Division of Informatics, University of Edinburgh

Institute for Communicating and Collaborative Systems

Current Directions in Computational Humour

by

Graeme Ritchie

Informatics Research Report EDI-INF-RR-0032

Division of Informatics **December 2000**
http://www.informatics.ed.ac.uk/

Current Directions in Computational Humour

Graeme Ritchie

Informatics Research Report EDI-INF-RR-0032

DIVISION *of* INFORMATICS
Institute for Communicating and Collaborative Systems

December 2000

To appear in "Artificial Intelligence Review"

Abstract :

Humour is a valid subject for research in artificial intelligence as it is one of the more complex of human behaviors. Although philosophers and others have discussed humour for centuries, it is only very recently that computational work has begun in this field, so the state of the art is still rather basic. Much of the research has concentrated on humour expressed verbally, and there has been some emphasis on models based on "incongruity". Actual implementations have involved puns of very limited forms. It is not clear that computerised jokes could enhance user interfaces in the near future, but there is a role for computer modelling in testing symbolic accounts of the structure of humorous texts. A major problem is the need for a humour-processing program to have knowledge of the world, and reasoning abilities.

Keywords : affective computing, artificial intelligence, humour, jokes, puns

Current Directions in Computational Humour*

Graeme Ritchie

Abstract

Humour is a valid subject for research in artificial intelligence, as it is one of the more complex of human behaviours. Although philosophers and others have discussed humour for centuries, it is only very recently that computational work has begun in this field, so the state of the art is still rather basic. Much of the research has concentrated on humour expressed verbally, and there has been some emphasis on models based on "incongruity". Actual implementations have involved puns of very limited forms. It is not clear that computerised jokes could enhance user interfaces in the near future, but there is a role for computer modelling in testing symbolic accounts of the structure of humorous texts. A major problem is the need for a humour-processing program to have knowledge of the world, and reasoning abilities.

1 Introduction

Over the past forty years, research into artificial intelligence has explored many areas of behaviour which would previously have been thought to be uniquely human. Initially, modelling of human activities tended to concentrate on the more well-defined and manageable examples, such as logical reasoning. However, as the discipline has matured and grown in confidence, attention has gradually turned to more and more aspects of "intelligent" or even "creative" behaviour. For example, the past ten years have witnessed a growth in attempts to characterise, in formal symbolic terms, human emotions (e.g. Bates (1994), Frijda and Moffat (1993), Frijda and Moffat (1994)).

One area which is particularly challenging is the modelling of humour. Although the mechanisms of humour have been discussed for thousands of years (see Chapter 1 of Attardo (1994) for a brief review), there has been relatively little work involving rigorous or detailed description of actual humorous mechanisms. At present, humour studies is very much a multi-disciplinary area, with contributions principally from philosophy, psychology, linguistics, sociology and literature, but with only a minimal computational or formal component. This is regrettable, as the techniques and methodology of artificial intelligence are well-suited to developing detailed, symbolic and testable models of something as intricate and multi-facetted as humour.

*An earlier version of this paper was presented at the 7th IEEE International Workshop on Robot and Human Communication, Takamatsu, Japan, October 1998.

Another unfortunate aspect of the relative lack of computational work is that artificial intelligence is (perhaps by accident) omitting an important aspect of human activity from the programme to understand intelligence. If we could develop a full and detailed theory of how humour works, it is highly likely that this would yield interesting insights into human behaviour and thinking. Indeed, it is implausible to suppose that a theory of humour could be developed prior to, and in isolation from, broader theories of human activity. It is not so much that we should build a humour theory and then see how it fits in with intelligence; rather, we can use the attempt to theorise about humour to help widen and deepen our study of human intelligence. There is no reason why we should exclude this one class of behaviour from our data.

As well as this *scientific* motivation for the study of humour, there is also the more practical, *engineering*, perspective, which will be discussed in Section 5 below.

Sceptics sometimes advance various supposed obstacles to devising a detailed theory of humour. These include the culturally dependent nature of much humour, and the fact that different people have different senses of humour. These (uncontroversial) observations merely point to the variety of factors that must be addressed, and the complexity that any general theory must have. They do not render the task impossible in principle, any more than the existence of different dialects prevents the development of a theory of how language works, or the variety of economic systems in the world makes it impossible to abstract economic principles.

2 Background

There is a wealth of literature on cultural issues, such as ethnic humour (Davies, 1990), and considerable research into psychological or physiological aspects of humour (e.g. Zillmann and Cantor (1976), Giles *et al.* (1976), Godkewitsch (1976), Fry (1994), Derks *et al.* (1997)). There is even a strand of research on measuring sense of humour (Ruch, 1996).

Traditional accounts of humour, which would be viewed as highly informal, discursive and anecdotal by practitioners in artificial intelligence, tend to mention a number of recurring themes. These include *incongruity*, *multiple perspectives* (or *ambiguity*), *surprise*, *psychological release* and *aggression*, although terminology may differ. (It appears from Attardo (1994) that many of these ideas can be traced back to Plato and Aristotle).

What is noticeable about these (many and varied) discussions is that they are not all considering the same aspect of humour. For example, those who consider aggression (or *superiority*) as the unifying essence of humour (e.g. Gruner (1997); see Fave *et al.* (1976) for some discussion) have little to say about why certain presentations of an idea are funny, while other presentations with equally aggressive content, are not. Minsky (1986) makes some preliminary remarks about how humour could be viewed from the artificial intelligence/cognitive science perspective, refining Freud's notion that humour is driven by our mental "censors" which control inappropriate thoughts or feelings. Minsky suggests that jokes are

illustrative examples of faulty logic, which enables humans to refine their reasoning censors in a relatively painless way. His remarks are thought-provoking, but say little about the internal structure of actual witticisms. On the other hand, attempts to analyse the internal structure of jokes (e.g. Attardo (1997)) do nothing to set jokes in their social or interpersonal context. Hence several of these proposals could simultaneously be true. It could be that the driving force for humour is hostility, but various verbal devices, such as ambiguity, can be used to that end. There is no reason why all aspects of every example of humour should be explicable in terms of a single principle (see section 0.2.2 of Attardo (1994) for a discussion of this "anti-essentialist" position).

Although there has been discussion of a wide variety of forms of humour (cartoons, slapstick, visual jokes, etc.), most of the work which begins to approach the formality needed for computational work is concerned with 'verbally expressed' humour, i.e. humour which is conveyed by speech or text in a natural language. Such humorous items may not be based on particular words (in the way that puns, for example, could be termed 'verbal humour'), but the content of the humour is expressed in language. (It is perhaps significant that the first International Workshop on Computational Humor (Hulstijn and Nijholt, 1996) chose as its topic "Automatic Interpretation and Generation of Verbal Humor", where this meant verbally *expressed* humour.)

Within this subarea, research has typically focussed on *jokes* rather than longer texts. Within the realm of jokes, there also tends to be a further concentration on what might be termed the "funny story", as opposed to spontaneous witticisms in conversational contexts (although particularly famous humorous remarks have become standard items for analysis, usually where these are relatively self-contained and not heavily dependent on context). There are vast numbers of literary analyses of short stories, plays, films and novels, but these are not normally couched in precise symbol or mathematical terms (although Hobbs (1990) makes a start). Hence the insights from literary theory are not yet in a form where their computational potential can be assessed.

Humour is a vast and complex phenomenon, manifested in a wide variety of forms, and methodologically it is may be helpful to restrict ourselves to some limited subarea if we are to make a start. Verbally expressed humour does encompass a large and varied proportion of humor. For the rest of the discussion here, I will confine attention to such phenomena (textual jokes) and the related research.

One advantage of focussing on verbally expressed humour is that there are at least partial theories or formal frameworks for describing matters of grammar, semantics and (to a lesser extent) pragmatics. That is, we may have some tools and building blocks ready to hand for stating whatever generalisations we decide to make. Moreover, these language models are generally finite symbolic systems, potentially implementable in software. If we were to attempt descriptions of, say, physical or visual humour, the descriptive vocabulary would be less obvious. (The problem might be comparable to the difficulty of notating movements in modern dance.)

As mentioned earlier, research has addressed various aspects of humour, ranging from the logical devices used within jokes to the deeper psychological motiva-

tions for joking. There is something of a consensus amongst a fairly wide range of such scholars about some very general features of verbally expressed jokes. Section A.II of Freud (1966) pointed out the importance to humour of packing two disparate meanings or views into a single text. One of the most influential presentations of this idea in recent years came from Koestler (1970), who defined 'bisociation':

> "... the perceiving of a situation or idea, L, in two self-consistent but habitually incompatible frames of reference, M_1 and M_2."

Koestler also discussed, informally, most of the other common themes, including surprise and emotional release. Hobbs discusses a similar device as underlying effective *poetic* imagery:

> "...two powerful but unrelated images are presented to us individually and we are forced to discover their relation."
> " ... juxtaposition seems to promise coherence and thus impels us to try to construct a coherence." (p.129 of Hobbs (1990))

As poetic imagery is not normally humorous, there must be more to humour than just the relating of otherwise disparate frameworks. Raskin (1985) suggests that the frameworks must be 'opposed' in one of a relatively few modes.

This basic idea of "two ideas compressed" has been developed by many authors into an analysis in which a joke consists broadly of a preliminary part, typically most of the text, which is sometimes called the 'set-up', followed by a (typically much shorter) portion, the 'punchline'. The set-up has some natural or obvious interpretation, and establishes certain expectations or predictions in the audience's mind. The punchline causes some form of conflict, by either forcing another (hitherto unseen) interpretation on the text, or by violating an expectation, or both. (As Deckers and Avery (1994) point out, the social or literary context of telling a joke creates a further expectation of a punchline rather than a logical ending, so that the apparent logical reversal in a joke may be more satisfying than what might appear to be a fulfilled prediction; a logical, "predictable" ending can even be puzzling.) Minsky suggests that this is the most common factor in humour:

> "... a scene is first described from one viewpoint and then suddenly – typically by a single word – one is made to view all the scene-elements in another, quite different, way."(p.10 of Minsky (1980))

Some simple (but not hugely funny) examples of this are the following:

> "I had gone to the south of France to finish my new book... I'm a very slow reader."(Frank Muir)
> "One morning, I shot an elephant in my pyjamas. How he got into my pyjamas, I'll never know." (Groucho Marx)
> "A young lady was talking to the doctor who had operated on her. 'Do you think the scar will show?' she asked. 'That will be entirely up to you," he said. (Quoted in Attardo (1994))

It is commonly suggested that the effect on the hearer of the suddenly revealed meaning is an attempt to "resolve" the apparent incongruity. (See Attardo (1997), Ritchie (1999) for discussion of 'incongruity resolution' approaches.) Notice the similarity to the quest for coherence within poetic interpretation postulated in the quotation above from Hobbs.

Even though these generalisations appear to be innocuously vague, Morreall points out (cited in footnote 1 to Attardo and Raskin (1991)), there are a variety of quite natural humorous texts or remarks which do not conform to this pattern, such as funny rhymes; excessive alliteration; certain verbal slips; tricks with apparent morphemes (*if it's feasible, let's fease it*); pragmatic incongruity; illogical remarks (*if you're going to smoke here, you'll have to either put out your pipe, or go somewhere else*); funny sayings in the style of epigrams (*you can get anywhere in 10 minutes if you go fast enough*). However, it is intuitively clear that there is a wide class of jokes ("funny stories") for which the scheme described above is at least plausible.

3 Linguistic treatments

There has been some research into verbally expressed humour within linguistics. These treatments typically attempt to develop symbolic accounts of the relationships between linguistic units such as words, phrases, meanings, etc. From the computational standpoint, this work is more interesting than traditional literary analysis, but still has a long way to develop before it reaches the level of detail and precision necessary for computational modelling. Although the broad field of linguistics does include some very formal treatments of language, the analyses of humour tend not to be so precisely defined.

Even if we limit our attention to attempts to analyse the structural properties of verbally expressed humour in a way which might lead to a symbolic processing model, there is little consensus on the necessary theoretical constructs, or even on what is the most insightful level at which to analyse humorous texts.

Oaks (1994) offers a catalogue of syntactic and lexical devices for creating ambiguity within jokes. Hetzron (1991) is typical of suggestions which focus on internal structuring, looking at the sequence of presentation of information in a joke. Norrick (1993) examines very broad surface attributes of jokes, and the importance of repetition. Curcó (1996) argues for the importance of a pattern of inferences based on relevance theory (Sperber and Wilson, 1986). Ephratt (1996) also gives an account in terms of pragmatics, but relying more on the notion of speech acts.

The proposals of Raskin (1985) have been highly influential in the study of verbally expressed humour. Raskin's original framework was essentially a formulation of the "incongruity resolution" approach (Section 2 above) using 'scripts', where a script is a structured configuration of knowledge about some stereotyped or familiar situation or activity (cf. the scripts of Schank and Abelson (1977) or the "frames" of Minsky (1975)). This has been developed further, into the General Theory of Verbal Humour (Attardo and Raskin, 1991). There the emphasis is on decomposing the types of knowledge used in the composition of a joke, to show that there is a hierarchy of types of knowledge, ranging from the very abstract and

general (choice of how the scripts are to be opposed) to the more concrete or specific (choice of particular words and phrases). It is claimed that there is empirical support for this stratification of knowledge, as people will tend to judge jokes as more or less similar depending on how far up the hierarchy of knowledge types their differences are located (Ruch *et al.*, 1993). However, it is not clear that the subjects' similarity judgements actually support the GTVH's particular account of what leads to the similarities.

Although the GTVH is more developed than most other linguistic theories of humour (see Attardo (1997) for the latest refinements), it is, from a computational viewpoint, a very early draft of a model (despite the optimistic tone of Raskin (1996)). Many of its basic constructs are not rigorously defined, and criteria for deciding when a particular joke embodies a particular script or a particular opposition are usually left to the analyst's intuitions.

4 Computational approaches

The state of the art in computational models of humour is not highly developed. There are very few implemented systems which process humorous text (and those that exist carry out very simple tasks), and there are only a few simple simulations of humour mechanisms. (Most of those discussed here are represented in the IWCH proceedings (Hulstijn and Nijholt, 1996).)

Utsumi (1996) outlines a logical analysis of irony, but this has not been implemented in any way. Ephratt (1990) has constructed a program which parses a small range of ambiguous sentences, and using some simple heuristics for how syntactic constituents should be arranged, detects an alternative, allegedly humorous reading. This is a rare example of an implemented exploration of the use of *syntactic* ambiguity in jokes, but it is extremely limited, and seems to be aimed solely at (single) sentences whose very ambiguity is deemed humorous, rather than the more elaborate constructions involving set-ups and punchlines. Ephratt's typical example is the sentence *A gold-miner is a person that has strong hands and boxes*, which can be seen as having two readings: ... *that has strong hands and has strong boxes* or ... *that has strong hands and that boxes*, with the latter being claimed to be the non-obvious, humorous interpretation. The data here could be contested.

Takizawa *et al.* (1996) have implemented a pun detecting program for Japanese, which accepts a sequence of phonemic symbols and produces possible analyses of this in terms of sequences of Japanese words, rating each word-sequence with the likelihood that it is a pun, based on various heuristics. Veale and Keane (1996) have a general approach to metaphor and analogy, which they describe in terms of a semantic network model. They claim that this can be used directly to describe various humorous forms, since something like Koestler's "bisociation" can be modelled by a form of matching configurations of nodes in such a network. They do not indicate what has been implemented, and with what results.

The JAPE riddle-generator (Binsted and Ritchie, 1994, 1997) is a program which is capable of producing rather simple punning riddles of the sort enjoyed by young children. Although such jokes do have a natural decomposition into a "set-up" and "punchline", they do not seem to be related in quite the same way as in, for

example, a funny story (Section 2 above). Given a set of symbolic rules about suitable meaning-combinations and textual forms, and also a large (but otherwise conventional) natural language lexicon, it can produce question-answer puns such as:

What do you call a quirky quantifier?
An odd number.

What's the difference between money and and a bottom?
One you spare and bank, the other you bare and spank.

What do you get when you cross a monkey and a peach?
An ape-ricot.

When these jokes were tested by showing them to schoolchildren (Binsted *et al.*, 1997), the better ones were judged to be jokes, to be as funny as some of the jokes found in human-written joke books of a similar genre, and certainly to be very different from various non-jokes included for control purposes.

While this performance is impressive in many ways, it has certain weaknesses when viewed in the broader landscape. The JAPE program has very little in the way of a theory underpinning it. Although there is a clear, implementation-independent statement of how it works (in terms of lexical entries, pattern-matching rules, etc.), these constructs are not tied to any real hypotheses about humour. Furthermore, it is not entirely clear how to generalise from this mini-model to other forms of humour. Although the authors, in a later paper (Binsted and Ritchie, 1996) draw a broad analogy between the structure of riddles and that of story puns, both these genres are rather specialised, and the similarities are at a very abstract level.

Katz (Katz, 1993, 1996) has proposed a neural account of what happens when a humorous stimulus such as a joke is processed by a hearer/reader. The set-up gives rise to an expectation that there will be a particular ending, so a neural unit (or units) corresponding to that expected ending are activated. The actual ending (punchline), differing from the prediction, activates a different neural unit. However, because this new activation happens so suddenly, and because the already established activation is supported by the earlier context, there is a phase in which both units are active, thus producing a particularly high level of activation. Katz argues that it is this transient surge in activation levels which produces the humorous (and usually pleasurable) effect. He points out that this neural description is consistent with the observations that the humorous effect of a joke can be improved by increasing the strength of the prediction (e.g. by having a particularly convincing set-up) or by increasing the plausibility of the relationship between the actual ending and the original set-up.

This theory is not necessarily in conflict with typical symbolic theories of humour, and indeed is supportive of the incongruity resolution account (see Section 2 above). Rather, it works at a different level of description, providing an implementation in neural constructs of some widespread observations about the way that verbally-presented jokes work. Its main weakness is that although Katz has carried out detailed checks on his model at the neural level, there are still no

intervening constructs to connnect his neurally-stated hypotheses with actual humorous stimuli: he does not feed jokes into the computer to test it. Also, he does not explain how humour differs from other stimuli which might produce sudden brief high activation levels. In a sense, he has offered an account of why jokes are stimulating, without explaining why they are funny.

5 Practical uses

What might computational models of humour be used for? If we had a workable model of humour, there are two obvious ways of implementing it: as a humour generator, or as a humour understander.

A computer generator of jokes could be of use to those who need a steady supply of jokes, particularly if quality was less important than quantity (e.g. those firms who, in Britain at least, insert riddles or other simple jokes into Christmas crackers). At present, computer systems are not in a position to outperform human joke-creators, but for very simple jokes aimed at children, it is conceivable that such a system could be workable in a few years (cf. the JAPE system, Section 4 above). Writing more complex jokes, such as comedy sketches, is much further away.

A joke-appreciating program is of less obvious use. It would have to be remarkably good before we would consider using it as a form of quality control for any human joke writing.

Let us consider both joke-generation and joke-appreciation from an HCI standpoint. Could computationally tractable theories of humour help in the building of better user-interfaces or more usable systems such as robots? If we are to cooperate with robots at work, or have intelligent agents as our constant advisors, perhaps some humour would make interactions more pleasant.

Binsted (1995) has argued that a computer system could be made more congenial by judicious use of humour generation in a user-interface. She suggests various situations in which a humorous remark from the system could ease the interactions – errors, poor system performance, offering hints, requests for clarification – and speculates that certain styles of humour would be appropriate – self-deprecatory remarks, observations about the situation, use of amusing phrases. However, she acknowledges that there are difficulties in such attempts, since humour (particularly imperfectly created examples) can easily irritate rather than relax the user. Stock (1996) also argues for the desirability of injecting humour into various sorts of computer applications, including education and entertainment, although he does not offer detailed proposals for how this would work. Binsted (personal communication) has also suggested that educational software for children could be based round the idea of machine-assisted joke generation, thus helping the user to explore word-meanings. This approach is advocated by McKay (2000), who has implemented a Prolog program ("WisCraic") capable of producing puns such as:

The obliging gardener had thyme for the woman.

The cruel deer-keeper broke the woman's hart.

The insolvent baker kneaded dough.

The program can also provide an outline of the semantic and phonetic associations which underly any one of its jokes, and McKay suggests that a system along these lines could help a second-language learner to explore and understand particular usages, particularly idioms. Takizawa *et al.* (1996) comment that "Kitagaki developed a pun generator to be applied to a human-friendly Japanese word processor", citing (Kitagaki, 1990, 1993). Loehr (1996) carried out a preliminary experiment in putting a joke-generating module (the JAPE program) within a user interface (Elmo), concluding that it is difficult to arrange automatic generation of a joke which is *relevant* to what the user is trying to do.

In considering systems like Loehr's Elmo, it is not clear that the behaviour crucially relies on jokes being computer-generated, as opposed to merely accessed online. Unless the process of generation is subtly influenced by the context, in particular the user's input, then the interface might as well produce jokes from a stored list (either computer-generated or taken from some other source). That is, if the "relevance" of a computer-generated joke is measured in some crude way such as keyword matching (as in Elmo), there are no benefits (and possibly some disadvantages) in having the joke generated on the spot rather than retrieved from some suitably cross-indexed database of jokes. A perfectly intelligent joke generator would change this argument, as it would be able to create new jokes which were directly pertinent to the user's situation, in a way that could not be simulated by a pre-computed list. We are a long way from knowing how to design such a generator.

There seem to be no proposals to enhance user-interfaces by allowing them to understand jokes by the human user, although this could fit into the more ambitious speculations about using intelligent agents as "companions" for the human in everyday activities. As with many advanced facilities, an imperfect version (which is all that will be achievable in the near future) is likely to be worse than none.

It will probably be some time before we develop a sufficient understanding of humour, and of human behaviour, to permit even limited form of jokes to lubricate the human-computer interface. The goal of creating a robot that is sufficiently "human" to use humour in a way that makes sense or appears amusing (other than inadvertently) is a long term one.

6 Methodology

The field of computational modelling of humour is so new and undeveloped that it is not even clear that it has any well-defined methodologies. For practical attempts at user-interfacing, as described in Section 5 above, evaluation should at least be relatively straightforward. Any interface which has supposedly been enhanced with humour could be compared in some controlled way with a non-humorous interface, and rated for ease of use, etc. In an engineering context, if the enhancements achieve their desired aim, they are a success.

For theoretical work which seeks to gain insights into the nature of humour, it is less clear how research can be organised and evaluated. At least for verbally expressed humour, some inspiration comes from the generative (Chomskyan) approach to linguistic theory, which has been dominant for forty years or so. As in the scientific study of language, the researchers must refuse to be daunted by the fact that the object of study is both highly complex and crucially "human"; they must step back from the everyday nature of the data and ask what mechanisms might generate it; and they must constantly check their generalisations against simple examples. In generative grammar, the researcher posits symbolic rules which would give rise to well-formed sentences. Then a comparison is made with the range of actual sentences in the human language being described, to check the correctness of the rules. Computationally, this can be implemented by a "grammar-tester" program, which allows the user to apply complex sets of rules in order to see their effect in terms of sentences (e.g. Friedman (1971), Friedman (1972)). Similarly, we could posit symbolic rules for jokes, and consider whether the predicted outputs from a joke-generator based on these rules are indeed funny. Whereas linguists were able to be relatively informal about assessing the grammaticality of their outputs, funniness is perhaps a more slippery concept, meriting more controlled and independent assessment. It is important to distinguish between building a joke-generator simply to test out a formalised theory, and attempting to create an agent which "can tell jokes". The former need only exercise a set of rules to see what jokes are generated, as in a grammar tester; the latter should produce appropriate jokes in a way which could be argued to exemplify some form of intelligence. The JAPE program (Section 4 above) is a rule-tester, not a prototype intelligent agent.

There is a sense in which the JAPE project is a miniature exemplification of the methodology outlined above. Symbolic rules were defined, with a specified formal interpretation, and an implemented joke-generator was built to test them; the ratings of funniness were made independently and scientifically.

An alternative method, instead of building a joke-generator, would be to create a joke-understander, a program which takes as input a joke and gives as output a rating of its funniness. For some theories of humour, this might be much more appropriate. For example, it is quite common for informal analyses of joke mechanisms to be phrased in dynamic terms, setting out what happens as the set-up and then the punchline are conveyed. If such an account was formalised in terms of processing steps, a joke-understander would be the natural implementation to test the model.

Where the theory is stated more statically or declaratively, and simply specifies particular relations which must hold between symbolic entities, then a generator and an understander would be equally natural test computations. However, the understander may in practice be more awkward. Leaving aside the need for various supporting facilities (such as a natural language parser), there is the issue of creating suitable test-data. The ultimate goal of a theory of humour may be to isolate factors which are both necessary and sufficient for something to be humorous (cf. comments in Section 2 above). Nevertheless, until the ideal theory is constructed in complete detail, we will always be working with partial or im-

perfect theories. We shall be positing rules which cover some (often small) subset of humorous phenomena. In such a situation, we will be claiming that our rules are *sufficient* for humour, but not *necessary*. To test such a hypothesis, a joke-generator is convenient, since all its output conforms to the rules, and we can use humans as joke-evaluators (a task they are well-suited to). If every output item is funny, our claim of sufficiency is met; every unfunny output counts as evidence against the sufficiency claim. On the other hand, how are we to assess the verdicts of a joke-understander? If it classes a funny input as unfunny on the grounds that the item does not conform to its rules, all we can conclude is that our rules do not cover all possible jokes; but we knew that already. If it classed an unfunny input as funny, this would refute the sufficiency of the rules, but this suggests quite intricate testing in which we have to find potential inputs which are not jokes (according to humans) but which might be characterised by the system's rules as being funny,

Until we require a joke-understander in its own right (for example, for user-interfaces, cf. Section 5), a joke-generator will be a more directly manageable test mechanism for declaratively stated symbolic rule systems.

7 World Knowledge, intelligence and the lexicon

One of the greatest challenges facing any attempt to automate the production or interpretation of humorous material is that human use of humour is typically built on a vast foundation of knowledge about the world, including not only facts but patterns of reasoning. If a computational model, no matter how unfaithful it might be to the details of the human mind, is to emulate such performance, it may well have to embody encyclopaedic knowledge coupled with sophisticated reasoning powers. As Raskin and Attardo (1994) and Raskin (1996) observe, if a system is to *understand* verbally conveyed jokes, it will require a powerful natural language processing system as one of its parts. This might seem to imply that all attempts at actual implementation will have to be postponed, pending solution of the central problems of artificial intelligence. However, it depends on the scope of the humorous phenomena that are being modelled. It is certainly true that a *universal* joke interpreter (or generator) would have to be dauntingly knowledgeable and intelligent. A more limited joke processing system could nevertheless survive with weaker "intelligence". Most of the pun-processing programs mentioned in Sections 4 and 5 above were examples of this less ambitious aim. For example, the very simple type of riddle created by the JAPE program involves very little real world knowledge, beyond the listing of simple properties such as are found in an ordinary lexicon. The JAPE program was deliberately tested with a general-purpose lexicon (Miller *et al.*, 1990) to avoid the accusation that the humorous capabilities had been smuggled in via a special-purpose lexicon; this restriction was necessary to support the claims being made for JAPE's rules. In a situation where scientific control was of less importance, it would be possible deliberately to enhance the lexical resources so as to render the creation of jokes more likely. In particular, special-purpose dictionaries based on particular topics (sport, politics, sex, etc.) could be employed to ensure the production of jokes concerning

that subject.

This issue still stands as a formidable obstacle. Working programs will be possible, for the present, only where the designer can isolate some limited style of humorous phenomenon which is manifested in some manageable medium (e.g. text), and in which the regularities appear to depend on relatively simple knowledge that can be coded up in some tractable fashion, ideally using existing resources (such as lexical databases).

This does not mean that computationally-oriented theoretical investigations must be abandoned. We can still contemplate the development of general abstract models of humour, including proposals regarding the processing of humorous items, with the eventual goal of computational implementation. What is more difficult at present is to produce concrete implementations which are useful or which yield interesting insights.

8 The way ahead

One of the first tasks we must take on is the articulation of some formal notions which would allow embryonic theories, or even particular insights, to be stated more rigorously. This should improve the clarity of proposals, and allow ideas to be compared without the distraction of terminological differences. It would be naive to hope for a universal representational framework suitable for describing not only all jokes but all potential theories of jokes. More realistically, it would be helpful if individual theories were stated formally, and theorists attempted to use the basic concepts of others wherever possible.

Also, we have to tackle the issue of empirical evidence. For a theory to be fully tested, it must at least make falsifiable predictions. At present, few theories of humour make even the semi-formal predictions that are customary in linguistics, where hypotheses can be checked against relevant sentence types. We are even further from being able to devise strictly rigorous experimental tests of theories of humour.

The overall message is that endeavouring to develop computational models of humour is a worthwhile enterprise both for artificial intelligence and for those interested in humour, but we are starting from a very meagre foundation, and the challenges are significant.

Acknowledgements: I would like to thank Professor Seiji Hata for making it possible for me to attend the ROMAN-98 International Workshop where an earlier version of this paper was first presented, and Professor Akifumi Tokosumi for further help. Thanks are due also to Salvatore Attardo, Kim Binsted and Dave Moffat for their comments on an earlier draft.

References

Attardo, S. (1994). *Linguistic Theories of Humour*. Mouton de Gruyter, Berlin.

Attardo, S. (1997). The semantic foundations of cognitive theories of humor. *HUMOR*, **4**(10), 395–420.

Attardo, S. and Raskin, V. (1991). Script theory revis(it)ed: joke similarity and joke representation model. *HUMOR*, **4**(3), 293–347.

Bates, J. (1994). The Role of Emotion in Believable Agents. Technical Report CMU-CS-94-136, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Binsted, K. (1995). Using humour to make natural language interfaces more friendly. In H. Kitano, editor, *Proceedings of the IJCAI workshop on AI and Entertainment*.

Binsted, K. and Ritchie, G. (1994). An implemented model of punning riddles. In *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*, Seattle, USA.

Binsted, K. and Ritchie, G. (1996). Speculations on story puns. In Hulstijn and Nijholt (1996), pages 151–159.

Binsted, K. and Ritchie, G. (1997). Computational rules for generating punning riddles. *HUMOR*, **10**(1), 25–76.

Binsted, K., Pain, H., and Ritchie, G. (1997). Children's evaluation of computer-generated punning riddles. *Pragmatics and Cognition*, **5**(2), 309–358.

Chapman, A. J. and Foot, H. C., editors (1976). *Humour and Laughter : Theory, Research and Applications*. Transaction Publishers, London, first edition.

Curcó, C. (1996). Relevance theory and humorous interpretations. In Hulstijn and Nijholt (1996), pages 53–68.

Davies, C. (1990). *Ethnic Humour around the World*. Indiana University Press, Bloomington, Indiana.

Deckers, L. and Avery, P. (1994). Altered joke endings and a joke structure schema. *HUMOR*, **7**(4), 313–321.

Derks, P., Gillikin, L. S., Bartolome-Rull, D. S., and Bogart, E. H. (1997). Laughter and electroencephalographic activity. *HUMOR*, **10**(3), 285–300.

Ephratt, M. (1990). What's in a joke. In M. Golumbic, editor, *Advances in AI: Natural Language and Knowledge Based Systems*, pages 43–74. Springer Verlag.

Ephratt, M. (1996). More on humor act: What sort of speech act is the joke? In Hulstijn and Nijholt (1996), pages 189–197.

Fave, L. L., Haddad, J., and Maesen, W. A. (1976). Superiority, Enhanced Self-Esteem, and Perceived Incongruity Humour Theory. In Chapman and Foot (1976), chapter 4, pages 63–91.

Freud, S. (1966). *Jokes and their relation to the unconscious*. Routledge & Kegan Paul, London. First published 1905.

Friedman, J. (1971). *A Mathematical Model of Transformational Grammar*. American Elsevier, New York.

Friedman, J. (1972). Mathematical and computational models of transformational grammar. In B. Meltzer and D. Michie, editors, *Machine Intelligence 7*, pages 293–306. Edinburgh University Press, Edinburgh.

Frijda, N. H. and Moffat, D. (1993). A model of emotions and emotion communication. In *Proceedings of RO-MAN 93: 2nd IEEE International Workshop on Robot and Human Communication*, pages 29–34.

Frijda, N. H. and Moffat, D. (1994). Modelling emotion. *Cognitive Studies*, **1**(2), 5–15.

Fry, W. F. (1994). The biology of humor. *HUMOR*, **7**(2), 111–126.

Giles, H., Bourhis, R. Y., Gadfield, N. J., Davies, G. J., and Davies, A. P. (1976). Cognitive Aspects of Humour in Social Interaction: A Model and Some Linguistic Data. In Chapman and Foot (1976), chapter 7, pages 139–154.

Godkewitsch, M. (1976). Physiological and Verbal Indices of Arousal in Rated Humour. In Chapman and Foot (1976), chapter 6, pages 117–138.

Gruner, C. (1997). *The Game of Humor*. Transaction Publishers, New Brunswick, N.J.

Hetzron, R. (1991). On the structure of punchlines. *HUMOR*, **4**(1), 61–108.

Hobbs, J. (1990). *Literature and Cognition*. Number 21 in Lecture Notes. Centre for the Study of Language and Information, Stanford, California.

Hulstijn, J. and Nijholt, A., editors (1996). *Proceedings of the International Workshop on Computational Humor*, number 12 in Twente Workshops on Language Technology, Enschede, Netherlands. University of Twente.

Katz, B. (1993). A neural resolution of the incongruity and incongruity-resolution theories of humour. *Connection Science*, **5**, 59–75.

Katz, B. (1996). A neural invariant of humour. In Hulstijn and Nijholt (1996), pages 103–109.

Kitagaki, I. (1990). A Fuzzy Determination Method of Generating a Laugh and Popularity/Inferiority: "Students' Holiday" and "the Lowest in Running". *Journal of Japan Society for Fuzzy Theory and Systems*, **2**(1), 100–104. In Japanese.

Kitagaki, I. (1993). Extraction of identity on pronounciation concerning play-on-word and evaluation of a tentative software: Aiming a wordprocessor of human friendliness. Technical Report HC92-65, Institute of Electronics, Information and Communication Engineers of Japan. In Japanese.

Koestler, A. (1970). *The Act of Creation*. Pan Books, London. First published 1964, Hutchinson & Co.

Loehr, D. (1996). An integration of a pun generator with a natural language robot. In Hulstijn and Nijholt (1996), pages 161–172.

McKay, J. (2000). *Generation of Idiom-Based Witticisms to aid Second Language Learning*. Master's thesis, Division of Informatics, University of Edinburgh, Edinburgh, Scotland.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K., and Tengi, R. (1990). Five papers on WordNet. *International Journal of Lexicography*, **3**(4). Revised March 1993.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston, editor, *The Psychology of Computer Vision*, pages 211–277. McGraw-Hill, New York.

Minsky, M. (1980). Jokes and the logic of the cognitive unconscious. AI Memo 603, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Cambridge, Mass.

Minsky, M. (1986). *The society of mind*. Heinemann, London.

Norrick, N. R. (1993). Repetition in canned jokes and spontaneous conversational joking. *HUMOR*, **6**(4), 385–402.

Oaks, D. D. (1994). Creating structural ambiguities in humor: getting English grammar to cooperate. *HUMOR*, **7**(4), 377–401.

Raskin, V. (1985). *Semantic Mechanisms of Humour*. Reidel, Dordrecht.

Raskin, V. (1996). Computer implementation of the general theory of verbal humor. In Hulstijn and Nijholt (1996), pages 9–20.

Raskin, V. and Attardo, S. (1994). Non-literalness and non-bona-fide in language: Approaches to formal and computational treatments of humor. *Pragmatics and Cognition*, **2**(1), 31–69.

Ritchie, G. (1999). Developing the incongruity-resolution theory. In *Proceedings of the AISB Symposium on Creative Language: Stories and Humour*, pages 78–85, Edinburgh, Scotland.

Ruch, W., editor (1996). *Special Issue: Measurement Approaches to Sense of Humor*. HUMOR, 9 (3/4).

Ruch, W., Attardo, S., and Raskin, V. (1993). Toward an empirical verification of the general theory of verbal humour. *HUMOR*, **6**(2), 123–136.

Schank, R. and Abelson, R. (1977). *Scripts, plans, goals and understanding*. Lawrence Erlbaum, Hillsdale, N.J.

Sperber, D. and Wilson, D. (1986). *Relevance: communication and cognition*. Blackwell, Oxford.

Stock, O. (1996). 'Password Swordfish': Verbal humour in the interface. In Hulstijn and Nijholt (1996), pages 1–8.

Takizawa, O., amd Akira Ito, M. Y., and Isahara, H. (1996). On computational processing of rhetorical expressions – puns, ironies and tautologies. In Hulstijn and Nijholt (1996), pages 39–52.

Utsumi, A. (1996). Implicit display theory of verbal irony: Towards a computational model of irony. In Hulstijn and Nijholt (1996), pages 29–38.

Veale, T. and Keane, M. (1996). Bad vibes: Catastrophes of goal activation in the appreciation of disparagement humour and general poor taste. In Hulstijn and Nijholt (1996), pages 133–150.

Zillmann, D. and Cantor, J. R. (1976). A Disposition Theory of Humour and Mirth. In Chapman and Foot (1976), chapter 5, pages 93–115.