Logical Forms in Wit

Abstract

This paper presents a preliminary theory for the logical structure underlying a certain class of jokes and witty comments. We show that there is a range of jokes which can be understood as intentionally poor speech acts (where the intention is conveyed by a variety of means, including nonsense, parody and self-reference). Constructing such utterances typically requires intelligence and creativity. Hence jokes may have evolved (via mate-selection forces) as a way of demonstrating mental fitness.

The goal of this work is both improving our understanding of humour (via formal theories of jokes), and techniques for the computer recognition and generation of jokes. It therefore focuses on structures that require relatively little world-knowledge. We identify several logical forms that give rise to jokes, and show how these forms correspond to violations of 'the rules of good speech' (e.g. Grice's Maxims). As a result of this analysis, several algorithms are proposed for joke generation. These algorithms have yet to be implemented and tested. Hopefully though, they will extend computer-generation to a wider range of jokes.

1 Introduction

Laughter is an intrinsically human thing. Comedy plays an important role in social relations, and a sense of humour is a key part of a person's character. An A.I. theory of humour is therefore an attractive goal. Firstly, for the light it could shed on this interesting aspect of human behaviour. Also, the ability to make and recognise jokes could aid Human Computer Interaction (HCI) in many areas.

Creativity is key to humour. One might suppose that a humorous A.I. could be created by storing examples, coupled with techniques[1] for assessing input (for joke recognition) or selecting jokes relevant to the current context (for generation). This is unlikely to work well at either task (at least, not for long on any one human). Originality is an important part of humour.[2] and jokes quickly go stale.[3]

This paper presents preliminary investigations on logical forms in *wit*, which we define to be those jokes characterised by ingenuity or verbal skill. We focus on abstract forms of joke, which may be more amenable to computer generation. We will identify several logical forms that give rise to jokes. These then suggest algorithms for joke generation – although these have yet to be implemented and tested.

1.1 Disclaimer

We do not attempt to give a full theory of humour, or even of one aspect of humour, but merely to identify some forms of joke which seem suitable for computational recognition and generation. Indeed, it seems unlikely that a complete theory of humour is possible. As Minsky observes, "One might suppose [the difficulty in defining what a joke is] is a mere surface difficulty, and hope that we may yet find a single underlying structure from which all funny things spring – some basic 'grammar of humour' or 'comical deep structure'. Not so, I fear; when we look deeper for that underlying structure of humour we shall still find a vexing lack of unity. I argue that this is a consequence of the way things usually evolve in biology.... [In] structures created by evolution, we find that only rarely does one evolutionary increment serve a single purpose... Behaviour emerges from a network of interdependent mechanisms, and one cannot expect any compactly circumscribed theory (or mechanism) completely to 'explain' any single surface component of behaviour."[10]

2 Background

2.1 Speculation

The purpose and origins of humour are beyond the scope of this paper. Nevertheless, we indulge here in some speculation which will inform the formal work that follows.

[1] For example, a statistical text comparison, such as HAL, between recent dialogue and the text of known jokes could be used for both tasks.

[2] Hence Sheridan's insult "The right honourable gentleman is indebted to his memory for his jests, and to his imagination for his facts."

[3] Whilst there is a lot of similarity between many jokes, this is typically disguised as much as possible. Strangely though, an 'old favourite' may remain funny – as long as it is not presented as a new joke.

Jokes serve both social and psychological roles, such as mockery, easing tensions and relief. It is these roles that are usually the focus of theories of humour (which [11] groups into the superiority, relief, & incongruity theories). However these roles could be served without jokes by simpler mechanisms, and indeed often are – c.f. Provine's study of laughter in everyday conversation, where very little observed laughter was actually related to jokes[13]. Hence these roles cannot explain the cleverness and imagination found in jokes.

Good thought vs. Poor Thought

Alexander Pope offers the following explanation: "True wit is nature to advantage dressed, What oft was thought, but ne'er so well expressed".[12] According to this view, wit displays ingenuity by definition, since it is intelligent elegant communication.

More recently, Marvin Minsky proposed the opposite view, that jokes are a device for teaching "fallacies in daily life".[10] According to Minsky, the role of humour is to help learn these things. Minsky also hypothesises that humour is mainly about taboos and the nonsensical, because our mind uses censors to suppress unproductive mental states and analogies to accelerate cognition. Thus, he says, humour fulfils a function of helping us in "recognising and suppressing bugs – ineffective or destructive thought processes."

A case can be made for either view, but in fact neither withstands too much scrutiny.

Whilst jokes typically display a sophisticated use of language, this is not because they are the epitome of communication. Jokes do often centre on accepted truths, and are frequently used to make a point in debate. However if jokes appear in text-books or lectures, it is to lighten the tone, rather than convey complex information. As Minsky observes, jokes can often involve false ideas, fallacious reasoning, nonsense and taboo-breaking. However Minsky's view that their purpose is to teach these things is backwards: To appreciate such a joke, we must already understand that the joke represents fallacy, nonsense, or taboo-breaking. A joke that taught these things would not be funny.

Humour as posture and pose

Perhaps the key aspect of a joke is that it demonstrates mental dexterity. We speculate that this was the original purpose of jokes: making jokes (and to a lesser extent, understanding jokes) is a way of demonstrating mental fitness. Thus a good sense of humour is evidence of a good brain, and hence the development of jokes may have been driven by the evolutionary force of *mate selection*[4]. We call this the *showing-off* theory of humour, and put it forward as an explanation for the ingenuity often seen in jokes. The primary prediction of this theory is that jokes should be difficult to form (something which is *not* predicted – and indeed is counter-productive – under other theories). This explains humorous songs, which are usually funnier than the same jokes would be if told in a prose style – even though rhythm & rhyme have no direct connection to humour.

[4]The process whereby a trait develops not for its direct survival value, but as a way of publicly demonstrating general fitness – which attracts a mate, thereby perpetuating the gene responsible.[16]

2.2 Computational approaches to humour

There has been a lot of informal analysis of what is funny written by comedians. However there has been very little work done on this topic from an A.I./computational perspective. We give a very brief overview of that work here; for a more detailed study, see [11].

There are a number of systems that use computers to produce comedy. Most of these systems work by randomly mangling text (e.g. *The Dialectizer* by S.Stoddard at http://rinkworks.com/dialect/) or randomly combining text fragments (e.g. *The Postmodernism Generator* by A.Bulhak at http://www.elsewhere.org/cgi-bin/postmodern/). In either case, humorous nonsense is created in a particular style. The results can be very funny. However these systems are not steps on the way to computational humour. They do not contain a model of humour, have no ability to recognise humour, and require considerable hand-coding of rules to create. Such systems embody the creativity of their author, rather than giving creativity to the computer.

Koestler put forward the theory that humour stems from sudden shifts in perspective.[6] These perspective shifts are often linked with shifts in status (e.g. disguised insults) or breaking taboos (e.g. where a sexual meaning is suddenly revealed). This is a general but powerful theory. It encompasses the influential incongruity-resolution theory of humour, where jokes consist of creating and then resolving incongruities[11]. It also encompasses the surprise theory, where jokes hinge on generating expectations, then deliberately breaking them. A broken expectation is one form of perspective shift.

There is a 'General Theory of Verbal Humour' (GTVH) in the literature. Ritchie states in [14] that the GTVH is more developed than any other theory, but still "no more than a very early draft". The GTVH identifies several aspects of humour. In the terms of the GTVH, this paper identifies a collection of forms in the *Logical Mechanism* of jokes. We acknowledge the importance of the other aspects of the GTVH, but do not attempt to model them here.

The most concrete work on computational humour is that done by Kim Binstead and Graham Ritchie. They looked at jokes based on ambiguity, and identified puns as a computationally tractable class of jokes (given current electronic resources).[1] This resulted in the JAPE pun-generator, which produces riddle-puns that fit into the incongruity-resolution theory. The jokes we examine do not fit under the incongruity-resolution theory, but do fit under Koestler's more general theory. They use a quite abstract type of perspective shift: from normal communication to poor communication.

3 Intentionally Poor Communication

Wit is a verbal form of humour, which suggests there may be connections to dialogue forms. We will explore here the idea that one class of wit (c.f. §1.1) is related to *poor speech*. Initially this seems strange: Wit is perhaps the very opposite of poor speech. Yet as we will show, it often involves breaking the rules of good speech. Wit is made challenging by the constraint that the utterance should be surprisingly poor. Typically this is achieved by initially looking normal, whilst signalling an intention to be ridiculous (i.e. wit consists of cleverly disguised obviously poor speech). Parallels exist here with deliberate comic stupidity, and the deliberate clumsiness of clowning.

This intention forms an interesting but vague constraint. It can be conveyed in a wide range of ways: by tone of voice (e.g. sarcasm), by context, by the character of the speaker (or in the case of mockery, the character of the target), by contrast with similar well-known utterances, by cross-reference, by self-reference, or by the artfulness of the joke's construction. The unifying factor seems to be that all of these signs distinguish the joke from a genuinely poor utterance. Note that context and character loosely correspond to the GTVH elements of *Situation* and *Target*. Situation and target also have an important effect on how funny a joke is (and to whom), which we will not examine here.

3.1 Examples: Breaking Grice's maxims

Grice's maxims are a well-respected list of rules for speech.[4] They represent an unwritten contract between speaker and listener, and are used in natural-language processing to explain the selection of what to say and what to omit, and how the listener arrives at the correct interpretation. Below we list Grice's maxims, plus example jokes that break them.

Maxim of Quantity:

1. Make your contribution to the conversation as informative as necessary.

Example joke that breaks this rule:

> Pres: You understand the reasons for absolute secrecy?
>
> Agent: Yes – but can you tell me *nothing at all* about my mission?[5]

Here intention is conveyed by the similarity to a familiar situation (secret agent adventures).

2. Do not make your contribution to the conversation more informative than necessary.

> Jewish Mother: Help, help! My son, the doctor is drowning![6]

Note how intention is conveyed in this joke by the targeting of a well-known stereotype.

Maxim of Quality:

1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

The Maxim of Quality is out of place in Grice's Maxims, since it is not a rule of conversation so much as a rule of good behaviour. Nevertheless, there are jokes based around violating this maxim. Some use contradictions, others use obvious lies:

> Farmer: Who's out there?
>
> Chicken Thief: Nobody here but us chickens.[6]

[5] Donald Westlake

[6] Source unknown.

> Captain Renault: What brought you to Casablanca?
>
> Rick: My health. I came to Casablanca for the waters.
>
> Captain Renault: The waters? What waters? We're in the desert.
>
> Rick: I was misinformed. [7]

There is a type of joke resembling this 2nd example which should be easy to generate. If an A.I. agent is given a question where either the answer is obvious or the agent does not wish to answer, then a sarcastic false answer will constitute a witty reply, especially if nonsensical. E.g.

> Q: Are you a machine?
> A: No I'm a new type of flamingo.

Nonsense of this form should be easy to generate; the only hard constraints are that the answer be of the correct grammatical type (e.g. Noun Phrase or Verb Phrase, depending on the question), and not be relevant to the question (which will usually be the case for a randomly generated utterance, but can also be checked using a sentence comparison technique such as HAL or LSA). Intention is shown by both the sarcasm and the context (which allows listeners to understand that the speaker is being deliberately difficult).

Maxim of Relevance: Be relevant

Plenty of humour involves non-sequiturs, e.g.:

> How many surrealists does it take to change a lightbulb? Giraffe.[6]

One large and important category in this area is that of *nonsense jokes*. Typically, these jokes consist of nonsensical happenings within a sensible framework. For example:

> A man at the dinner table dipped his hands in the mayonnaise and then ran them through his hair. When his neighbour looked astonished, the man apologised: "I'm so sorry. I thought it was spinach."[8]

Here, the 'strange act - apology' framework is sensible; but the apology is nonsense. Computer generation of this category could be difficult, since typically these jokes involves considerable common-sense knowledge (in order to recognise what is genuinely nonsensical). However it may be that tractable subsets can be identified.

Maxim of Manner:

1. Avoid obscurity of expression.

Euphemisms (obscure references to unpleasant/taboo subjects) are a reliable source of humour, although most such jokes are too obscene to reproduce here. Government euphemisms are also commonly mocked.

> Civil Servant: 'The matter is under consideration' means we have lost the file. 'The matter is under active consideration' means we are trying to find the file.[9]

[7] *Casablanca*, M.Burnett & J.Alison.

[8] Reported by Freud[3].

[9] *Yes Minister*, Jonathan Lynn

2. Avoid ambiguity.

Several forms of ambiguity exist, with related jokes: aural (e.g. puns – a form of humour explored in [1]), lexical (e.g. A man walked into a bar. Ouch.[6]), grammatical parsing (e.g. Every minute, somewhere in the world a woman gives birth. We must find this woman & stop her.[6], & anaphora resolution (e.g. in slapstick comedy: When I nod my head hit it.[6]). Expectation failure and surprise only occur if one reading is quite likely and the other highly unlikely. This makes ambiguity jokes hard to generate reliably, since we cannot usually judge the relative plausibility of statements (although one could generate candidate jokes and use a human operator to filter them).

3. Be brief (avoid unnecessary wordiness).

Plenty of jokes are based on unnecessary wordiness. Most commonly this in the form of long lists (e.g. Monty Python's famous parrot sketch rant), but can also be used as a way of caricaturing certain archetypes (e.g. bureaucrats).

4. Be orderly.

> Melchit: Let's play a word game.
>
> Blackadder: O.K. Rearrange the following words to make a sentence: Face, Your, Sodding, Shut. [10]

4 A Logic for Wit

Based on the idea of intentionally poor communication, we now develop a set of logical forms that can produce jokes. We adopt a semi-formal presentation here (i.e. using logical expressions, but without formally defining all the concepts used). It would be premature at this stage to develop a rigid formalism, as this would probably require adaptation after further investigation. However, Grice's maxims are too informal to be used in computational work without some formalisation. We will instead work with a more formal .

4.1 Resources

We note the existence of the following resources:

- Online thesauruses, allowing synonyms and antonyms to be found.
- WordNet an augmented thesaurus, giving relations such as 'X is a kind of Y', 'X is a part-of Y'.[7]
- Corpuses and statistical corpus analysis tools. These include co-occurence analysis techniques such as LSA[8] or HAL[9], and word-scores such as those provided by the MRC[2].
- We will assume the availability of a parser capable of converting between logical and natural language forms. This allows us to focus on logical forms. The issue of surface language is important, but is largely independent of the logical analyses we present here.
- We will assume the availability of an inference engine capable of 'common sense reasoning'.

Only limited versions are actually available of the last two resources. The success of joke generation programs will depend to a fair extent on how effectively these limited systems can be utilised.

[10] *Blackadder II*, B.Elton & R.Curtis.

4.2 Notation

For simplicity, we use true/false predicates here. A more sophisticated approach might be to use fuzzy logic, allowing utterances to be slightly poor & slightly funny.

- We introduce the relation *suggests*, written $A \rightsquigarrow B$, & interpreted as "proposition A suggests proposition B". This a weaker form of *implies*, used to capture the misdirection which is key to many jokes. In general determining $A \rightsquigarrow B$ requires real-world human knowledge, and is therefore beyond current A.I. However we will identify cases where $A \rightsquigarrow B$ depends only on the logical form of the joke itself, and these cases form the heart of this project.
- According to the theory of perspective shifts, order is important, and this can be seen in the setup/punchline structure of many jokes. Since the order in which a joke is told may not be the same as the ordering of elements in the logical form, we introduce an ordering relation $A - B$ for "B follows A" or "B is deduced after A". Note that A, B need not be well formed formulae.
- Define an *utterance skeleton* to be an ordered set of formulae fragments, e.g. $J = A - B - C$.
- Define a *joke form* to be an utterance skeleton that gives rise to jokes.
- Write $\Gamma_S(t)$ for the speaker's knowledge at time t, $\Gamma_L(t)$ for the listener's knowledge (or rather, the speaker's model of the listener's knowledge. We will allow knowledge to include counter-factual and hypothetical information. If an utterance is ordered $A - B$, then we have $A \in \Gamma_L$ when B is said.
- Define a measure for the complexity of an utterance: $\texttt{complexity} : utterances \rightarrow \mathbb{R}^+$. Complexity can be approximated by sentence length. A more sophisticated measure would take into account the obscurity of the words (measurable via corpus analysis) and the non-linearity of the parse-tree.
- Define a predicate $\texttt{deliberately}{-}\texttt{poor}(X)$. As noted in §3, intention can be conveyed in many ways. We do not have a theory for determining $\texttt{deliberately}{-}\texttt{poor}(X)$. Instead we present a collection of specific situations where it holds.
- Whilst we will talk of logical relations and inferences, we must take into account the speed and difficulty of such calculations. We write $!P$ to mean 'P can be calculated very quickly (i.e. in the flow of conversation) by a normal person'. Initially, let us assume that $!\exists S$ holds if $\texttt{complexity}(S)$ is small, and $!A \models B$ if there is a short proof (e.g. less than 5 steps) in a 'suitable logic'.

4.3 The maxim of normal communication

We suggest the following axiom is true for normal speech acts. It can be seen as a mixture of elements of Grice's maxims on quantity plus something of the maxims on Manner.

S is a poor utterance given Γ_S, Γ_L, if $!\exists S'$ such that $!\Gamma_S \models S'$ (i.e. S' is true) and $!\Gamma_L, S' \models S, \texttt{complexity}(S') \ll \texttt{complexity}(S)$

S is a poor utterance given given Γ_S, Γ_L, if $!\exists S', S''$ such that $!\Gamma_S \models S'$ and $!\Gamma_L, S' \models S, S'', \texttt{complexity}(S') \leq \texttt{complexity}(S), \texttt{relevant}(S'')$

Maxim of Normal Communication: In normal communication, poor utterances only occur by mistake.

This maxim is useful because it identifies utterances that create expectations which we can reliably predict without sophisticated world knowledge:

Corollary: If A is a poor utterance given $B \in \Gamma_S$, then $A \rightsquigarrow B \notin \Gamma_S$.

4.4 An axiom of wit

We will assume here that there are only three forms of speech act: information exchange, mistakes, and jokes. This means that deliberately poor utterances are jokes. I.e., given an utterance J, if J is a poor utterance and $\texttt{deliberately}-\texttt{poor}(J)$, then J is a joke. Using this rule, we now identify a number of joke forms. How *funny* a joke is will depend on several aspects (e.g. target, situation, & surface language) which we do not examine here.

5 Joke Forms

We now identify a number of joke forms which break the axiom of normal communication.

5.1 Obvious tautologies

Simple tautologies[11] are a clear violation of the Maxim of Normal Communication – indeed they are such an obvious violation as to almost always be intentional. The tautology can be very simple indeed, if the joke is being used to mock a stupid character. E.g.

> Football Manager: 'Cause football is about nothing, unless it's about something. And what it is about -
>
> Interviewer: Yes, yes?
>
> Football Manager: Is football. [12]

Stating definitions

One form of tautology-based joke consists of a statement $S = P - Q$, where Q is part of the definition of P, and hence Q is a tautology given that P has already been stated. Examples include:

> It is dangerous to make predictions, especially about the future.[13]

> Almost all acts of self-abuse are committed by someone who knows the victim, and these cases are hardly ever reported.[14]

Note the presence of parody, quantifier abuse ('almost all' used instead of 'all' – c.f. §5.2) and taboo-breaking in this last joke.

[11] I.e. tautologies which can be recognised as such almost instantaneously.

[12] Peter Cook

[13] S.Gorn.

[14] C.Beecham.

Generation: It should be possible to generate such utterances from dictionary definitions. E.g. From "humor, n. 1. The quality that makes something laughable or amusing; funniness."[15], we could produce "I like humour, especially when it's laughable, amusing or funny". Doing this automatically involves two stages: Identifying a suitable fragment of the definition (which might be done by using the MRC to identify keywords), and constructing an utterance (which could either be done via a set of templates, or (better) a small generative grammar for this task). A closely related form is $S = P \text{if} Q$ where $Q \in \texttt{synonyms}(P)$, as in the joke:

> "I could have been smart too if i'd had the brains."

These jokes should be simple to generate, given either a template or grammar system for generating the surface form.

To be, or not to be?

Another form of obvious tautology is $S = P \vee \neg P$, as in Badly Drawn Boy's non-observation:

> Sometimes it feels like my soul is being sucked out through my arse. And sometimes, it doesn't.

. Note that $P \vee \neg P$ can be a speech act meaning "I do not know". This means it is not necessarily a poor utterance, unless $\texttt{complexity}(P \vee \neg P) \ll \texttt{complexity}$("I do not know").

Generation:

Utterances of this form can be created by simply appending a negation to any utterance P where P is reasonably complex (so that $\texttt{complexity}(P \vee \neg P) \ll \texttt{complexity}$("I do not know")). Given the simplicity of this generation mechanism, we might expect (from the showing-off theory of humour) that other factors will be required to make such jokes genuinely funny.

5.2 Quantifier abuse

Forall nothing

The utterance S = '$\forall x \in X$' suggests $\exists x \in X$. This is known as 'existential import'. Although modern logic has opted against it, everyday language is ambiguous on the point: $\forall x \in \emptyset$ is considered technically correct, but odd. Indeed, a universally quantified statement strongly suggests the existence of *multiple* objects. We can derive this suggestion from the Maxim of Normal Communication If $X = \emptyset \in \Gamma_S$ then let $S' = S'' =$'$X = \emptyset$'. We have $\texttt{complexity}(S') \leq \texttt{complexity}(S)$, and we can assume $\texttt{relevant}(S'')$. Similarly for $X = x$, stating this would normally be more informative, and hence should be preferred.

Breaking this suggestion gives rise to joke form $J = \forall x \in X.P(x)$ where $!X \models |X| < 2$, with the parts ordered $\forall x - P(x) - \in X$ in the utterance. Examples include Henry Ford describing the choice in cars he offered ("Any colour you like as long as it's black"), or

> Cat: I was with you right up until the beginning.[16]

Note that if the ordering of the joke fragments was changed to $\forall x - \in X - P(x)$ then we would still have a poor utterance

[15] First definition returned by `www.dictionary.com`.

[16] *Red Dwarf*, Grant & Naylor.

for $X = \emptyset$, but of a different type. It would now be a tautology (since for any P, $X = \emptyset \Rightarrow \forall x \in X.P(x)$). However this would also give jokes (c.f. §5.1). Example instances of this form of tautology include:

> They had nothing. But they were willing to risk it all.[17]

> Free advice is worth every penny.[6]

Recognition / Generation: Computer recognition of either form of joke will be limited by the limitations of current parsers, but should otherwise be straightforward. Computer generation may be harder, requiring some method for generating situations where these utterances can occur. Generation will largely be the same for both the expectation-confounding ($\forall x - \in \emptyset - P(x)$) and the tautology ($\forall x - \in \emptyset - P(x)$) form.

Instances of X where $|X| = 1$ are easy to create. Instances of $X = \emptyset$ can be created using 'basic terms' for $\emptyset$: "nothing", "no-one". They can also be made using A, B such that $A \Rightarrow \not B$, since $\emptyset = \{x' \text{s.t.} A(x) \& B(x)\}$. Such pairs of contradictory predicates can be generated from antonyms. If a related third predicate $C(x)$ can be found, then the interconnection may make the resulting joke (either $\forall x - C(x) - x \in \{x' \text{s.t.} A(x) \& B(x)\}$ or $X = \{x' \text{s.t.} A(x) \& B(x)\}.\forall x \in X.C(x)$) funnier.

Alternatively, in an A.I. agent, generation could be done opportunistically when the conversation presents a suitable situation.[18]

Giant exceptions

The statement "Every x but y" suggests that the positive instances greatly outnumber (or are more important than) the negative ones. This is closely related to the previous class of jokes, and can also be derived from the Maxim of Normal Communication. Violating it gives rise to the joke form: $J = \forall x \in X/E$ where $|E| \sim |X|$. Examples include:

> Oscar Wilde: I can resist anything but temptation.

> Doctor: Apart from being dead, there's nothing wrong with him.[19]

> Lewis Carroll: I am fond of children - except boys

> Everyone agreed, apart from the diehard misogynists (typically 74% of the population).[20]

Generation / Recognition: These jokes require more world knowledge than the 'forall nothing' class in order to judge relative sizes / importance. Generation can begin with the larger category (the main set), and then try to recognise the possible subsets of this category (e.g. start with 'children', and deduce children = boys + girls). WordNet has a relation (hyponyms) for 'X is a kind of Y'. Unfortunately this is erratic, including many obscure entries (e.g. 'changeling' is a type of child), but the results could be filtered against frequency of occurrence in a corpus. Dictionary definitions could also be used (e.g. using `www.dictionary.com`, which combines several dictionaries, gives 'person', 'son', & 'daughter' as the nouns referenced by more than one dictionary when describing 'children'). Generation could also begin with the exception, and find a superset using WordNet's hypernyms relation (e.g. hypernyms of 'boy' gives 'male' followed by 'person' as enclosing categories). Given a main set and a subset, we can test for the relative size of the subset using frequency of occurrence in a corpus.

An alternative algorithm involves lists of cases. If the size of a list is comparable to the number of all possible list entries, then presenting it as a list of exceptions should constitute a joke. This situation could arise in an AI agent reporting problems (e.g. a mechanic agent reporting faulty car parts might say "There is nothing wrong with this car apart from the engine, transmission, brakes & stereo.").

Exists everything

Statements of the form $\exists x \in X.P(x)$ suggest $\not\forall x \in X.P(x)$. Our theory therefore suggests jokes of the form $J = \exists x \in X.P(x)$ where $!\forall x \in X.P(x)$. This seems to be a rarer class. One example of such a joke is:

> Politician: ...which will result in nuclear holocaust. Now some people might say that's a bad thing... [21]

5.3 Contradictions

Stating $P - Q$ where $!P \Rightarrow \neg Q$ is a poor utterance. We can derive this from the Maxim of Normal Communication using $\perp \Rightarrow P \& \neg P$ and $\texttt{complexity}(\perp) \ll \texttt{complexity}(P \& \neg P)$ – especially if $\texttt{complexity}(P)$ is large. It is also a violation of Grice's maxim of quality (c.f. §3.1). The finest examples of this class are those jokes that are both contradictory and self-justifying:

> Saul Gorn: I am a firm believer in optimism, because without optimism, what else is there?

> There are 3 types of mathematician: those who can count and those who can't.[6]

Unfortunately, they are also likely to be very hard to generate.

Oxymorons

A much easier class of contradiction is the use of oxymorons – when a modifier is combined with a contradictory object. E.g.

> Tom Lehrer: The Harvard mathematics department was a hotbed of celibacy.

Generation: Given a noun P, we can generate an oxymoron by selecting $Q \in \texttt{antonyms}(P)$ then finding adjective A such that $A - Q$ is common, but $A - P$ is either nonexistent or very rare (this would be done using a corpus). The sentence fragment $A - P$ should be an oxymoron. Alternatively, we could find adjectives A, B such that $B \in \texttt{antonyms}(A)$ and $B - P$ is common whilst $A - P$ is rare. These two methods are basically equivalent, but likely to return different utterances, giving a wider range for selection. Oxymorons involving verbs

[17] Tagline for *The Commitments*, Roddy Doyle.

[18] Which is how much of humour occurs, and makes the humour funnier – c.f. §6.5.

[19] *Day of the Tentacle*, published by LucasArts.

[20] *The Algebraist*, Iain Banks

[21] *Not The 9 O'Clock News*, BBC.

and adverbs can similarly be generated. We suspect that oxymorons are always recognisable as jokes rather than mistakes, so signalling intention is not necessary.

Conjunction of conflicting statements

Another form of joke, closely related to contradiction is the juxtaposition of conflicting statements. For example:

> The Texan turned out to be good-natured, generous and likable. In three days no one could stand him.[22]

> Nately had a bad start. He came from a good family.[2]2

Conflicting utterances should be straightforward to generate. However showing that such statements are not mistakes may be harder. In the examples above, what identifies these utterances as jokes rather than junk is that, although the statements are in conflict, they can be resolved to give a satisfactory explanation. This is likely to be difficult to formalise, and so it may not be possible to reliably generate such jokes.

5.4 Convoluted statement

There are jokes based around breaking the Gricean axioms of manner: (1) "Never use a big word when a diminutive alternative would suffice.", (2) "Don't more use words than necessary; it's highly superfluous"[6] (note that both of these maxims can be deduced from the Maxim of Normal Communication).

This is an interesting class of jokes as it seems particularly well suited to computerisation. However such utterances can be mistaken for the accidental overly-detailed statements of a bore. They are therefore unlikely to be funny unless the humorous intention can clearly be conveyed. This can be done if (a) the utterance is delivered by an agent playing up to some suitable caricature (e.g. an officious bureaucrat, or a pompous pseudo-intellectual), or (b) the convolution is pushed to extremes (limited by producing an utterance that is still intelligible to the normal listener). This second case is particularly applicable when trying to express what the speaker feels is an important point to someone who is not listening.

Generation

Excessive detail can be generated by replacing generic terms with specific ones (found via WordNet), and/or inserting extra adjective phrases. The simple tactic of replacing a single adjective with a string of synonyms should be quite effective. Obscure wording can be created by using a thesaurus to find synonyms, ranked by frequency-in-a-corpus.

5.5 Double negatives

Stating $\neg\neg P$ for P is a clear breach of normal communication, but it is also rather artificial unless disguised. Sarcasm is one such disguise.

Generation: We have implemented a system for converting between normal and sarcastic comments in simple sentences. This uses a shallow parse to identify the adjective in a phrase. A thesaurus is then used to find antonyms, which are ranked using the *imageablity* rating from the MRC Psycholinguistic Database. Preliminary tests suggest this is quite effective. For example, "You look terrible" becomes "[sarcastic] You look wonderful". Conversely, the same process will produce un-sarcastic versions of sarcastic statements. Here the meaningfulness rating is more appropriate for determining word-choice. With this, "[sarcastic] You look wonderful" becomes "You look poor". Preliminary experiments suggest that this system is quite robust for simple sentences and common terms. However the MRC scoring was needed to avoid the selection of obscure/weak substitutions.

Inappropriate metaphors

There is a more sophisticated form of double-negative, where a metaphors uses a negative example to imply a negative statement. E.g. using "savage as a gerbil" for "not savage". A good, if unrepeatable, example of this is Douglas Adam's meta-metaphor: "The big yellow somethings hung in the air in almost exactly the way that bricks don't."

Generation: Inappropriate metaphors could be generated using the following process. Given the target statement $\neg P$ (where P is an adjective):

- Find an antonym Q of P using a thesaurus.
- Find an exemplar[23] X of Q. Finding exemplars should be possible using co-occurrence techniques and a dictionary to identify nouns. For example, LSA – which links words that occur in similar contexts and can list the most-strongly linked – should be appropriate here.

  These two steps will give many possible choices for X. One way of rating examples would be to use the MRC meaningfulness score for Q, the strength of link between X and Q, and the MRC imageablity score of X – combined by multiplication.
- This gives the humorous description "As P as an X".

6 Future Work

6.1 Implementation & testing

The obvious next stage in developing this work is to implement the algorithms we describe. This would allow them to be tested. Hopefully this would validate the ideas we have presented – but however successful an implementation is, it is sure to suggest areas for improvement.

6.2 More joke forms

We have identified a set of joke forms based around intentionally poor communication. However it is unlikely that this is an exhaustive set. Further exploration of the idea of intentionally poor communication will probably lead to more such joke forms being identified.

6.3 Formalisation

If implementation and testing are successful, then we should try to formalise the theory. In particular, it would be interesting to try and automatically derive the individual joke forms from a formal statement of the idea of intentionally poor communication. This could also prompt the identification of new joke forms.

[22] *Catch 22*, J.Heller.

[23] A typical example.

6.4 A richer theory of discourse

In this paper, we assumed that there were only three types of speech acts (information exchange, mistakes, and jokes). A richer theory of discourse (i.e. one containing more speech acts, and relating these to human interaction) would allow us to filter out some non-jokes, and should allow us to identify more joke forms.

6.5 Measuring funniness

In this paper we have only considered what makes an utterance into a joke. We have not examined what makes some jokes funnier than others. Developing a metric to measure the funniness of a joke will be important to most applications of computational humour. There are many factors involved in this – including some that are personal to an individual. Below we mention three factors that seem potentially well-suited to computer modelling.

Relevance

Wit is characteristically spontaneous, generated in response to the flow of conversation, and is considerably funnier as a result. This phenomenon can be seen as a result of the showing-off theory of humour (c.f. §2.1). This suggests that wit should be generated within situations, rather than presented as stand-alone jokes, and it's funniness can partly be judged by the relevance of the witty comment to the current context. This can be measured by using a co-occurence technique to compare the witty statement with recent conversation.

Taboos

Jokes that break social taboos are generally funnier, albeit riskier (e.g. an absurd *rude* adjective will probably be funnier than a merely absurd adjective). Although taboos are very complex psychological phenomenon (varying strongly between cultures), they may be reasonably susceptible to crude automated metrics. Certain subjects are reliably shrouded in taboos: sexual affairs, the scatalogical (toilet-related) and violence. These subjects typically use an identifiable set of words/concepts. Hence a simple metric for whether a joke is taboo-breaking would simply look for taboo-related keywords.

This also suggests a simple formulae for generating taboo-breaking concepts: link a taboo term to an anti-taboo term (i.e. a symbol of purity): e.g. "nun-punching", "kick-boxing granny", "bishop-shit", etc.

Parody

An element of parody (mimicking a well-known person, situation or form of utterance) can greatly strengthen a joke's appeal. This is for several reasons. Parody strengthens false-suggestions, emphasises the absurd (by providing a contrasting real-world example), increases the cleverness of a joke, and mocks the original person/situation/utterance. Although it is ambitious, we might hope to use statistical corpus analysis to both recognise and generate parody.

7 Conclusion

We have found a relatively abstract form of humour that is potentially computationally tractable. This is the class of jokes and witty statements which draw their humour from deliberately breaking the rules of normal conversation – a class of humour we summarise as *ingeniously poor communication*. Our analysis is based on the perspective-shift theory of jokes. We present an overarching framework for this: that jokes emerged as a means of publicly displaying intelligence. This can be justified on evolutionary grounds, as an instance of mate selection forces.

The work presented here is in a preliminary stage, and there is clearly much more to be done. Hopefully the algorithm schemes we have described here will prove successful when implemented. If so this work will considerably extend the range of computational humour. It contains many limitations. Yet if we are optimistic, these too may advance the field by spurring future developments. After the pioneering work of Binstead & Ritchie on puns, creative computational humour has languished. This is a shame, as it is clearly an area of great interest and possibility. Perhaps more than anything, what this field needs is fresh work.

References

[1] K.Binstead & G.Ritchie "Computational rules for generating punning riddles" in Humor: the international journal of humor research, 1997.

[2] M.Coltheart "The MRC Psycholinguistic Database" in the Quarterly Journal of Experimental Psychology, 1981.

[3] S.Freud "Jokes and their Relation to the Unconscious", 1905.

[4] H.P.Grice "Logic and Conversation" in Logic and Grammar, 1975.

[5] I.Kant "The Critique of Judgement", 1790

[6] A.Koestler "The Act of Creation", 1964.

[7] C.Fellbaum (editor) "WordNet: An Electronic Lexical Database", MIT Press, 1998.

[8] T.Landauer, P.Foltz, & D.Laham "Introduction to Latent Semantic Analysis" in Discourse Processes, 1998.

[9] K.Lund "Through Context to Global Co-occurrence and Meaning: The HAL Odyssey" presented at the Psychonomic Society Symposium, 1997.

[10] M.Minsky "Jokes and their Relation to the Cognitive Unconscious" in Cognitive Constraints on Communication, 1981.

[11] M.P.Mulder & A.Nijholt "Humour Research: State of the Art", 2002.

[12] A.Pope "An Essay On Criticism, part 2", 1711.

[13] R.R.Provine "Laughter punctuates speech: Linguistic sopcial and gender contexts of laughter", published in Ethology, 1993.

[14] G.Ritchie "Current directions in computational humour" in A.I. Review, 2001.

[15] H.Spencer "The physiology of laughter", 1860.

[16] A.Zahavi "Mate selection: a selection for a handicap" Journal of Theoretical Biology, 1975.