Division of Informatics, University of Edinburgh

Institute for Communicating and Collaborative Systems

Describing Verbally Expressed Humour

by

Graeme Ritchie

Informatics Research Report EDI-INF-RR-0012

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

Describing Verbally Expressed Humour

Graeme Ritchie

Informatics Research Report EDI-INF-RR-0012

DIVISION *of* INFORMATICS
Institute for Communicating and Collaborative Systems

April 2000

In Proceedings of AISB Symposium on Creative and Cultural Aspects and Applications of AI and Cognitive Science, Birmingham, UK.

Abstract : In pursuit of the long-term goal of developing a general theory of humour, it is reasonable to study certain limited forms of humorous artefact in detail. One obvious class of humour to consider is verbally expressed humour, and in particular jokes. We propose a methodology for exploring this subarea. The central idea is to devise detailed symbolic descriptions of the internal linguistic structure of classes of jokes, at a suitable level of abstraction. These descriptions are intended to make explicit the semantic and pragmatic factors (broadly interpreted) that are relevant to the humorous effect of the subclass of joke in question, and also to contribute an accumulation of analysed data over which more general theorising may occur. An analogy is drawn with established practice in linguistics.

Keywords : jokes, linguistic description, verbal humour, humour theory

Describing Verbally Expressed Humour

Graeme Ritchie
Division of Informatics
University of Edinburgh
80 South Bridge
Edinburgh
Scotland EH1 1HN

g.d.ritchie@ed.ac.uk

Abstract

In pursuit of the long-term goal of developing a general theory of humour, it is reasonable to study certain limited forms of humorous artefact in detail. One obvious class of humour to consider is verbally expressed humour, and in particular jokes. We propose a methodology for exploring this subarea. The central idea is to devise detailed symbolic descriptions of the internal linguistic structure of classes of jokes, at a suitable level of abstraction. These descriptions are intended to make explicit the semantic and pragmatic factors (broadly interpreted) that are relevant to the humorous effect of the subclass of joke in question, and also to contribute an accumulation of analysed data over which more general theorising may occur. An analogy is drawn with established practice in linguistics.

1 Motivation

The ability to comprehend, appreciate and produce humorous artefacts such as jokes is central to human culture and social interaction, and hence the area of humour (including both humorous activities and the artefacts involved) merits scientific study. Exploring humorous activity and objects in rigorous detail may throw light on (and interact with the study of) a variety of aspects of human behaviour, such as cognition, physiology, social conventions, means of communication. Artificial intelligence is well placed to pursue such an enquiry, as its methodologies and techniques have been developed to assist with the detailed symbolic modelling of complex human behaviour. The work here is an attempt to lay the foundations of such an investigation.

Although there is no accepted theory of humour, there have been numerous observations and proposals regarding the nature of humour, and these are often clustered into a tripartite division of *incongruity*, *superiority*, and *relief* "theories" (Raskin, 1985). Attardo (1994) generalises these labels to *cognitive*, *social*, and *psychoanalytical* (Figure 1). Another perspective would be to say that cognitive/incongruity approaches concentrate on the humorous *stimulus*, social/hostility approaches consider the *interpersonal* effects, and psychoanalytical/relief proposals emphasise the *audience's reaction*. All of these are interesting and valid, but distinct, aspects of the phenomenon of humour. If we are develop a complete theory of humour and its use, all of these facets must be considered. In particular, it will be necessary to have a good account of *the types of humorous stimulus that exist, how they are structured and how they function*. It is this question – the nature of the stimulus – that is addressed here. This does not constitute a complete theory of humour, but it is certainly a necessary step towards a full investigation, since it would be difficult to seek empirical support for a theory of humour use without some properly detailed analysis of the data. If we are, for example, to find correlations between types of stimuli and human reactions, then we require an account of the stimuli, to structure and guide our experimentation. The work here makes a start on dissecting the humorous stimuli, in one particular subarea of humour, namely *verbally expressed* humour. The term *verbal humour* (Raskin, 1985) is avoided here, as it is sometimes used in a narrower sense, roughly denoting plays on words (Attardo, 1994).

Cognitive	Social	Psychoanalytical
Incongruity	Hostility	Release
Contrast	Aggression	Sublimation
	Superiority	Liberation
	Triumph	Economy
	Derision	
	Disparagement	

Figure 1: Types of "theory" (Attardo, 1994)

That is, the focus is restricted to humour conveyed in language, as opposed to physical or visual humour, but not necessarily playing on the form of the language.

This restriction makes the task slightly more manageable, while still leaving a wide and rich range of phenomena to be considered.

A further simplification is choosing to study individual *jokes* in isolation from the context of use or the speaker/hearer involved. (Jokes could be loosely defined as short texts deliberately designed to elicit humorous response, often in a manner unrelated to a specific context; however, nothing here depends on having an initial definition of jokes in general.) Starting by tackling jokes out of context does not embody a claim that humorous effects are not dependent on factors such as context, personal opinions, and culture. Rather, it is an attempt to make some progress by (at least initially) not attempting to study simultaneously all the factors involved in a complex phenomenon. There is also a deliberate claim here that there will be regularities in the stimuli involved in humour (the joke texts) which should be documented and described before we can proceed to correlate these with anything else, or to devise more elaborate hypotheses about the whole humour mechanism. This is again analogous to descriptive linguistics, in which there has been a great deal of effort devoted to analysing the structural properties of *sentences*, which are in a sense the counterpart of jokes here. The hope is that if we can develop a good account of how jokes operate, then we can proceed to apply a similar methodology to other forms of verbally expressed humour.

It is important to note at the outset that this paper does *not* propose a theory of humour, not even of verbally expressed humour, nor even of jokes. What it does is to outline a methodology for approaching the construction of a theory of jokes, thence a theory of verbally expressed humour, and eventually a theory of humour. Of course, any methodological approach implicitly embodies some theoretical hypotheses, but the assumptions adopted here are relatively minimal (see Section 12 below).

2 Description and theory

If we are to develop a general theory of verbally expressed humour, it must be based on data. However, it is not feasible to proceed from raw data (e.g. a large and unanalysed collection of jokes) to a complete theory in one step. Some preliminary analysis of the data is required first. An analogy can be made here with generative linguistics, where (within the Chomskyan paradigm at least) there is a quest for a theory of universal grammar. Linguists rely, in their theory development, not on undifferentiated and unannotated data (sentences). Instead, universal theories build on previous analyses of language: fragments of grammar, or comparisons of particular constructions across languages. A large amount of pre-theoretical sorting out and dissecting of the data must occur before proceeding to highly abstract and over-arching theories of language. The position taken here is that humour research can usefully proceed in an analogous fashion (Figure 2). Before we can construct a genuinely empirical general theory of (verbally expressed) humour, we must carry out a significant amount of groundwork which involves analysing our primary data. Below, we make some suggestions about how such descriptive work could proceed, with some illustrative analyses of simple jokes.

Linguistics	**Humour research**
Strings	Grammatical texts
Sentences	Jokes
Grammaticality	Being a joke
Sentence type	Subclass of jokes
Structural description	Description of a joke
Grammar rule(s)	Pattern for class of joke
Theory of grammar	Theory of jokes
Theory of language use	Theory of humour use

Figure 2: Analogy with linguistics

3 What is an analysis?

The overall idea behind the framework here is simple: in order to make clear and explicit the various factors that contribute (or might contribute) towards the humorous effect of a piece of text, one should specify in some detail the various abstract objects that are posited as underlying the texts (e.g. symbolic representations of meaning), the various properties that these objects have (e.g. denoting a taboo subject) and the various interrelations which hold between them (e.g. one meaning being more obvious than another, one word sounding similar to another). An analysis of a joke is then a precise listing of this information for the joke, at a suitable level of abstraction. This last point is important: we have to rely (at least at this stage of methodological development) on the intuition and judgement of the analyst to propose abstract entities which are relevant to the humorous effect, while ignoring irrelevant details. For the moment, the assessment of the suitability of the components of the joke analysis will be left to the informed opinion of other humour researchers, but in time we should evolve a more sophisticated methodology in which independent criteria can be brought to bear. Appealing again to the analogy with generative linguistics, when a syntactic rule is posited for a particular class of sentences, there are various non-subjective criteria that can be applied to argue for or against the adequacy or elegance of the rule. In describing verbally expressed humour, analogous criteria must be developed, for example by arguing that the joke analysis offered will delineate some natural class of jokes (cf. Ruch et al. (1993)). Classes of joke can then be characterised by abstracting from individual descriptions to form more general patterns, in a manner analogous to creating grammar rules which define the structure of classes of sentences.

4 Basic objects

We shall start from a few basic data types, and build from there, introducing new primitives only as particular types of joke seem to demand them.

Most discussions of jokes do not make it explicit what their assumed primitive alphabet is. Since jokes are conveyed sometimes in speech and sometimes in writing, either phonetics or orthography could be chosen. For many jokes, the choice is immaterial. For some jokes, the spoken delivery is essential in order to create some form of ambiguity. Only very rarely is it necessary to use a written form in order to create the desired effect. In general, it is up to the analyst to define which alphabet is to be the formal representation for a joke, but here we shall assume that there is such an alphabet, and that any string over that alphabet constitutes a *text*. That is, we will use the technical term "text" to cover any sequence of symbols, whether a complete joke (or well-formed sentence) or not. We shall also assume that there is a *similarity measure* which indicates how similar two texts are, in some primitive sense. (identity of texts will simply be identity within the set of strings over the alphabet). This will be useful in analysing jokes involving puns or ambiguity. It will then be possible to define various kinds of near-equality relation between texts, based on degrees of similarity (cf. the "paraphony" and "hahaphony" of Dienhart (1999)).

Our initial set of object-types is then:

(a) ALPHABET: a set of basic symbols from which jokes are (at the simplest level) made up; the analyst should make it clear whether written or spoken symbols are intended.

(b) TEXT: a TEXT is a sequence of elements from the chosen alphabet. Hence any substring of a TEXT is a TEXT. There is a *similarity measure* which indicates how similar two TEXTs are.

(c) MEANING: a MEANING is what might be termed the *literal meaning* or *semantic structure* of a TEXT. It takes no account of any inference or contextual information which might flesh out or interpret the meaning of the actual words used.

(d) INTERPRETATION: A TEXT may also have an associated INTERPRETATION, which will be dependent upon (but not identical to) the MEANING of the TEXT (or its parts). It can be thought of as an interrelated and consistent set of propositions, with more content than the bare MEANINGs. This is intended to be a broader kind of meaning (of a passage of text or of some sequence of events, for example), which may involve much reasoning, filling in of implicit information, unwarranted addition of assumptions, etc.

(e) DESCRIPTION: This is a semantic structure which encodes some attributes which could be true of an entity; it can *describe* a MEANING.

These classes of object are primitive in the sense that they are defined solely in terms of the relationships they enter into with other primitive objects. MEANINGs, INTERPRETATIONs and DESCRIPTIONs are all SEMANTIC ITEMS, sometimes abbreviated below to "*SemItem*". All of the above entities are linguistic or abstract representational forms. There will also be entities denoting objects or situations within some world (real or imaginary); see Section 10 for two simple examples.

5 Properties and relations

In addition to some directly linguistic relationships between our basic objects, there will be a large set of attributes which are relevant to recording the humorous mechanisms within a joke. Thus our conceptual repertoire will range from the relatively straightforward (e.g. one TEXT is a substring of another), through conventional linguistic notions (e.g. a TEXT may have zero or more MEANINGs), to quite difficult and non-trivial properties (e.g. a MEANING is absurd, or an INTERPRETATION conveys a taboo idea). This paper will *not* attempt to give full and precise definitions of these predicates, although we will provide informal glosses of those which we use, in order to make the examples intelligible. Supplying detailed definitions for all the predicates involved in descriptions of jokes constitutes a central and substantive part of developing a theory of jokes, and thence a theory of humour (cf. (Ritchie, 1999, Section 4.5)). The first step in our methodology is to postulate a range of these constructs, and see if we can account for joke structure by using them consistently. This decomposes the research into stages, with the final definitions of these conceptual building blocks being postponed until we have an idea of the set of primitives that we need.

Two illustrative and typical examples might be as follows:

$absurd(\langle SemItem \rangle)$: This is true if $\langle SemItem \rangle$, a SEMANTIC ITEM, is in some way odd or bizarre.

$conflicts(\langle Meaning \rangle, \langle SemItem \rangle)$: $\langle Meaning \rangle$ will not merge with $\langle SemItem \rangle$ to form a coherent INTERPRETATION.

6 Structural descriptions

To set out a description of the linguistic content of a joke, we need to state exactly what abstract objects we are positing and what the relationships are between them. Many authors have remarked on the way in which certain jokes use the final line (punchline) to reveal an unexpected meaning for the initial text, in a way that implies, evokes, or describes an image that is odd in some way ((Raskin, 1985), (Attardo, 1994, Chapter 2), (Ritchie, 1999)). To illustrate the descriptive approach here, we can borrow

this informal idea and state it in our terminology. (Notice that the examples in this paper are mostly chosen for their brevity and simplicity rather than the excellence of their wit.)

(1) Why do birds fly south in winter?
It's too far to walk.

The relevant workings of this joke could be summarised thus:

> There is a TEXT T = "Why do birds fly south in winter? It's too far to walk." There are subsequences T_1 = "Why do birds fly south in winter?", T_2 = "It's too far to walk", and MEANINGS M_1, M_2, M_3 such that M_1 is a more **obvious** MEANING than M_2 for T_1, M_3 is the MEANING of T_2, M_3 **conflicts** with M_1 (or perhaps an INTERPRETATION derived from M_1), M_3 is **compatible** with M_2, and there is an INTERPRETATION I of M_2+M_3 which is **absurd**.

(where words in **bold** font indicate properties or relations which this description relies on).

7 Structural patterns

The above summary describes the content of one particular example joke. However, a necessary next step is to abstract from such itemisations to create more general patterns, which will describe classes of jokes which are similar in their internal workings. Here we will call these *structural patterns*. For the sake of a simple notation, we will adopt a sorted version of first order predicate logic (FOPL).

The above example can be seen as an instance of a broader class characterised by the following (where T is the text of the joke):

$$\begin{array}{l}\exists T_1, T_2 : text; M_1, M_2, M_3 : meaning; \\ I : interpretation \\ such\ that \\ substrings(T, T_1, T_2) \wedge meaning(T_2, M_3) \\ \wedge\ obviousmeaning(T_1, M_1, M_2)\ \wedge \\ conflicts(M_3, M_1) \wedge compatible(M_3, M_2) \wedge \\ form\_interpretation(M_2, M_3, I) \wedge absurd(I)\end{array}$$

Here we have introduced further predicates:

$obviousmeaning(\langle Text\rangle, \langle Sem_1\rangle, \langle Sem_2\rangle)$: This is true if $\langle Sem_1\rangle$ is a more obvious interpretation of $\langle Text\rangle$ than $\langle Sem_2\rangle$.

$compatible(\langle Meaning\rangle, \langle SemItem\rangle)$: $\langle Meaning\rangle$ can merge with $\langle SemItem\rangle$ to form a coherent INTERPRETATION.

$form\_interpretation(\langle Sem_1\rangle, \langle Sem_2\rangle, \langle Sem_3\rangle)$: $\langle Sem_1\rangle$ merged with $\langle Sem_2\rangle$ forms $\langle Sem_3\rangle$.

8 Notation

Although we have adopted FOPL as our notation for expressing the information about a joke, this in no way implies that the content of jokes is "logical" in any ordinary sense of the word; the FOPL notation is merely acting as our metalanguage: a convenient and concise way of writing down statements which involve abstract *objects*, *properties* and *relations* (see the summary of assumptions in Section 12 below). The use of a logical representation should facilitate the detection of classes and subclasses of jokes, as such inclusions (or similarities) will be reflected in logical expressions which subsume or overlap with one another.

FOPL has its merits as a meta-notation, but it could be rather verbose if used to state structural patterns in the manner shown above. For example, the existence of substrings has to be stated every time, and will not usually be very complex, so it may not be necessary to have the full expressive power of logic merely to state this information. Also, patterns will always be of a form in which some existentially quantified variables are introduced, and then some constraints are placed on them. Moreover, there are some recurring interrelations (e.g. that a meaning is associated with a particular stretch of text). Therefore, it is worthwhile developing a more succinct and perspicuous notation (which could in principle be expanded into FOPL).

Firstly, we can indicate the segmentation of the text into parts using brackets and subscripts:

$[_1$Why do birds fly south in winter?$]_1$
$[_2$It's too far to walk.$]_2$

The subscripts can be used in terms indicating the various objects involved, so that $\mathcal{M}(1)$ is the MEANING of TEXT labelled 1 (which in turn is notated as $\mathcal{T}(1)$). Where there are more than one possible MEANING for a particular TEXT $\mathcal{T}(N)$, these will be indicated by $\mathcal{M}(N_a)$, $\mathcal{M}(N_b)$, etc. An INTERPRETATION formed from the MEANINGS $\mathcal{M}(1), \ldots \mathcal{M}(n)$ will be written $\mathcal{I}(1, \ldots, n)$. Also, we can assume that all variables mentioned (or implicitly used) in the logical expressions are existentially quantified. The properties of items, and relationships between items, can still be written using the notation of FOPL. The example description can then be given as:

$$\begin{array}{l}obviousmeaning(\mathcal{T}(1), \mathcal{M}(1_a), \mathcal{M}(1_b))\ \wedge \\ conflicts(\mathcal{M}(2), \mathcal{M}(1_a))\ \wedge \\ compatible(\mathcal{M}(2), \mathcal{M}(1_b)) \wedge absurd(\mathcal{I}(1_b, 2))\end{array}$$

This has the merits of building the more basic linguistic relations into the notation, thus rendering the more substantive and humour-related predicates more prominent. That is, the notation is *not* being used to express interesting theoretical claims (as is sometimes advocated in early Chomskyan linguistics); instead, it is introduced to push the less interesting structure into the background and let the potentially significant predicates appear clearly.

To extend this notation to structural patterns as well as descriptions of individual jokes, we can indicate the decomposition of the text into substrings by a diagram of labelled brackets showing the relative positioning of the segments. That is, each pattern will have a "header" which shows the shape of the text, numbering its subparts; for example:

$[_1 \ldots]_1 [_2 \ldots]_2$

indicates a text made up of two substrings. The above example (1) is then an instance of the following pattern:

$[_1 \ldots]_1 [_2 \ldots]_2$
$obviousmeaning(\mathcal{T}(1), \mathcal{M}(1_a), \mathcal{M}(1_b)) \wedge$
$conflicts(\mathcal{M}(2), \mathcal{M}(1_a)) \wedge$
$compatible(\mathcal{M}(2), \mathcal{M}(1_b)) \wedge absurd(\mathcal{I}(1_b, 2))$

In this case, the application of the pattern to the specific example is encoded entirely in the binding of the text segments 1, 2, etc.

9 Descriptive jokes

Some very brief examples may help to demonstrate the approach advocated here. These make central use of the DESCRIPTION data type introduced earlier.

(2) Why is coffee like the soil?
It is ground. (Pepicello and Green, 1984)

This example could be described as:

$[_1 \ldots]_1 [_2 \ldots]_2$
$compares(\mathcal{T}(1), M, N) \wedge$
$yields\_description(\mathcal{M}(2), D) \wedge$
$describes(D, M) \wedge describes(D, N)$

providing that we have the following predicates:

$yields\_description(\langle Meaning \rangle, \langle Description \rangle)$: This is true if the $\langle Description \rangle$ can be extracted from the $\langle Meaning \rangle$.

$describes(\langle Description \rangle, \langle Meaning \rangle)$: This is true if $\langle Description \rangle$ describes $\langle Meaning \rangle$.

$compares(\langle Text \rangle, \langle Meaning_1 \rangle, \langle Meaning_2 \rangle)$: This is true if $\langle Text \rangle$ implies or states that $\langle Meaning_1 \rangle$ and $\langle Meaning_2 \rangle$ are similar.

To describe example (3)

(3) What is grey, has four legs, and a trunk? A mouse on vacation. (Rothbart, 1977)

we require the following predicate:

$obviousdescription(\langle Desc \rangle, \langle Meaning_1 \rangle, \langle Meaning_2 \rangle)$: This is true if both $\langle Meaning_1 \rangle$ and $\langle Meaning_2 \rangle$ are described by the DESCRIPTION $\langle Desc \rangle$, but this is more obviously the case for $\langle Meaning_1 \rangle$ than for $\langle Meaning_2 \rangle$.

The structural pattern is then:

$[_1 \ldots]_1 [_2 \ldots]_2$
$yields\_description(\mathcal{M}(1), D) \wedge absurd(\mathcal{M}(2))$
$\wedge \; obviousdescription(D, M, \mathcal{M}(2)))$

Example (4)

(4) What do you call a strange market? A bizarre bazaar. (Binsted, 1996)

can be described as:

$[_1 \ldots]_1 [_2 [_3 \ldots]_3 [_4 \ldots]_4]_2$
$yields\_description(\mathcal{M}(1), D) \wedge$
$describes(D, \mathcal{M}(2)) \wedge soundalike(\mathcal{T}(3), \mathcal{T}(4))$

where the definition of $soundalike$ can be based on the similarity metric for TEXTs.

10 Narrative jokes

A large class of more complex jokes are those which rely on narrative (i.e. "funny stories"). To describe the internal workings of such jokes, particularly those which have a "butt" or "target", we need to introduce data-types denoting entities (concrete or abstract) within the world of the story. For the moment, we will restrict ourselves to an EVENT-SEQUENCE, which a TEXT $narrates$, and a data-type CHARACTER for denotations of individuals within a story.

For example, consider (5).

(5) Russian officers in an Eastern European country go to a tavern. They order beer. The waiter places coasters on the table and serves the beer. Later they order another round. The waiter returning with the beer finds no coasters. 'OK,' he tells himself, 'these are collectors,' and puts down another set of coasters. When the third round is ordered and brought out, there are again no coasters. Angry, the waiter puts the beer down on the table, but places no more coasters. One of the Russian officers protests: 'What's this? No more crackers?' (Hetzron, 1991, p.62)

This could be approximated with the pattern:

$[_1 \ldots]_1 [_2 \ldots]_2$
$narrates(\mathcal{T}(1), E) \wedge$
$obviousinterpretation(E, I_1) \wedge$
$conflicts(\mathcal{M}(2), I_1) \wedge adopts(\mathcal{I}(2), C, I_2, E) \wedge$
$different(I_1, I_2) \wedge absurd(I_2)$

assuming we use the following predicates:

$narrates(\langle Text \rangle, \langle Events \rangle)$: $\langle Text \rangle$ recounts the EVENT-SEQUENCE $\langle Events \rangle$.

$obvious interpretation(\langle Events\rangle, \langle Interp\rangle)$: The natural INTERPRETATION for $\langle Events\rangle$ is the INTERPRETATION $\langle Interp\rangle$.

$adopts(\langle InterpA\rangle, \langle Char\rangle, \langle InterpB\rangle, \langle Events\rangle)$: In INTERPRETATION $\langle InterpA\rangle$, the CHARACTER $\langle Char\rangle$ adopts the INTERPRETATION $\langle InterpB\rangle$ for $\langle Events\rangle$.

along with some notion of "distinctness" of INTERPRETATIONS.

This emphasises the difference between MEANINGS and INTERPRETATIONS. The idea that the coasters have been consumed by the soldiers, is part of the INTERPRETATION of the narrative, although it is not stated as literal meaning. Also, the idea (imputed to the soldiers) that the coasters are crackers is an interpretation of the waiter's actions (within the world of the story), and is not a literal meaning of any fragment of text.

Although the analyses given above suggests that the relevant property of the soldiers' interpretation is that it is $absurd$, the pattern could perhaps be generalised to cover more jokes in a fairly natural way. Consider (6), the central example from Raskin (1985).

(6) 'Is the doctor at home?' the patient asked in his bronchial whisper.'No,' the doctor's young and pretty wife whispered in reply.'Come right in.'

This broadly follows the same pattern as (5). It could be argued that the amusing interpretation adopted by the character in this story is not so much $absurd$ as $taboo$ (with its implication of adultery). We could generalise the structural pattern by replacing $absurd$ with $inappropriate$, which we would define to be a disjunction of $absurd$, $taboo$, and perhaps other properties found to render interpretations amusing. Notice that *some* property is essential, otherwise the stories would be simple tales of misunderstandings, with no humorous effect. Raskin argues (in keeping with his semantic script-based theory of humour) that in (6) the important ingredient is not located in the mistaken interpretation alone, but in a form of comparison (*script opposition*) with the more obvious and natural interpretation. If we were to accept this idea as being a possible ingredient in making tales of misunderstanding funny, then the last term in the above pattern could be changed from $absurd(I_2)$ to $(inappropriate(I_2) \vee contrast(I_1, I_2))$ where $contrast$ embodies the appropriate form of opposition.

(Notice that we have not yet covered the satirical or mocking aspect of (5), as our structural patterns give no indication that this joke casts a slur on Russian officers. Such indirect or inferred content goes beyond the current paper.)

It is interesting to see how this pattern can also cover some apparently simpler jokes, such as (7) and (8).

(7) 'Mr Fields, do you believe in clubs for young people?' 'Only when kindness fails.' (Shultz, 1976) and elsewhere.

(8) A lady went into a clothing store and asked 'May I try on that dress in the window?'
'Well,' replied the sales clerk doubtfully, 'don't you think it would be better to use the dressing room?' (Oaks, 1994), citing from Clark (1968)

It might seem natural to focus on the fact that these examples rely on linguistic ambiguity (lexical ambiguity in (7), syntactic structural ambiguity in (8)), and to posit a pattern such as:

$$[_1 \ldots]_1 [_2 \ldots]_2$$
$$obvious meaning(\mathcal{T}(1), \mathcal{M}(1_a), \mathcal{M}(1_b)) \wedge$$
$$conflicts(\mathcal{M}(2), \mathcal{M}(1_a)) \wedge$$
$$compatible(\mathcal{M}(2), \mathcal{M}(1_b)) \wedge absurd(\mathcal{I}(1_b, 2))$$

However, this would miss a generalisation. In these very short stories, the utterances attributed to specific characters (the questioner in (7) and the lady in (8)) are in fact events – linguistic in nature – which are being narrated, and the ending of the joke involves some other character imposing an interpretation on these linguistic events which is not the obvious interpretation, and which is inappropriate in some way. Hence they both fall under the more general pattern given for (5) (as amended to go beyond $absurd$ in its last line). The linguistic ambiguity in (7) and (8) is then viewed solely as a means to an end, since it is the ambiguity which allows the different possible interpretations of the linguistic events.

Taking this perspective is different from some more traditional humour analyses (cf. (Attardo, 1994, Chapter 2)), in which a major dividing line is drawn between jokes which depend for their effect on the language in which they are expressed (*verbal* jokes) and jokes which are more easily translatable into other languages because the exact phrasing is not crucial (*referential* jokes). In such a taxonomy, (7) and (8) would be verbal, while (5) and (6) would be referential. That would obscure the generalisation, which we conjecture is a useful one, that all these four stories share a common mechanism which is significantly involved in their status as jokes.

11 Possible implementations

The work reported here is very preliminary, and does not, at present, involve computational implementation. However, the emphasis on formalisation and detail is intended to lead towards fuller symbolic models which could be implemented and tested. There are various ways in which this work could lead to implementation.

A rule tester. In the same way that a linguist could make use of a *grammar testing program* to check their rules (e.g. Friedman (1971)), software could be constructed to apply rules to data (much as was done with the JAPE system (Binsted et al., 1997)). If every output item is deemed to be a joke (by a suitable set of human subjects), our

claim of sufficiency is met, and every non-joke output counts as evidence against the sufficiency claim.

A joke understander. It is conceivable that a program could be built which takes natural language input, and rates each text as to its funniness, or (in terms of the proposals here) its joke-hood. This would have the added drawback that it would need an advanced natural language analysis system with relatively wide coverage (which is not easy to achieve at present). Also, there would be methodological complications, given that the approach here starts with subsets of jokes and gradually develops a description of a wider class of jokes. If such a system 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 that was already known. 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,

A humorous program. It would be natural to jump to the conclusion that computationally-oriented work on humour is aimed at producing, or might be used to produce, an intelligent joke-telling and joke-appreciating software agent. While this would be an impressive achievement, it would, to work properly, require more than a theory of joke structure. We would need a full theory of the use of humour in context, which is not the immediate aim of the descriptive work proposed here.

12 Moving towards a theory

The proposal here is that developing detailed descriptions of subclasses of joke is a viable approach to building a theory of what constitutes a humorous stimulus, (or at least a verbally expressed one). Theoretical progress will come gradually by finding generalisations across the various classes, and perhaps constructing a hierarchy of classes and subclasses. Such a taxonomy, based on systematic accounts of the mechanisms of jokes, would be more interesting than a simple classification of jokes into categories. Also, the explicit itemisation, within structural patterns, of the necessary properties and relationships will clarify what concepts have to be explicated as part of a fuller theory. Theoretical work can proceed by developing definitions of these predicates; in particularly, the frequently proposed concept of *incongruity* could be dissected in this way.

It may also be possible to discover some internal structuring within the structural patterns themselves. It may be that the conditions summarised in a structural pattern are of a few different sorts, some of which are more directly "structural" (e.g. $obviousmeaning, yields\_description$), with others being more concerned with the properties of interpretations that render them amusing (e.g. $absurd$). At the present stage of development, this is still vague speculation.

Developing a range of structural patterns for jokes will provide, at any stage of the ongoing work, a *provisional* theory in which a text is claimed to be a joke *if* it falls into one of the classes described; that is, the interim theory will offer a sufficient condition for joke-hood, but not (until it is complete) a necessary condition. Even if complete in this sense, it would *not* be a theory of *humour*. In the same way that a theory of grammar would not be a full theory of language (since it would not cover issues such as language use), a theory of joke structure would be just one step towards a broader theory.

The framework set out so far does, as mentioned in Section 1, implicitly adopt some theoretical positions. Some of the more basic of these are as follows:

(a) It is reasonable to posit abstract linguistic objects such as "meanings", and relationships such as "more obvious than" to represent the linguistic content of a text.

(b) There are properties of, and relationships between, these objects, such as "absurd", which can be defined simply in terms of the abstract objects themselves (i.e. without contextual information).

(c) If a suitable set of these primitives can be developed, then it will be possible to define, solely in terms of these concepts, subclasses of texts which constitute jokes.

(d) The status of a text as a joke can be determined from the text itself in isolation (in contrast to the funniness of the joke, which might be dependent on various more complex factors such as context, timing, hearer's sense of humour, etc.).

(e) The status of a text as a joke is relevant to its use or potential use in a humorous way; that is, a theory of the structure of jokes in isolation is a contribution to building a broader theory of humour.

These are very minimal theoretical claims, and leave open many decisions about the nature of joke structure. This is, at the current stage, desirable. The approach advocated here requires further empirical study in order to develop some substantive theoretical proposals which are based on the evidence. In that way, some general progress should come about.

Another essential development will be some criteria to guide and restrict the introduction and use of data-types and relationships, as the descriptive exercise is rather unconstrained otherwise. As noted in Section 3, there is an analogy with descriptive linguistics, in which guidelines have evolved for comparing one putative analysis with another.

Acknowledgements

Thanks are due to the members of the Alfa Informatica group at the University of Amsterdam, where this work was largely carried out.

References

Salvatore Attardo. *Linguistic Theories of Humour*. Mouton de Gruyter, Berlin, 1994.

Kim Binsted. *Machine humour: An implemented model of puns*. PhD thesis, University of Edinburgh, Edinburgh, Scotland, 1996.

Kim Binsted, Helen Pain, and Graeme Ritchie. Children's evaluation of computer-generated punning riddles. *Pragmatics and Cognition*, 5(2):309–358, 1997.

David Allen Clark. *Jokes, Puns and Riddles*. Doubleday, New York, 1968.

John M. Dienhart. A linguistic look at riddles. *Journal of Pragmatics*, 31(1):95–125, 1999.

Joyce Friedman. *A Mathematical Model of Transformational Grammar*. American Elsevier, New York, 1971.

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

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

William J. Pepicello and Thomas A. Green. *The Language of Riddles*. Ohio State University Press, Columbus, Ohio, 1984.

Victor Raskin. *Semantic Mechanisms of Humour*. Reidel, Dordrecht, 1985.

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

Mary K. Rothbart. Psychological approaches to the study of humour. In Anthony J. Chapman and Hugh C. Foot, editors, *It's a Funny Thing, Humour*, pages 87–94. Pergamon Press, Oxford, 1977.

Willibald Ruch, Salvatore Attardo, and Victor Raskin. Toward an empirical verification of the general theory of verbal humour. *HUMOR*, 6(2):123–136, 1993.

Thomas R. Shultz. A Cognitive-Developmental Analysis of Humour. In Anthony J. Chapman and Hugh C. Foot, editors, *Humour and Laughter : Theory, Research and Applications*, chapter 1, pages 11–36. Transaction Publishers, London, first edition, 1976.