The Structure of Forced Reinterpretation Jokes *

Graeme Ritchie
Division of Informatics
University of Edinburgh
Edinburgh, Scotland
g.d.ritchie@ed.ac.uk

Abstract

There is a very common type of joke, much discussed in the literature, in which the punchline forces a reinterpretation of the preceding set-up. We sketch the discourse mechanisms which would be needed to give an adequate account of these jokes. The aim is to illuminate the internal workings of such jokes, and also to outline what linguistic and logical mechanisms might be needed by a theory of joke structure.

Keywords: computational humour, ambiguity, jokes, incongruity-resolution

1 Aims

We take "computational humour" to refer to:

(a) the design, implementation and evaluation of programs which are capable of producing humorous artefacts or behaviours;

(b) the design, implementation and evaluation of programs which are capable of responding appropriately to humorous artefacts or behaviours;

(c) the use of computational concepts for studying and theorising about humour.

With the aiming of contributing to all three aspects of computational humour, we present an analysis of a much-discussed class of jokes at a greater level of detail and formality than in previous work. This may illuminate the humorous phenomenon itself and also act as a basis for implementations. Another aim is to clarify the basic concepts that humour research requires from disciplines such as artificial intelligence and linguistics.

Although we shall introduce notions of "proposition", "inference", etc., it is important to note that we are *not* putting forward a particular text-understanding model (or a logic), novel or otherwise. We are stating *the interface we need* from such a model – i.e. the mechanisms it should be able to support – but not the manner in which this is achieved.

This work is part of a larger programme to examine the *delivery mechanisms* used in jokes (i.e. linguistic devices, pragmatic effects, etc.) so as to clarify the role of these structuring techniques in the presentation of jokes. That is, the longer term aim is to sort out which factors can be attributed to the structure of jokes and which are inherent in the notions being conveyed.

2 The Class of Jokes

Many authors have remarked upon the fact that some jokes involve the following arrangement. There is an initial text (the *set-up*) which has potentially more than one interpretation, although one interpretation is more obvious than others. The end of the joke (the *punchline*) is in some way incompatible with this obvious interpretation, but summons up one of the less obvious meanings; (1) is an example.

*This work was carried out in the Language Technology Group, Department of Computing, Macquarie University, Sydney, supported by a Research Fellowship from the Leverhulme Trust.

(1) 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)

Some authors have even gone so far as to imply that *all* jokes take this form (Dascal, 1985; Dolitsky, 1992; Norrick, 2001), and it is at the heart of the SSTH (Raskin, 1985). Ritchie (1999) suggested that a more detailed analysis of this class of joke would involve the following concepts:

M_1: the first (more obvious) interpretation of the set-up text
M_2: the second (hidden) interpretation of the set-up text
M_3: the meaning of the punchline.
M_4: an interpretation formed by integrating the meaning of the punchline with M_2.

The properties and relationships relevant to humour would then be:

OBVIOUSNESS: M_1 is more likely than M_2 to be noticed by the reader.
CONFLICT: M_3 does not make sense with M_1
COMPATIBILITY: M_3 does make sense with M_2
CONTRAST: there some marked difference between M_1 and M_2 (or possibly M_1 and M_4).
INAPPROPRIATENESS: M_4 is inherently odd, eccentric or preposterous (ABSURD), or deals with matters not conventionally talked of openly, such as sexual or lavatorial matters, or forbidden political sentiments (TABOO).

(This is a slight refinement of the version in Ritchie (1999), where CONTRAST was called COMPARISON). Notice the difference between CONFLICT and COMPATIBILITY, which may in principle rely on a different meaning of "does make sense with", and not simply be opposites or negations of each other. CONFLICT is the symptom that the punchline meaning does not fit in, thus signalling a need for reinterpretation; COMPATIBILITY describes the relationship between punchline meaning and set-up meaning in the eventual resolution. These might rely on different semantic/discourse/inferential relations.

The current paper picks up that outline and formalises it in more detail.

3 Examining the Basic Concepts

3.1 Meanings and Interpretations

In Ritchie (1999), these jokes were referred to as *sudden disambiguation* jokes, and the various interpretations of the set-up were regarded as linguistic meanings (as could be the case in (1)). Consideration of a wider class of examples shows that this joke structure exists with "interpretations" which are something wider than linguistic meanings, as in (2).

(2) '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.' (Raskin, 1985)

Intuitively, the above account does describe (2), in that the punchline triggers a reinterpretation of the set-up. However, this joke does not involve linguistic ambiguity or two "meanings" in the normal linguistic sense of the term; rather, it involves two ways that the audience (and one of the characters in the joke) could make sense of the other character's actions. It is just about possible to relabel this as "pragmatic ambiguity", by claiming that there is ambiguity about the interpersonal import of the query *Is the doctor at home?*, but this is a rather contrived argument, since it would class any action which could have more than one outcome as "ambiguous". The difference in the two "meanings" of the query occurs at some higher level of the character's plan, and is not directly part of (for example) the illocutionary force of the utterance. Also, there are other, often longer, instances of jokes where there is no way to label the confusion as linguistic ambiguity. Jokes of this sort involve alternative interpretations for sequences of events within the narrative, and ambiguous utterances are simply one kind of event open to multiple interpretation. We shall therefore use the term *forced reinterpretation*, FR, rather than 'sudden disambiguation', for this class of joke.

Related to this is the fact that understanding the text, and therefore the humour, requires interpretation beyond the literal meanings of sentences, using some form of inference. Humour-relevant properties (e.g. incongruity) may be in the inferred content rather than in the literal meanings. Even in (1), the punchline *don't you think it would be better to use the dressing room?* does not directly *state* the idea that the lady had been suggesting using the window as a changing room, but *implies* it (in some loose sense); that is, this is the overall interpretation when punchline and (the less obvious meaning of) the set-up are taken together. Although we will use the term *inference* to refer to this "fleshing out" of the text's meaning, it is not logical inference in the normal sense (although it may include some valid deductions). It involves the filling-in of assumed or likely connections or consequences of the literal meanings, comparable to the *abduction* of Hobbs et al. (1993) and the *script-based* processing of Schank and Abelson (1977).

We shall adopt as a primitive type of item the *proposition*. This can be thought of as a fact or hypothesis or statement, which could be formalised as a well-formed formula in a suitable logic. For the moment, this is just a primitive notion. A set of propositions may be *consistent*, which we also leave as a primitive concept. Being consistent is not central to our analysis. In particular, (logical) consistency is not taken as related to incongruity, CONFLICT or COMPATIBILITY: it is merely a condition we impose upon sets of propositions inferred from a text.

The meaning of a sentence will be a set of propositions, the input to inference will be a set of propositions, and the output of inference is discussed below.

3.2 OBVIOUSNESS

Any logical (or quasi-logical) mechanism that is used to describe the imparting of information during a joke should allow *defeasible* inferences, with *non-monotonic* effects (Charniak and McDermott, 1985). That is, we must be able to express the fact that certain propositions *normally* (or *by default*) imply other propositions, but that these implications are not immutable in the face of later information. The formal logic must allow what is sometimes referred to as *belief-revision* (Gärdenfors, 1988). We will build this into our formalisation in two places. Firstly, the new information in each succeeding part of the text will be combined with previous propositions using an *update* operator, $\bigoplus$, so that $P \bigoplus P'$ is the set of propositions produced by updating (in some suitable way) the existing set P with some new information P'. This is intend to encapsulate the normal operation used during incremental understanding of a text, and its definition will depend on the precise knowledge-representation scheme adopted. That is, we delegate the problem of normal (non-humour-related) belief-revision within a discourse to the architects of the full language-understanding model. Also, there is an element of defeasibility in our handling of OBVIOUSNESS (which is perhaps closer to a notion of "salience" or "prominence"). Our "inference" process will yield not one interpretation, but an *ordered set* of possible interpretations, where the ordering reflects the degree of OBVIOUSNESS. This ordered set will always contain a *maximal* element, so that there is some notion of the "most obvious" interpretation.

3.3 CONFLICT

In the sketch so far, the punchline can be recognised by the fact that it does not make sense (CONFLICTS) with the current (most OBVIOUS) interpretation of the preceding text. Nevertheless, it must be possible for at least the literal meaning of the punchline text to be interpreted semantically (despite its lack of fit with the textual context), or this CONFLICT would not be detectable. Inspection of various examples, including (1) and (2), suggest that the difficulty involves *discourse coherence*: a response does not seem to answer the question, or a character's utterance does not make complete sense with respect to the hitherto OBVIOUS interpretation (cf. Jurafsky and Martin, 2000, Sects 18.2, 19.4). That is, the CONFLICT relation should *not* be identified with logical inconsistency, nor with the kind of semantic clash that results from violation of selectional restrictions (e.g. applying a verb to a semantically unsuitable kind of object). In this type of FR joke, the punchline *assumes* a particular, hitherto hidden, perspective, giving an effect very much like the failure of a *presupposition* (van der Sandt, 1988). The audience, in these cases, has to make sense of the punchline, by "working backwards" (cf. Hobbs et al. (1988, 1993)). In

our formal model (below) we shall approximate this by representing the literal meaning of a text segment as having *two* components. As well as the core meaning (the set of propositions conveyed by the segment) there will be *felicity conditions*, which are propositions which must be checked in a suitable way against the current context before the core meaning can be considered. This gives a crude model of discourse factors such as presupposition, focus maintenance, rhetorical structure, etc., which we will encapsulate by saying that a set of propositions P *can accommodate* a set of felicity conditions FC if there is no discourse incoherence. (The use of 'accommodate' is related to our handling of COMPATIBILITY, below.)

Very little of the formalisation here depends on using an explicit set of felicity conditions to check coherence, so it would be possible to enhance the model with some subtler notion of discourse coherence, or with a more refined notion of the types of 'incoherence' used in such jokes.

3.4 COMPATIBILITY

Once the punchline has been detected, there is then the question of how the meaning of the punchline relates to the possible interpretations of the set-up. The usual account is to say that the punchline evokes a less OBVIOUS interpretation of the set-up.

In jokes where (we suggest) the CONFLICT relation which marks the punchline is discourse incoherence, it also seems to be the case that the punchline is *discourse coherent with a less obvious interpretation of the set-up.* That is, COMPATIBILITY is some form of discourse coherence. For example, in (1), if the lady-character had intended to use the shop-window as a changing room, then the clerk-character's response would be completely coherent. That is, the OBVIOUS interpretation does not fulfil the punchline's assumptions, but the less OBVIOUS interpretation does.

Related to this is the notion of *accommodation*, which has been discussed in the literature alongside presupposition. When a piece of text presupposes some proposition (potential fact), then there are various ways that the hearer of the text may react. If the hearer believes the proposition to be untrue, then he may not accept the text as comprehensible; if he believes the proposition to be true, he will accept the text as comprehensible (even though he may dispute the truth of the text itself); if he has no information one way or another about the truth of the presupposed information, he may accept it as true, since the speaker appears to be assuming it. This latter case – acceptance of presupposed material in the absence of conflicting knowledge – is known as *accommodation.*

As well as occurring during routine text understanding, this effect can be used in certain jokes. If the joke-teller employs an character stereotype within the joke (e.g. a mean Scot, a stupid Irishman), the hearer may, for the sake of the joke, accommodate this information, even if he/she has not encountered the particular stereotype before. That is, accommodation allows the joke either to convey the insulting message or to assume it, depending on the hearer's state of belief.

We will introduce an accommodation operator $\odot$ to represent this absorption of assumptions into existing knowledge.

3.5 ANOTHER VERSION OF CONFLICT AND COMPATIBILITY

There are other examples where neither CONFLICT nor COMPATIBILITY involve discourse coherence. Consider example (3).

(3) John and his wife Mary were having a shower together in their upstairs bathroom when the doorbell rang. Mary heard the bell, got out of the shower, wrapped a towel around her, went downstairs, and opened the door. Their neighbor Charlie looked at her from the doorway, and said, 'Oh. I see that I got you out of the shower. Sorry about that.' 'That's all right,' Mary said, 'What do you want?'
'Not too much... my goodness you have beautiful skin. It's so pink from the shower. Mary, if I was to give you a hundred dollars, would you remove the towel from your upper body?' Mary thought about it for a minute, figured why not, for a hundred bucks, and removed the towel from her breasts. 'Wow,' Charlie exclaimed, 'they are truly beautiful. Listen, for

another hundred bucks would you consider taking the towel all the way off?'
'Why not,' Mary thought, 'that's a lot of money,' and she dropped the towel completely to the floor. Charlie had a good look, complimented her again on her fine looking body, reached into his pocket, took out two hundred dollars, gave it to her, and left. As she got back up stairs and was getting back into the shower, John asked her who was at the door. 'Just Charlie,' she said, as she started to rub his back. 'Charlie, eh,' said John, 'Did he give you the two hundred dollars he owed me?' (from `www.jokes2000.com`)

In this example, the final (punchline) utterance does not *presuppose* any particular interpretation of preceding events. It simply provides new information which causes the audience to reinterpret those events. It is not clear what signals that a punchline has arrived – perhaps the ending of the text.

In terms of COMPATIBILITY, what seems to be happening is that the new information renders one of the (hitherto) less OBVIOUS interpretations more OBVIOUS, making it now the most OBVIOUS. That is, if we assume that interpreting the set-up results in a set of possible interpretations, ordered by their OBVIOUSNESS, then the content of the punchline, taken together with the set-up, yields a different ordering.

Notice that many jokes where the punchline is signalled by discourse incoherence also involve linguistic ambiguity, whereas (3) does not. It is not clear if this is significant; that is, whether FR jokes which rely on linguistic ambiguity in the set-up always make use of discourse incoherence to indicate the arrival of the punchline. (The converse does not hold, as (2) seems to use the discourse incoherence device but not linguistic ambiguity.) The evidence here suggests that discourse incoherence is not universally present as a punchline marker.

4 A FORMALISATION

We will now attempt to set out a relatively precise statement of the FR class of jokes, expanding the remarks in section 3 into a fuller formalisation.

4.1 RELATIVITY

The status of a text as a joke (and also its degree of funniness) is always relative to some body of knowledge: cultural knowledge, social context, personal beliefs of the joke audience, etc. We will therefore include in our formalisation a structured entity called an *joke-interpreter*. This is not intended in any way as a realistic model of a human listener; rather, it is a bundling together of the miscellaneous factors relative to which a joke is to be interpreted or judged. A *joke-interpreter* will be taken as consisting of the following two components:

Linguistic Knowledge. We shall abstract over the details of linguistic levels, etc., and simply stipulate that the Linguistic Knowledge must somehow furnish a *text-to-semantics* mapping M, which takes two arguments – a text string and a set of propositions – and returns a set, ordered by their OBVIOUSNESS, of *literal meanings*. A literal meaning will be a pair of sets of propositions, where the first of the pair is a (possibly empty) set of *felicity conditions* for that sentence (i.e. presupposed facts which must match the context if the sentence is to be coherent in the discourse), and the second set represents the *core meaning* of the sentence (i.e. what is being asserted, queried or ordered by the sentence). Informally, the second argument to M (the propositions) will represent the knowledge (beliefs, etc.) that the audience uses when semantically analysing the text so far. Each of the meanings returned by M corresponds to a bare meaning of the sentence, without further inference or contextual reasoning. (We ignore here any linguistic notions of temporary discourse structures, such as might be used for processing pronoun references; the simplifying abstraction here is that we are interested only in the result after such transient data items have been used and disposed of.)

Persistent beliefs. This is a set PB of propositions, representing knowledge/beliefs about the world, including cultural assumptions, social prejudices, etc. These items are referred to as

'persistent' rather than 'permanent' because they are subject to change, but they typically survive across different instances of joke-interpreting or story-hearing. (The analogy is with persistent data structures in computer systems, which, although updatable, remain in the system from one computing session to another.)

It is arguable that the inference mapping, I, should form part of the joke-interpreter, to make the inference relation relative to the audience and context. However, for the moment we have left it out, thereby assuming that formal inference is general.

4.2 Interpreting Texts

We shall assume that a joke text consists of the concatenation of one or more text strings $T_1, \ldots T_n$, where each T_i is a suitable input to the text-to-semantics mapping M in the joke-interpreter (roughly, it is a sentence or complete phrase). We envisage an arrangement whereby each T_i is first mapped to a set of possible literal meanings, inference is carried out on the most obvious of these, then the process continues from this state for the next T_i in sequence. Initially, M acts on T_1 starting from PB (i.e. $M(T_1, PB)$), producing an ordered set of candidate literal meanings, $(LM_1, <)$, where each element of LM_1 is of the form (FC_1^j, CM_1^j), a pair consisting of the felicity conditions and core meaning. The next step is to select (FC_1^{max}, CM_1^{max}), the maximum element of $(LM_1, <)$, which is the "most obvious" literal meaning. If PB can accommodate FC_1^{max}, then inference can take place, starting from $(PB \odot FC_1^{max}) \bigoplus CM_1^{max}$; that is $I((PB \odot FC_1^{max}) \bigoplus CM_1^{max})$. (As stated earlier, '$\odot$' is the "accommodation" operator, and '$\bigoplus$' is the "belief-update" operator.) As with the text-to-semantics mapping, we are positing that inference produces an *ordered* set of results, reflecting the varying degrees of obviousness of the interpretations of the given meaning; that is, $I((PB \odot FC_1^{max}) \bigoplus CM_1^{max})$ will be of the form $(R_1, <)$ where R_1 is a set of sets of propositions. The maximum element from this set, S_1, is then the interpretation so far of the text, and can be used as an input to the interpretation of T_2, in place of PB. In this way, interpretation proceeds segment-by-segment, with obviousness, felicity conditions and inference involved at each step. (It might be helpful to assume that M is sufficiently subtle that it can take account of the relationship between FC_i and S_{i-1}, to the extent that FC_i^{max} is unaccommodatable in S_{i-1} only if there is no alternative meaning of the textstring which can be accommodated in PB; that is, FC_i^{max} will normally be accommodatable unless *all* the available literal meanings are unaccommodatable).

We can define this more precisely and less procedurally as follows.

Definition: An *interpretation* of a text $T_1, \ldots T_n$, relative to a joke-interpreter (M, PB), is a set S_n of propositions such that there is a sequence of consistent sets of propositions $S_0, \ldots S_n$, where:

(a) $S_0 = PB$

(b) for $1 \leq i \leq n$, there is an element (FC_i^j, CM_i^j) of $M(T_i, S_{i-1})$ such that S_{i-1} can accommodate FC_i^j.

(c) for $1 \leq i \leq n$, S_i is an element of $I((S_{i-1} \odot FC_i^j) \bigoplus CM_i^j)$.

That definition allows for the literal meaning used at each step and the inferences drawn from that meaning to be less than maximally obvious: any item of the relevant set may be chosen. The next definition is more specific.

Definition: The *most obvious interpretation* of a text $T_1, \ldots T_n$ relative to (M, PB), is a set S_n of propositions such that there is a sequence of consistent sets of propositions $S_0, \ldots S_n$, where:

(a) $S_0 = PB$

(b) for $1 \leq i \leq n$, S_{i-1} can accommodate FC_i^{max} where (FC_i^{max}, CM_i^{max}) is the maximum element of $M(T_i, S_{i-1})$.

(c) for $1 \leq i \leq n$, S_i is the maximum element of $I((S_{i-1} \odot FC_i^{max}) \bigoplus CM_i^{max})$.

Any interpretation which is not the most obvious interpretation is said to be *a less obvious* interpretation.

Notice that although each S_i is stipulated to be consistent, there is no requirement that $S_i \subseteq S_{i+1}$ or that $S_i \cup S_{i+1}$ be consistent, so there could be changes of facts at a step in the sequence, if the accretion of information is non-monotonic.

4.3 Forced Reinterpretation Jokes

For simplicity, we assume that the punchline is the final (nth) segment of the text, and adopt all the notation of section 4.2 above. First, we shall formalise the condition we have referred to as Conflict, assuming that this relies on some sort of presupposition failure or discourse incoherence. We also assume here that this will involve the most obvious literal meaning of the punchline. These definitions are relative to a joke-interpreter (M, PB).

Definition: A text $T_1 \ldots T_n$ has *punchline conflict* in interpretation S_{n-1} if:

(a) S_{n-1} is the most obvious interpretation of $T_1 \ldots T_{n-1}$ based on a sequence $S_1, \ldots, S_{n-1}$

(b) S_{n-1} cannot accommodate FC_n^{max} where (FC_n^{max}, CM_n^{max}) is the maximum element of $M(T_n, S_{n-1})$.

Definition: A text $T_1 \ldots T_n$ has *punchline resolution* with interpretation $S_n^{'}$ if:

(a) There is a less obvious interpretation $S_{n-1}^{'}$ of $T_1 \ldots T_{n-1}$, based on a sequence $S_1^{'}, \ldots, S_{n-1}^{'}$;

(b) $S_{n-1}^{'}$ can accommodate FC_n^{max}, where (FC_n^{max}, CM_n^{max}) is the maximum element of $M(T_n, S_{n-1}^{'})$;

(c) $S_n^{'} = I((S_{n-1}^{'} \odot FC_n^{max}) \bigoplus CM_n^{max})$.

Definition: A text T constitutes a *discourse-coherence misunderstanding*, relative to to a joke-interpreter (M, PB), if T has both punchline conflict and punchline resolution.

These definitions cater for the (wide) class of FR jokes in which both Conflict and Compatibility are based on discourse coherence.

To cover jokes such as (3), we need some further definitions. The idea will be to define a situation where information extracted from the punchline (initially interpreted in the context of the set-up) would, if inserted earlier in the text, have led to a different interpretation of the text: for example, if (3) had included the sentence *Charlie owed John two hundred dollars* at an early stage of the narrative. This involves having an interpretation sequence $S_1^{'}, \ldots, S_{n-1}^{'}$ which initially is the same as the original most obvious interpretation sequence, but has the punchline information added at some intermediate point, with subsequent stages then making inferences from that enhanced base of propositions.

Definition: Let $T_1 \ldots T_n$ be a text. Suppose $T_1 \ldots T_{n-1}$ has a most obvious interpretation S_{n-1} based on a sequence $S_1, \ldots, S_{n-1}$. Suppose that the maximum element of $M(T_n, S_{n-1})$ is (FC_n, CM_n) and CM_n (the core meaning part) contains a set of propositions P such that the following holds. There is a sequence of sets of propositions $S_1^{'}, \ldots, S_{n-1}^{'}$ such that

(a) $S_{n-1}^{'} \neq S_{n-1}$

(b) for some k, $1 \leq k \leq (n-1)$:

  (i) for $1 \leq i < k$, $S_i = S_i^{'}$;

  (ii) $S_k^{'}$ is the maximum element of $I(((S_{k-1}^{'} \odot FC_k^{max}) \bigoplus P) \bigoplus CM_k^{max})$ where (FC_k^{max}, CM_k^{max}) is the maximum element of $M(T_k, S_{k-1}^{'})$;

  (iii) for $k < i \leq n-1$, $S_i^{'}$ is the maximum element of $I((S_{i-1}^{'} \odot FC_i^{max}) \bigoplus CM_i^{max})$, where (FC_i^{max}, CM_i^{max}) is the maximum element of $M(T_i, S_{i-1}^{'})$.

Then $T_1 \ldots T_n$ has *punchline revision* with interpretation $S_{n-1}^{'}$.

4.4 The Critical Ingredient

The definitions in section 4.3 above do *not* fully define FR jokes, or even subclasses of FR jokes, because they do not demand any factor that will result in humour. These definitions characterise misunderstandings, and not all misunderstandings are humorous. Jokes require something further, perhaps Contrast or Inappropriateness. The definitions above offer *necessary* conditions for membership of these particular subclasses of FR jokes, but not *sufficient* conditions.

> *Conjecture:* In a text which constitutes a discourse-coherence misunderstanding or has punchline revision, if the resulting interpretation is Absurd or Taboo, the text constitutes a joke.

Of course, this defers as subsidiary research problems the definitions of Absurd and Taboo. Also, even for this narrow class of joke, matters are more complex. Consider (4).

(4) While on a trip in a remote country area, Peter and Bob were caught in a blizzard. They found a farmhouse and asked the occupant, a very attractive woman, if she could give them accommodation. She agreed to put them up. Nine months later, Peter got a letter from the woman's attorney. He went to Bob and said, 'Bob, do you remember that good-looking woman at the farm we stayed at?' 'Yes, I do.' 'Did you happen to go to her room in the middle of the night and have sex with her?' 'Yes, I admit that I did.' 'Did you use my name instead of telling her your name?' Bob blushed and said, 'Yeah, I'm afraid I did.' 'Well, thanks!' said Peter. 'She just died and left me everything!' (shortened from `www.jokes2000.com`)

This joke has the same inferential pattern as (3). However, it is hard to argue that the "interpretation" established by the punchline is somehow Absurd or Taboo. Intuitively, the audience is led to expect an scenario which is mildly improper (accidental pregnancy and deception over paternity) but then receives a *less* Taboo interpretation. There may be a case here for some form of Contrast as the extra "kick" which transforms a mere change of interpretation into a joke.

5 Another Joke Class: Drawing Out Implications

Now that we have set out a formal account for FR jokes, we can start to examine other types of joke from a similar perspective. We have already pointed out that the location of the incongruity in an FR joke may not be the literal meaning, but some inference from there. There are also non-FR jokes where inferring an incongruity seems to be the central joke-creating device. Example (5) makes use of a common device in narrative jokes, misinterpretation by a character.

(5) On a US highway, a traffic cop sees a car puttering along at 22mph. Thinking this is abnormal behaviour, he gets the car to stop, and he goes to speak to the driver. The occupants are four old ladies. The driver says 'Officer, I don't understand. What seems to be the problem? I was driving exactly at the speed limit! Look, there's the sign – 22 mph'. The officer explains to her that '22' on the sign is the route number, not the speed limit. The woman smiles in embarrassment. 'Oh, thank you,' she says, 'It's a good thing you didn't see us a few minutes ago, on Route 119.' (shortened from `www.jokes2000.com`)

Although this joke involves misinterpretation by a character, it is not misinterpretation of information presented in the set-up, and revelation of this misinterpretation does not provide the punchline, so it is not in our FR class of joke. The punchline supplies further information which is not in itself humorous or incongruous, but which, through the provision of a simple statement, permits the inference of an amusing consequence of the already-established misinterpretation. There is a sense in which this inference, or something very similar to it, could be made, once the misconception has been stated: driving at a speed numerically equal to the route number is bound to result in some very high speeds (particularly in the countries such as the USA where speeds are stated in miles per hour). However, the punchline both provides a concrete instance of this possibility, and draws attention to that consequence. The fact that the driver-character's

behaviour is based on a misconception is not central to the working of the joke. It is the general rule adopted by this character, together with the punchline statement, which allows the inference.

Definition: Using all the terminology and notation set out earlier, a text $T_1 \ldots T_n$ *has a punchline inference* iff:

(a) the text $T_1 \ldots T_{n-1}$ has a most obvious interpretation S_{n-1};

(b) the maximum element of $I((S_{n-1} \odot FC_n^{max}) \bigoplus CM_n^{max})$ contains a set A of propositions which

- are not a subset of CM_n^{max} (i.e. they must be inferred);
- are not a subset of S_{n-1} (i.e. prior to the punchline, they were not OBVIOUS).

As with the other "delivery mechanisms" in section 4, this definition merely describes a subclass of natural language text, which could equally well be used for non-humorous discourse. Once again, to reach the status of a joke, we require some further component, such as ABSURDITY.

> *Conjecture:* In a text which has punchline inference, if the resulting proposition set (A in the above definition) is ABSURD or TABOO, the text constitutes a joke.

In this example, we are proposing that the idea of an old lady driving at 119mph is (sufficiently) ABSURD. Implicit in our two *Conjectures* is the idea that an ABSURDITY may or may not be funny in itself, but if presented using a suitable "delivery mechanism", it is humorous (see section 6 below for further remarks).

6 DISCUSSION

We have not proposed any algorithms, nor have we offered a software design. What we have tried to show – in some detail – is how certain types of joke can be analysed in terms of a fairly conventional text-interpretation mechanism, coupled with some (yet to be defined) key notions of INAPPROPRIATENESS and CONTRAST. Our analyses of the joke classes here could form the basis of a computational model, using suitable compromise modules for text-understanding and inference, and some preliminary definition of INAPPROPRIATE (and perhaps CONTRAST). All of the basic text-understanding facilities we have proposed (interpreting relative to a base of knowledge, using inference at each step, having some notion of discourse coherence) are relatively uncontroversial. Jokes of the sort considered here (and probably most "funny stories") do not rely on special text processing. They are normal texts in most respects, though with the slight abnormality that what we have called 'discourse-coherence misunderstanding' jokes include as an integral part something that would normally mark a text as pragmatically awkward or even ill-formed. Nevertheless, such infelicities could occur in non-joke texts, with the hearer having to make an additional effort to understand them.

Even where we assume concepts which can be plausibly claimed to be linguistic notions (i.e. not part of humour theory), the position is not simple. 'Discourse coherence' is far from being a single, uniform, well-understood notion, involving as it does ideas of presupposition, topic and focus, rhetorical structure, etc.

One of the many unfinished issues is the notion of ABSURDITY. Some examples of ABSURDITY (such as the old lady driving at high speed) could be argued to be *incongruity*, since they involve two ill-matched ideas being forced together. In that case, our sketch of these jokes could be loosely characterised as involving "incongruity conveyed by a suitable delivery mechanism". Even if we allow ourselves to defer or delegate the definition of ABSURD as one of the building-blocks for FR (and other) jokes, we have to be careful that we are indeed referring to the same concept on all occasions. Example (6) can be analysed as a 'discourse-coherence misunderstanding' joke.

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

For this to be a joke, it should, under our analysis, contain a further ingredient of INAPPROPRIATENESS, probably ABSURDITY. In this example, the (factually correct) punchline offers a

(presumably incorrect) reason for the birds' behaviour. In what sense is this ABSURD? The image of flocks of birds trudging across continents might be ABSURD, but that is only indirectly invoked, as even the punchline assumes that this does *not* happen. Moreover, is this notion of ABSURDITY similar to the geriatric racing in example (5)?

As noted earlier, the SSTH (Raskin, 1985) uses a variant of the FR picture to describe all humour conveyed verbally. In spite of that, the formalisation proposed here is not redundant, for three reasons. Firstly, the SSTH is still relatively ill-defined, and therefore its exact predictions are unclear. Secondly, the nub of the SSTH is a claim that the critical humorous element in a joke is a particular form of what we have labelled CONTRAST (namely, *script opposition*), but we wish to explore the workings of jokes without that theoretical (and as yet unproven) commitment. Thirdly, SSTH also has a built-in commitment to a particular semantic view of language-processing, based on *scripts*, and we wish to see how analyses can be framed without that prior assumption. In a sense, the SSTH account and the analyses here are comparably underdeveloped: both depend upon having a model of language-understanding (which SSTH intends to achieve using scripts), and both propose a rather ill-defined but crucial concept which creates humour, script-opposition or INAPPROPRIATENESS.

The formalisation presented here is still preliminary and tentative. To some extent, one of the methodological goals of this outline is to stimulate debate at this level of detailed description. Those who wish to contest these analyses are invited to respond using comparable formality and precision, so that an investigation can proceed by informed argument.

REFERENCES

Charniak, E. and McDermott, D. (1985). *Introduction to Artificial Intelligence.* Addison-Wesley, Reading, Mass.

Dascal, M. (1985). Language use in jokes and dreams: sociopragmatics vs. psychopragmatics. *Language and Communication*, 5(2):95–106.

Dolitsky, M. (1992). Aspects of the unsaid in humor. *HUMOR*, 5(1/2):33–43.

Gärdenfors, P. (1988). *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* MIT Press, Bradford Books, Cambridge, MA.

Hobbs, J., Stickel, M., Appelt, D., and Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, 63:69–142.

Hobbs, J., Stickel, M., Martin, P., and Appelt, D. (1988). Interpretation as abduction. In *Proceedings of 26th Annual Meeting of the ACL*, pages 95–103.

Jurafsky, D. and Martin, J. (2000). *Speech and Language Processing: an Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition.* Prentice-Hall, New Jersey.

Norrick, N. R. (2001). On the conversational performance of narrative jokes: toward an account of timing. *HUMOR*, 14(3):255–274.

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.

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.

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

van der Sandt, R. (1988). *Context and Presupposition.* Croom Helm, London.