it only existed in the head of the subject it would be a useless tool for either describing the world or interacting with others). At both extremes of the objective/subjective dialectic meaning starts to break down. In between these extremes meaning can be more or less unique to the subject and about the world in general. In most cases the more it is about the world in general the less useful it will be to an individual subject, we normally want to know what things mean in relation to us. An abstract description of the world, though more accurate in its portrayal of how things actually are, may be less useful.

3.2 Relation to IR

The problem of IR can be seen as a product of this tension. Each user is a subject and will have a particular view of their information need and a particular context to their problem. This then has to somehow be manipulated into a query which will locate information from the objective domain. This process is problematic for two reasons: firstly it is necessary to find a balance between using general, or more objective terms, and specific, or more subjective terms in order to get the appropriate level of detail and generality in the results; secondly it is necessary to try and think of terms that others will have used to describe the information you want. The first difficulty is one of finding the most useful position on the objective/subjective dialectic, the second is related to this in that it involves thinking outside the user's subjective view of the problem and analysing how others may have described it.

This is difficult, indeed Blair [9] argues that locating terms that others have used in the documents you want is the central problem in IR. This is because there are so many different ways of describing the same (or similar things). One object (in the case of IR this could be a particular topic) can be written about or indexed in quite a wide range of different ways. This is problem for both automatic indexing in that it relies upon the author's use of terms and for manual indexing in that it relies upon the indexer's choice of terms. Automatic indexing does not provide an objective solution because its raw material is text which must have an author. As symbols, words are always something other than that which they represent and no form of manipulation can erase this quality. Attempts can be made to narrow down the possible meanings of terms through disambiguation though work by Sanderson [10], concluded that longer queries, rather than disambiguation technology were the most effective method of reducing ambiguity. Query expansion can expand on the meaning of the query by adding related terms, often though this can result in the retrieval of documents that are too different from the original query to be relevant.

3.3 Possible Improvements

Firstly, if the objectivity and subjectivity aspect of meaning is a dialectic, a total resolution is not a realistic goal. A text cannot simultaneously be analysed in a way which is both objective and subjective. The decision as to the best position in the dialectic for a particular IR process can be taken on a number of different levels.

3.3.1 Query Level (how to look)

At the query stage it could be made more transparent to the user that they need to balance their need for specific information about their particular information need and their need for general information on their topic. The subjective approach emphasises that all information retrieved must closely relate to the information need, the more objective approach focuses less on the detail and more on the general context of the problem. There is no optimum solution that will work for each information need, it is question of making it clearer which approach is best suited to which situation. Documents could be indexed at different levels of detail with each level representing a different balance between objectivity and subjectivity. An interface could provide the user with clear information about the advantages and disadvantages of each level.

Mizzaro's [11] work on the different aspects of relevance, in particular his concept of abstract characteristics of documents such as quantity (number of documents and their length) and comprehensibility (level of difficulty of documents required) seems to relate quite strongly to the dialectical tension between objectivity and subjectivity. The decision, for example, to have a high level of difficulty in one's results would normally involve an increase in subjectivity and detail with a corresponding loss in a general or objective overview

3.3.2 System Level (where to look)

This can be seen as a decision as to whether a general source should be chosen with a wide range of information, though perhaps not in great depth or a very specific source of information which is much more focused but also narrower in remit. The general information source could be seen as the most objective and the more specialised the more subjective. Medical information provides a useful illustration of this problem, information on a medical condition from the British Medical Journal is likely to be far too technical for the ordinary patient, a more appropriate source is likely to be the web-site of a self-help group. In terms of effective IR it is the inability to make this kind of judgement which often frustrates users and effective signposting is vital. This can take the form of clear explanations of the level of information contained in different databases and pointers towards sources of more or less detail.

4 Individual versus Shared

This dialectic is concerned with language as a communication device i.e. the way in which we use meaning in our relationships with others. It is connected to the tension between objectivity and subjectivity in that objectivity can be seen as the view that is most shared and subjectivity as the view that is most individual.

4.1 Theoretical Problem

The question is to what extent meaning can be said to be something that happens in someone's head and to what extent is meaning a public shared act. How do I know that other people have understood what I mean and why is that sometimes people disagree about meaning. This aspect of the problem of meaning is connected to the problem of epistemology i.e. how do we know things. In this case how can I know what I mean and how can I know what others mean (or indeed that there are others). What is it that we are most certain of, is it my own meaning or is it the fact that most of the time other people seem to use similar meanings to me.

This dilemma can be illuminated by examining the shift from the Cartesian to the Wittgensteinian view of the self. Descartes [12] argued that what we are most certain of is our individual experience of thinking, we are essentially a thinking mind and we deduce our own existence (*cogito ergo sum*) and that of the world and other people from this experience. I know what I mean and I theorise that other people seem to share these meanings most of the time, meaning is at foremost both an individual and intellectual activity. Humans are primarily thinking beings who just happen to inhabit bodies and share lives with other humans.

In a challenge to this Wittgenstein [4, 13] argued that what we are most certain of is not our individual consciousness but our communal life and language. Meaning is not an individual mental act that we then decide also happens to other people, it is rooted in our shared form of life (*lebensform*). Meaning must be rule-based for it to work and there must be a public check that we are using it correctly, for example meaning would break down if we all changed our names every hour in a random fashion. The whole function of a name requires consistency of application, if someone gets your name wrong it would be a bizarre response to change it, rather you would correct them. The reason that meaning works well most of the time is that we share many things in common and it is easy for us to see how others use language. We do not have to deduce meaning in the same way that we do not actually deduce that others exist, we just naturally behave as though they did. In terms of Wittgenstein's philosophy humans are primarily physical creatures with a shared pattern of life, our language and intellectual activities are dependant on this.

It is clear then, that a perspective on the extent to which meaning is individual or shared is connected to quite complex issues about the nature of the individual and our experience. In one sense it is a continuum on which our position constantly shifts. The factors that move us one way or the other are difficult to quantify and sometimes the process will break down and meaning will fail to function in a particular communication or situation.

4.2 Relation to IR

The dialectic between the individual and shared meaning clearly has many different levels in IR as shared could at a minimum be two people and at a maximum absolutely everyone. If it is seen as essentially a problem of communication i.e. how do we use language to convey and receive information, then it can help illuminate two aspects of IR which can also be seen as communication activities.

Firstly the process of querying could be seen as a very clumsy attempt to have a conversation with the documents in the collection. The chances of this being a fruitful conversation could be said to depend on how much the user 'has in common' with the documents or how likely it is that they will have a shared 'view' of meanings. It is also a two way process, though one partner clearly has zero social skills, in that the user can spend time refining their query on the basis of documents retrieved and the indications they give about how information is ordered in the system. This will normally have a very beneficial effect on results. Alternatively it is possible to try and design the system so that it can assist in this process though a computer can only have very limited success in 'understanding' the user.

Secondly this dialectic contributes to clarifying the question of to what extent the user looks for information on their own. In one sense, of course, it is incoherent to discuss somebody looking for information totally on their own, because, even if the user faces the IR system without help, most information needs arise from a situation which to some extent is social. Other people will have shaped the situation which prompted the query. There is, however, still a useful distinction to be made between an individual user searching in isolation and one who uses other people such as colleagues or information professionals to help. IR research often seems focused on enabling the individual user to do without human assistance though the rationale behind this is not clear as recent research shows that professional help does improve results [14]. It may be that collaborative effort before a search increase the chances of successful communication with the system because other people's perspectives have been incorporated. If IR can be seen as an attempt to maximise the possibilities of shared meaning then this would suggest that the involvement of others will normally be helpful.

Finally, an understanding of the tension between the individual and the shared in meaning can help illuminate the problem of what information actually is. The cognitive tradition's [15-17] emphasis on information as the extent to which a person's mental state changes assumes a Cartesian view of the self and mental activity. The idea that what is important about information is how it changes what is inside one person's head seems dependent on the assumption that it is the individual and their own experience of information that is the most important factor in the meaning process. In contrast a Wittgensteinian view of individuals would tend to place more emphasis on the importance of the shared understanding of information and how that can be publicly observed. In this case the traditional test collection methodology with its general approach to relevance judgements could be said to be dependent on a Wittgensteinian rather than a Cartesian perspective.

4.3 Possible Improvements

4.3.1 System/User Matching

Attempts to improve the shared meanings between system and user is difficult and success is likely to depend on the scope of both. It becomes more difficult when the user group is very diverse, Ellis's recent paper describing the problems of IR on the web where the user is unknown highlights a persistent problem in generic information systems [18]. IR researchers who draw on the later work of Wittgenstein [9, 19] argue that shared meaning is a result of shared activity and therefore that IR systems would work much better if they were very focused on a particular subject area and limited to a particular group of users who were connected by shared practice (e.g. a particular profession).

In evaluating this proposal it is important to distinguish two separate factors: firstly there is the argument that specialists will find focused and detailed information sources more useful than general ones; secondly there is the argument that they find them more useful because they share meaning with the system as a result of common shared activity with the authors of the documents and presumably sympathetic system designers. In my view the first argument is clearly correct, the more specialist a user is the more they will require detail at the expense of generality. It is not necessarily the case, though, that this increased usefulness is because of an increase in shared meaning based on shared activity. In fact research shows that specialists are more likely than novices to disagree on relevance judgements, suggesting that shared agreement actually decreases with level of speciality [20, 21]. In terms of natural language processing, however, and indeed most kind of artificial intelligence, attempts to manipulate language in a sophisticated way are much more successful in limited domains than they are in generic systems [22]. This would suggest that research effort should be focused on these limited specialist areas rather than attempts to solve the problem of IR in general.

4.3.2 Improving Collaboration.

Encouraging collaborative information seeking is to a large extent a management or organisational issue rather than one of system design though some systems are more likely to encourage it than others [23]. I would argue that the IR community could usefully challenge its focus on the lone user and its distaste for the traditional model of assistance from informational professionals or others. If IR is a problem rooted in the inherent tensions of meaning then collaboration can improve the chances of clarification, for example a medical librarian is likely

to be able to assist a user in searching through Medline due to their knowledge of the context and use of the information it contains. It is this kind of knowledge that makes shared meaning more likely rather than leaving the novice user to fruitlessly use terms that would never occur in academic medical journals. Attempts to incorporate context, which in the final analysis can be seen as the experience of being a human in the world, into computer systems do not seem to show signs of progressing beyond narrow domains so in the near future a stronger focus on cooperation seems the most realistic method of improvement.

5 Conclusions

In this paper I have attempted to provide clarification of the theoretical tensions in IR and in particular to show that work in meaning and language always comes with some 'metaphysical baggage' about the nature of the individual subject and how we interpret the world. In attempting to progress in IR, it is actually these issues that require resolution, not technological problems. As our relationships to each other and the world operate in a constant tension which is manifested in how our language works, such a resolution is not possible. Incremental improvements are most likely to come from research which focuses on clarification and elucidation rather than a generic solution.

6 References

1. Frege G: On Sense and Reference. 1892. In: Moore A W (ed) Meaning and Reference. OUP, Oxford, 1993, pp 23-42
2. Russell B: Descriptions. 1919. In: Moore A W (ed) Meaning and Reference. OUP, Oxford, 1993, pp 192-207
3. Wittgenstein L. Tractatus Logico Philosophicus. Ogden C K (trans). Routledge, London and NY, 1922 (reprinted 1996)
4. Wittgenstein L. Philosophical Investigations. Anscombe G E M (trans). Basil Blackwell, Oxford, 1958 (reprinted 1988)
5. Eagleton T. Literary Theory: An Introduction. Oxford, Basil Blackwell, 1983 (reprinted 1989)
6. Cleverdon C W, Mills J, Keen E M. Factors Determining the Performance of Indexing Systems, Vol 1. Design, Vol 11. Test Results, Aslib Cranfield Project, College of Aeronautics, Cranfield, 1966
7. Nagel T. The View from Nowhere. OUP, New York, 1986
8. Kant I. Immanuel Kant's Critique of Pure Reason. Smith Kemp N (trans), Macmillan Education Ltd Basingstoke, 1989
9. Blair D C. Language and Representation in Information Retrieval. Elsevier Science Publications, Amsterdam, 1990
10. Sanderson M. Word Sense Disambiguation and Information Retrieval. PhD thesis. University of Glasgow, 1997
11. Mizzarro S. How many relevances in Information Retrieval. Interacting with Computers 1998; 10 (3):305-322
12. Descartes R. Meditation on First Philosophy. In: Wilson M D (ed) The Essential Descartes, Meridian, New York, 1969, pp 154-223
13. Kerr F. Theology after Wittgenstein. Basil Blackwell, Oxford. 1986
14. Jokic M. Analysis of user's searches of CD-ROM databases in the national and university library in Zagreb. Information Processing and Management 1997; 33 (6):785-802
15. Brookes B C. The foundations of information science parts I & II. The Journal of Information Science 1980; 2: 125-133; 209-221
16. Belkin N J. The cognitive view in Information Science. Journal of Information Science 1990; 16: 11-15
17. Ingwersen P. Information Retrieval Interaction. Taylor Graham, London, 1992
18. Ellis D, Ford N, Furner J. In search of the unknown user: indexing, hypertext and the World Wide Web. Journal of Documentation 1998; 54 (1):28-48
19. Brier S. Cybersemiotics: a new interdisciplinary development applied to the problems of knowledge organisation and document retrieval in information science. Journal of Documentation 1996; 52(3):296-344
20. Cuadra C A, Katter R V. Experimental Studies of relevance judgement: final report. System development corporation, Santamonica, CA, 1967
21. Rees A M, Schultz D G. A field experimental approach to the study of relevance assessments in relation to document searching. 2 vols. NSF Contract No C-423, Center for Documentation and Communication Research, School of Library Science, Case Western Reserve University, Cleveland, 1967
22. Smeaton A. Progress in the application of Natural Language Processing to Information Retrieval. The Computer Journal. 1992; 35 (3):268-278
23. Winograd T, Flores F. Understanding Computers and Cognition: a new foundation for design. Ablex Publishing Corporation, New Jersey, 1986