

Cleverness versus funniness

Greg Lessard
French Studies
Queen's University

Michael Levison
Computing and Information Science
Queen's University

Chris Venour
Computing and Information Science
Queen's University

Abstract

The traditional distinctions between verbal and non-verbal humour and between lexical and encyclopedic bases of humour are examined by means of the analysis of corpus and generated data and shown to be more complex than is usually assumed. Tendencies found in the data analysed are used as the basis for proposing a distinction between cleverness and funniness, the first based on the factor of surprise, the second based on script opposition, as proposed by Raskin. It is suggested that use of the distinction would be of value in the evaluation of computer-generated humour.

Keywords: Humour, Puns, Natural Language Generation

1 INTRODUCTION

Traditionally, humour researchers have accepted a distinction between verbal and non-verbal humour, based on criteria such as the use of linguistic materials versus the use of gestures, images, etc., the possibility or impossibility of alternate formulations, ease or difficulty of translation, and cultural contingency versus non-specificity of culture. A prototypical example of verbal humour is provided by anagrams of the simplest sort, in which the letters of a word are reordered to form another word of the same language. Use of the device is strictly language-internal and untranslatable. A prototypical example of non-verbal humour is provided by the slapstick routine based on a man in the middle who holds a ladder and who, when turning it, hits first one then the other of two men preceding and following him. No verbal devices are required to understand this and it is in principle capable of being found funny by a human speaking any language.

Another facet of the same distinction turns on the opposition between verbal humour based on manipulation of lexical relations between words, versus that which makes use of encyclopedic or 'real world' knowledge. Examples of the first are found in riddles based on single word substitutions, such as in the following, which make use of relations of homonymy or synonymy:

What do you call a naked bruin? A bare bear.
What has a tongue and can't speak? A shoe.

Examples of the second can be found in other riddles which require more complex knowledge of how the world works, such as:

What did the boy octopus say to the girl octopus? I want hold your hand, hand, ...
How many Xs does it take to change a light bulb? Ten, one to hold the bulb and nine to turn the ladder.

as well as in jokes of the more traditional sort.

Verbal humour of the first kind would represent an in principle enumerable (albeit large) set, based on the systematic enumeration of all the lexical relations holding between all items of a given lexicon, while verbal humour of the second kind would represent an open set relying on serendipity, the ability to creatively establish links based on world knowledge, or the ability to construct a humorous scenario.

Taken together, these two distinctions suggest a difference in kind between basic forms of verbal humour such as puns and simple riddles, which turn on lexical relations, and more complex phenomena such as jokes, which turn on real-world or encyclopedic knowledge. However, categories leak, and meaning appears to be one of the most corrosive substances in causing such leaks. Thus, in the case of anagrams, a variant exists in which the re-ordered letters form an explicit or implicit judgement on the real-world entity referred to by the base form. The following examples, taken from www.genius2000.com, illustrate the mechanism:

Houses of Parliament – Meet piranhas so foul
House of Commons – O home of honest scum
Tony Blair MP – I’m Tory Plan B

In these examples, there exists a semantic (and evaluative) relation between the base form and the anagram and to a limited extent a ‘script opposition’ between the two terms. Similarly, the slapstick routine mentioned above takes on new significance if the characters hit by the ladder are one to the left of the ladder-carrier and one to the right, particularly if their clothing marks them as prototypically members of a political group. In fact, despite their formal differences, both the anagrams and the slapstick routine can come to carry the same message: ‘the indistinguishability (and low value) of political figures or parties’.

These simultaneous differences and similarities suggest a problem for the analysis of humour. Should we treat anagrams, puns, and jokes as more or less complex points along a single dimension, and measure their effect in terms of the factor of ‘funniness’, or should we see a difference in kind between basic verbal humour and more complex forms of humour such as that found in jokes? The question has some importance for the computational generation of humour, since presumably a difference in kind would also require a difference of approach. To gain a clearer sense of the details involved, we will begin by studying in more detail two sets of examples, one a corpus of human-generated puns, and the other a corpus of computer-generated puns.

2 TWO SETS OF EXAMPLES

The first set of examples is drawn from a web page (<http://thinks.com/words/tomswift.htm>) which provides 374 examples of Tom Swifities, a type of pun based on a quoted utterance having both a formal and semantic link to the quoting utterance, as in:

I hate chemistry, said Tom acidly.

In what follows, we will use the terminology proposed in Lessard and Levison (1992, 1993, 1995) to identify the parts of a Tom Swifty: the **pivot** represents the element of the quoting sentence which describes the nature of Tom’s affirmation (in the example above, *acidly*), the **base** represents the subset of the pivot which enters into a semantic link with one or more elements of the quoted utterance (here, the base is *acid*) and the **target** represents the element or elements of the quoted utterance linked to the base (here, *chemistry*, which is linked to *acid*). The relation between base and pivot is called the **formal bridge**, while that between base and target is called the **semantic bridge**.

There are actually a number of formal variants on the canonical Tom Swifty, including those where the pivot is represented by another part of speech:

I think I’ve caught one, said Tom with *baited* breath. (cf. bated breath)
I have to cut the grass again, Tom *moaned*. (cf. mow-ned)

but all share the same essential characteristics.

As a first step in the analysis, the collection from thinks.com was tagged with a number of categories, including the following:

Number of bridges

m = multiple semantic bridges

The prisoner escaped down a rope, said Tom *condescendingly*. (prisoner – con, down – descend)

Types of simple lexico-semantic relations (see Cruse 1986)

syn = a semantic bridge based on synonymy

I have no idea, said Tom *thoughtlessly*. (idea – thought)

ant = a semantic bridge based on antonymy, of whatever subtype

You must be my host, Tom *guessed*. (host – guest)

hyp = a semantic bridge based on hyponymy

This tuna is excellent, said Tom *superficially*. (tuna – fish)

mer = a semantic bridge based on meronymy (a part-whole relation)

I've still got two fingers left, said Tom *handsomely*. (finger – hand)

seq = sequential or cyclic relations

We've just brought gold and frankincense, the Magi *demurred*. (gold, frankincense, myrrh)

Types of complex lexical or semantic relations

p = a paraphrase in which one word of the base is linked to a phrase or sentence of the target

It's where we store the hay, Tom said *loftily*. (where we store the hay = loft)

enc = a relation in based on world knowledge

Do you play the guitar?, Tom asked *callously* (playing guitar may produce callouses on the fingers)

npr = a relation in which understanding of the Tom Swifty depends on knowledge of the characteristics of an individual:

Henry the Eighth, Tom said *unthinkingly*. (un-thin-king)

One upon a time there was a beautiful princess, Tom began *grimly*. (Grimms' fairy tales)

In the first instance, it is the distinction between simple and complex lexical relations which will concern us. We will assume that the first (synonymy, antonymy, etc.) are lexicalized and thus relatively static, as well as being language-specific. Complex lexical relations such as encyclopedic links or relations based on paraphrase or traits associated with proper names, on the other hand, are not lexicalized, but rather produced dynamically, and are not language-specific, in the sense that many different formulations may be imagined. Compare, for example, the following variants of the examples appearing above:

Look above the stable, Tom said loftily.

I like the spacious new apartment, Tom said loftily.

I'm a hardworking stonemason, said Tom callously.

And they lived happily ever after, said Tom grimly.

Once upon a time there were two brothers, said Tom grimly.

Analysis of the 374 examples of Tom Swifties in the thinks.com corpus reveals the following distribution of the types:

enc	92 (25%)
npr	38 (10%)
p	52 (14%)
simple	192 (51%)

It is likely that the npr class should be considered a subset of the enc class, and the relationship between enc and p remains to be elicited. Be that as it may, it remains that between a quarter and a half of the examples cannot be constructed on the basis of simple lexical links. And yet, these encyclopedic, paraphrase and proper name Tom Swifties are interspersed among the others with no sense of a difference in kind. This suggests that an adequate computational model of Tom

Swifties should not limit itself to lexical links but must seamlessly move from lexical, to proper name, to paraphrase and to encyclopedic generation devices.

Consider now a second corpus of puns generated and then analysed in Venour (1999). Known as homonym common phrase puns (HCPP) these make use of idioms ("kick the habit", "pass the buck", "jump ship") or collocations ("knead the dough", "serial killer", "tip the waiter") which contain a homophone. Some typical examples:

John is violent. He is razing cattle.

John ate a dollar coin. He is passing the buck.

As is the case with Tom Swifties, HCPPs can be formally defined in terms of a target phrase, a base, and a pivot. The pivot in the first example above is the common phrase "raising cattle", the base is "raising", whose homophonous meaning "razing" (defined as "completely destroying") is reinforced by the target phrase "John is violent". (The word "raising" actually has a second homophone: "raising" in the sense of lifting up, which might be used to create the pun "John is lifting up cows. He raises cattle.")

The HCPPs discussed in Venour were generated using a version of the VINCI natural language generation environment (Levison and Lessard, 1992; and more generally www.cs.queensu.ca/CompLing). Prior to generation, the following procedures were used to construct the relations which underlie the puns. Note in passing that in all cases, care was taken to ensure that the lexicon would be "general and neutral", i.e. not humour-specific (Binsted and Ritchie, 1994).

a) As a first step, seventy adjectives with noun homophones and seventy nouns with noun homophones were taken from a list on the web (Cooper, 1999). Since these contained only homophones spelled differently, they were supplemented from another source (Franklyn, 1966). Sixteen words were obscure enough that the persons judging the puns would be unlikely to recognize them. These were removed.

b) The second step involved picking collocations of these words from the Oxford English dictionary. Each of the homophones was looked up, and the first common phrase agreeing with our syntax requirements was chosen. If a common phrase with the proper syntax could not be found for a word or its homophone, both words were deleted. The resulting phrases were added to the lexicon as units, as were the individual words of the phrases.

c) For each of the words, volunteers were given a questionnaire and asked to provide words related to the various nouns, adjectives and verbs in the lexicon. It is interesting to note the diversity which this produced. Thus, apart from the usual basic lexical relations like synonymy, antonymy, meronymy, etc., we find encyclopedic relationships such as *sailor – pier*, *diver – coral*, *grave – bier*. In the first two cases, the relationship between the two words is based on typical locations associated with the first term, while in the third example, the relationship is based on shared membership in a common semantic domain (cemetery and burial rituals and objects).

d) In a last step, other words were added to the lexicon, including determiners, conjunctions, prepositions, etc. called for by the syntax specifications.

After the addition of morphological and syntactic specifications, the VINCI system was used to generate possible HCPPs. Of the 11 HCPP schemata identified by Venour, three were used. In all, 50 jokes were generated. These were distributed to 16 volunteer judges, each receiving either the first or the second half of the list. One volunteer did not follow the instructions and that reply set was discarded. Thus, 375 votes were cast in total. The judges were asked to evaluate each joke on the following scale, taken from Binsted and Ritchie (1994):

- 1: Not a joke. Does not make sense.
- 2: Recognizably a joke but a pathetic one.
- 3: OK. A joke you might tell a child.
- 4: Quite good.
- 5: Really good.

Some typical average scores are given below, with examples for each range of scores.

1-2

The butcher commits a carelessness. A gross negligence.
Joan visits a grave in the basement. A bier cellar

2-3

The diver joins a coalition. A choral society.
A store-keeper boards a ship. A sale boat.

3-4

The sailor earns a diploma. A berth certificate.
The juvenile studies a writer. A minor poet.

4-5

The pheasant breathes oxygen. Fowl air.
The sailor bears a stress. Pier pressure.

The average score for all the jokes was 2.81, between pathetic and child-like. This statistic, however, obscures the fact that a significant number of good jokes were generated. Nearly half (22 out of 50) scored between 3-5; and about one-third of the total votes were 4 or 5.

What is more interesting here is the fact that there is no particular correlation between degree of funniness attributed to these examples and the mechanisms used to produce them. More precisely, encyclopedic sources of humour are neither more nor less funny than those based on standard lexical relations such as synonymy, as the following table shows:

<i>Generated pun</i>	<i>Lexical relations</i>	<i>Avg score</i>	<i>Standard deviation</i>
The housewife captures a murderer. A cereal killer.	Encyclopedic/synonymy	1.43	0.5
Joan looks at a musical sign on the station. A bass clef.	Hyponymy/encyclopedic	1.43	0.76
John gives a peck in the home. A buss shelter.	Synonymy/hyponymy	1.71	1.46
The chimney sweep digs a burrow. A grate hole.	Encyclopedic/synonymy	1.86	1
Joan visits a relation on the mound. An aunt hill.	Hyponymy/synonymy	1.88	0.5
Joan visits a grave in the basement. A bier cellar.	Encyclopedic/synonymy	1.88	0.82
The butcher commits a carelessness. A gross negligence.	Encyclopedic/synonymy	2	1.46
The social worker hates a foe. A hostel enemy.	Encyclopedic/synonymy	2	1.07
The cleaner loves an individual. A pail person.	Encyclopedic/synonymy	2	1.34

The vocalist arrives at a settlement. A bass camp.	Hyponymy/synonymy	2.13	0.69
Joan bears stress on the boardwalk. Pier pressure.	Hyponymy/synonymy	4	1.15
The textile worker fulfils a requirement. A dyer need.	Encyclopedic/synonymy	4	0.75
The general performs an operation. A major surgery.	Sequence/synonymy	4	0.5
Joan kisses a hero at the disco. A knight club.	Encyclopedic/synonymy	4.25	0.82
The sailor bears a stress. Pier pressure.	Encyclopedic/synonymy	4.38	0.47

3 DISCUSSION

One of the fundamental challenges for the analysis of computationally generated humour has been to determine how good it is, and to correlate the reactions elicited in humans with characteristics of the system used to produce different sorts of humorous utterances. The examples discussed above suggest that, at least in the case of Tom Swifties and HCPPs, significant use is made of all sorts of relations to generate puns, and there is no particular correlation between the nature of the mechanisms used to produce puns (static lexical relations versus encyclopedic knowledge) and the degree of funniness attributed by humans to the products of these relations.

This suggests that perhaps we should rethink the perspective adopted. Instead of asking whether puns of the sort are **funny**, perhaps we should instead be asking whether they are **clever**. In other words, we should ask whether goodness is correlated with the ability to pleasantly **surprise** the reader or hearer. (One might speak of an **aha** versus a **haha** factor.) Note that this still requires a delicate balancing act. On the one hand, a pun based on a simple and predictable lexical relation is not particularly clever:

It's freezing in here, said Tom coldly

but on the other, a pun which requires a level of knowledge unavailable to the reader or hearer will also fail. For example, the following, taken from the first corpus above:

I'm from a Humberside port, said Tom ghoulishly

relies on the reader or hearer knowing that there is a city called Goole in this region, knowledge that is far from universal. If this knowledge is not available to the hearer or reader of the pun, the pun will fail.

Nevertheless, if we take as a starting point that cleverness is a function of the number and possibly the nature of the links between the elements of the pun, then empirical analysis is possible using computer-generated data and human raters. Variables would include the kinds of links, and also their number. Among other things, we noted above that Tom Swifties at least may include more than one simultaneously operating link (see the *m* tag above). The following examples illustrate the range of values observed to date in corpus data:

I wonder why the hive's still empty, said Tom belatedly.	(bee – late; <i>m</i> = 2)
I've gone back to my wife, was Tom's rejoinder.	(re – joined – her; <i>m</i> = 3)
I had to ask her to leave the yacht because she was too heavy, said Tom excruciatingly.	(ex-crew-she-ate; <i>m</i> = 4)

As an initial working hypothesis, we could suggest that if the type of lexical relation is held constant, then the number of active links in a pun will be correlated with its cleverness. In a second test, if the number of links is held constant, we can

ask whether cleverness is correlated with different types of links. For example, it appears, intuitively, that synonym-based Tom Swifties are less clever than those based on more complex lexical relations.

Another related point has to do with the nature of encyclopedic links. It may be that two distinctions are required: one between simple lexical relations and encyclopedic relations, and another between script oppositions and hermetic games. Consider for example the anagram presented at the beginning of this paper:

Tony Blair MP - I'm Tory Plan B

This creation is funny, in that it summarizes a political judgement with respect to the Labour government in the UK. The funniness may be a function of its value judgement coupled with the 'script opposition' (Raskin, 1985) implicit in the simultaneous occurrence of the name of a Labour PM with the name of the party he opposes. The Tom Swifties and HCPPs found in the corpora discussed above, however, even those based on proper names and those which are based on encyclopedic relations, do not possess this script opposition, but represent rather hermetic games. That is not to say that the addition of a script opposition is impossible in the case of a Tom Swifty. Consider, for example:

I'm sticking to my Labour principles, said Tony Blair rightly.

However, it does mean that instead of answering a single large question (how funny is this element of verbal humour, and why?), we are left with two simpler questions: (a) how clever is this pun, and how does this correlate with its formal characteristics; and (b) does this pun instantiate a script opposition and does this correlate with its funniness? In fact, this also provides some hope for the computational modelling of puns. If we can assume that there is no particular difference in kind between the various sorts of relations which underlie cleverness of the class of Tom Swifties or HCPPs (and probably riddles as well), then we can begin to think about a single formalism designed to capture all such relations.

Of course all of this is very preliminary. Among other things, it is unclear what correlates can be found for these two measures beyond the simple verbal labels. It might be that cleverness is based on a non-involved attitude (we spoke earlier of 'hermetic games') while funniness requires involvement on the part of the hearer. Ideally, there would be two distinct neurological correlates involved. A second question concerns the relation between the two measures of cleverness and funniness. For the moment, it would seem most prudent to assume that the two are orthogonal, in that we could assume examples which instantiate each of the following different combinations, to varying degrees:

Not clever and not funny	(a simple Tom Swifty)
Clever but not funny	(a complex Tom Swifty)
Not clever but funny	(slapstick, or a simple joke)
Clever and funny	(a clever joke, or a pun with funny implications)

Certainly, much work has already been done which would be applicable to this issue, including, for example, the formalism for the representation of world knowledge proposed within the CYC project (Lenat, 1989) and in new initiatives in lexical modelling such as the notion of the generative lexicon (Pustejovsky, 1995). Our own work is based on use of the VINCI natural language generation environment, under development since 1986. Originally conceived to create drill exercises for language learning, the system provides a collection of metalanguages (semantics, syntax, lexical items and lexical relations, inflectional and derivational morphology, phonology) and an interpreter allowing the generation of utterances based on grammars written in the various metalanguages. The two principal tasks of the VINCI system are sentence or text generation and word creation. Word creation involves the systematic application of word-formation rules (lexical transformations) to an existing lexicon to obtain all possible new forms which the rules specify. Sentence generation involves the creation of phrases or sentences in a language specified by the user, either at random or under the control of semantic or formal constraints, including semantic expressions or traits, frequency, orthographic or phonological characteristics, and lexical relations.

In our current research, we are experimenting with complex lexical items which embody not just basic lexical relations (links to synonyms, antonyms, hyperonyms, derived forms, etc.) but also encyclopedic information with respect to individuals. All of this information is captured by means of a common system of attribute classes and values enhanced by a partial ordering mechanism, devices for the construction and deconstruction of complex attributes, and a translation mechanism for movement between a logical formalism and attribute specifications.

REFERENCES

- [1] Binsted, K. (1996). Machine humour: An implemented model of puns. Dissertation, University of Edinburgh, Scotland.

- [2] Binsted, K. and G. Ritchie (1994). A symbolic description of punning riddles and its computer implementation. Research Paper 688, University of Edinburgh, Scotland.

- [3] Cooper, A. (1999) Alan Cooper's Homonyms. Available at <http://www.cooper.com/alan/homonym.html>.

- [4] Cruse, D.A. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.

- [5] Franklyn, J.H. (1966) (Ed.) *Which Witch? Being a grouping of phonetically compatible words*. Hamish Hamilton, London, 1966.

- [6] Lenat, D., R.V. Gupta (1989). *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project*. Reading, MA: Addison-Wesley.

- [7] Lessard, G. and M. Levison (1992). Computational Modelling of Linguistic Humour. *ALLC/ACH92*, Oxford, England.

- [8] Lessard, G. and M. Levison (1993). Computational Modelling of Riddle Strategies. *ACH/ALLC93*, Georgetown, USA.

- [9] Lessard, G. and M. Levison (1995). Linguistic and cognitive underpinnings of verbal humour. *International Cognitive Linguistics Association conference*, Albuquerque, NM.

- [10] Levison, M. and G. Lessard (1992). A System for Natural Language Generation. *Computers and the Humanities*, 26:43-58.

- [11] Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.

- [12] Raskin, V. (1985). *The Semantic Mechanisms of Humour*. Dordrecht, Reidel.

- [13] Venour, C. (1999) The Computational Generation of a Class of Puns. Master's thesis, Queen's University, Kingston, Ontario.