Automatic Interpretation and Generation of Verbal Humor

Proceedings of the Twelfth
Twente Workshop on Language Technology

joint with

International Workshop on Computational Humor

J. Hulstijn & A. Nijholt (eds.)
Automatic Interpretation and Generation of Verbal Humor: Proceedings Twente Workshop on Language Technology 12, joint with International Workshop on Computational Humor / J. Hulstijn, A. Nijholt Enschede, Universiteit Twente, Faculteit Informatica

ISSN: 0929-0672

trefw.: computational humor, non-literal language use, irony modelling, artificial intelligence, cognitive science, language processing, language technology

© Copyright 1996; Universiteit Twente, Enschede
SPONSORS AND SUPPORT

Main Sponsor
NOG Verzekeringen, Amsterdam

Other Sponsors

Board of the University of Twente
Department of Computer Science

KPN Research, Leidschendam

KNAW, Amsterdam

TNO-TPD, Delft

Medialab, Schellinkhout
PREFACE

TWLT is an acronym of Twente Workshop(s) on Language Technology. These workshops on natural language theory and technology are organised by Project Parlevink (sometimes with the help of others), a language theory and technology project of the Department of Computer Science of the University of Twente, Enschede, The Netherlands. Each workshop has proceedings containing the papers that were presented. For the contents of these proceedings consult the last pages of this volume.

Previous workshops.
TWLT4, Pragmatics in Language Technology. 23 September, 1992.
TWLT5, Natural Language Interfaces. 3 and 4 June, 1993.
TWLT6, Natural Language Parsing, 16 and 17 December, 1993.
TWLT7, Computer Assisted Language Learning, 16 and 17 June 1994.
TWLT8, Speech and Language Engineering, 1 and 2 December 1994.
TWLT9, Corpus-based Approaches to Dialogue Modelling, 9 June, 1995
TWLT10, Algebraic Methods in Language Processing, 6-8 December, 1995.

TWLT12/IWCH'96 has been organized by a program committee consisting of Anton Nijholt (chair, Neth.), Salvatore Attardo (USA), Douglas Hofstadter (USA), Franciska de Jong (Neth.), Victor Raskin (USA), Graeme Ritchie (UK), Oliviero Stock (Italy) and Akira Utsumi (Japan).

TWLT12/IWCH'96 was aiming to bring together research results in the area of computational humor, with an emphasis on the computational interpretation and generation of verbal humor. One assumption underlying the organisation of this workshop was that in future human-machine interaction, humans will demand a naturalness and effectiveness that requires also the incorporation of models of probably all human cognitive capabilities, including the handling of humor. Apart from human-machine communication, robotics, automatic text interpretation and machine translation are other areas that will also profit from research on computational humor.

The workshop was held in De Vrijhof and the Collegezaalcomplex at the campus of the University of Twente. Just as with the previous workshop programs, there were presentations by a select group of internationally well known researchers. The general aim was to provide a platform for the presentation of new developments and for the exchange of ideas between people working in the area of computational humor. The workshop was preceded by a symposium on computational humor rounded off with a panel discussion. There was also an essay-contest for students.

A workshop is the concerted action of many people. We are grateful to the authors for their efforts and contributions. In addition we would like to mention here the people whose efforts have been less visible during the workshop proper, but whose contribution was evidently of crucial importance. Charlotte Byron and Alice Hoogvliet-Haverkate took care of the administrative tasks (registration, hotel reservations, etc.). TWLT13, the next workshop in the series, will take place on February 13 and 14, 1997. Its preliminary title is Cheats, Failures and Catastrophes in Language Technology. We hope it will match the success of this and the previous workshops.

Joris Hulstijn and Anton Nijholt

September, 1996
SYMPOSIUM

The workshop was preceded by a symposium on berekenende humor (computational humor) on Wednesday, September 11. A number of well-known researchers from the AI community and the field of linguistics presented their views on the theme. One of the aims of the symposium was to make the topic, computers and humor, accessible to a larger audience. The symposium was organized to celebrate the 15th anniversary of the Department of Computer Science of the University of Twente.

The symposium was introduced by Anton Nijholt, followed by scientist and publicist Hugo Brandt Corstius. He is well known for his witty and often sharp comments on language and politics that appear weekly in various national newspapers and magazines. The results of an essay-contest on Computational Humor were introduced by the chairman of the jury, Dutch author Gerrit Krol. Oliviero Stock presented his views on the usability of humor in the interface, taking inspiration from the Marx Brothers episode 'Password Swordfish'. He was followed by Marvin Minsky, one of the founding fathers of AI and Cognitive Science as we know it, but also at the root of most computational humor implementations. Then we were joined by Douglas Hofstadter by means of an interactive audio-visual link between Bloomington, USA and the symposium site at Enschede. Hofstadter talked on creativity and the 'Eliza effect'.

The afternoon was concluded by a panel discussion, lead by computational linguist Graeme Ritchie. The panel focused on where computational humour might go from here. Computation humor is a new, and very small, field. There is a need for firm results, standard techniques, and even agreement about what constitute the central problems. For instance, what specific questions or subareas should we tackle? Are there some areas of humour that we should avoid for the moment? What methodologies should we adopt? How does computational humor relate to other areas, such as modelling creativity or emotion? Do we have to wait for advances in other fields, such as AI, before computational models of humour will be possible?

The workshop ended on Friday afternoon with a lecture by mathematician John Allen Paulos. In his lecture the connection between humor and cognition in general was explored.

SEMINAR

On Saturday, September 14, an extra seminar was held on the analysis of humorous texts, lead by Salvatore Attardo and Wladyslaw Chlopicki. The seminar revolved around the analysis of three humorous texts with which the participants were expected to have familiarized themselves in advance. Different approaches to the humorous material in the texts were demonstrated and discussed. The texts included examples of irony and complex intertextual references. The emphasis was on the processing of the humorous.

ESSAY CONTEST

Berekenende Humor (computational humor) was also the topic of an essay competition for Dutch computer science students. Students were asked to write an essay on computational modeling of humor and possible applications. The members of the jury were Gerrit Krol (chairman; author), Peter Wesly (philosopher), Michael Stechouder (linguist), Anton Nijholt (computer scientist) and Joris Hulstijn (secretary of the jury; computer scientist). During the symposium the jury presented the results of the contest and rewarded the prizes.
CONTENTS

'Password Swordfish': Verbal Humor in the Interface
   Oliviero Stock 1

Computer Implementation of the General Theory of Verbal Humor
   Victor Raskin 9

Why do People use Irony? - The Pragmatics of Irony Usage
   Akira Ito and Osamu Takizawa 21

Implicit Display Theory of Verbal Irony: Towards a Computational Model of Irony
   Akira Utsumi 29

On Computational Processing of Rhetorical Expressions - Puns, Ironies and Tautologies -
   Osamu Takizawa, Masuzo Yanagida, Akira Ito and Hitoshi Isahara.

Relevance Theory and Humorous Interpretations
   Carmen Curcó 53

From ALIBI to COLOMBUS. The Long March to Self-aware Computational Models of Humor
   Ephraim Nissan 69

Humor Theory beyond Jokes: The Treatment of Humorous Texts at Large
   Salvatore Attardo 87

A Neural Invariant of Humour
   Bruce Katz 103

Why is a Riddle not Like a Metaphor?
   E. Judith Weiner 111

No Laughing Matter: The Cognitive Structure of Humour, Metaphor and Creativity
   Tone Veale 121

Bad Vibes: Catastrophes of Goal Activation in the Appreciation of Disparagement
   Humour and General Poor Taste
   Tony Veale and Mark Keane 133

Speculations on Story Puns
   Kim Binsted and Graeme Ritchie 151

An Integration of a Pun Generator with a Natural Language Robot
   Dan Loehr 161

Humorous Analogy: Modeling 'The Devils Dictionary'
   Cameron Shelley, Toby Donaldson and Kim Parsons 173

More on Humor Act: What Sort of Speech Act is the Joke?
   Michal Ephratt 189
PASSWORD SWORDFISH: VERBAL HUMOUR IN THE INTERFACE

Oliviero Stock
IRST
38050 Povo, Trento, Italy
stock@irst.itc.it

ABSTRACT
Humour will be a necessity in future interfaces, especially in the area at the crossroads of entertainment and education, the so called edutainment. Some considerations on the state of the art in natural language processing and on computational humour prospects are presented, as well as some ideas for the introduction of certain types of computational humour in seductive interfaces. In doing that, reference is made to some of the sparkling exchanges of the Marx Brothers.

1. INTRODUCTION
In a wonderful scene of ‘Horsefeathers’, impossible to reproduce in its force, Baravelli (Chico Marx) is at the door of a clandestine bar. Professor Wagstaff (Groucho Marx) must utter the right password to be let in. Baravelli has told him it is the name of a fish. Groucho, first insists it is 'Mary' with the argument that she drinks like a fish, then tries with 'sturgeon'. Baravelli (who has a strong Italian accent...) says: 'Are you crazy? Sturgeon is a doctor cuts you open when you're sick. I give you one more chance.' Prof. Wagstaff: 'I got it. Haddock.' Baravelli: 'At's funny, I gotta haddock too.' Then this theme goes on until Baravelli says: 'You can't come in here unless you say swordfish. Now I give you one more guess.' Prof. Wagstaff: 'I think I got it. Is it swordfish?' So he can come in and shuts the door. In the passage, Baravelli is left out: 'I wanna come in.' Prof. Wagstaff: 'What's the password?' Baravelli: 'Aw, you no fool me. Heh! Swordfish.' Prof Wagstaff: 'No. I get tired of that. I changed it.' [Anobile 1971].

While I was trying to recover an old file from a remote server some days ago, the system asked me for the password. I was not sure what password it was expecting, tried some I used who knows when. After three attempts I was kicked off. The system manager later told me he changed the password. I did not laugh.

Yet there are many situations with the computer that could be very hilarious, starting from the old famous utterance: 'syntax error' in response to the submission of a complex, inextricable program, fruit of the work of days and nights.

Another funny situation happened to me years ago, when for the first time we were demonstrating the AlFresco System (a dialogue system meant to give information on Fourteenth Century Italian Art) to some important visitors. It was an initial version of the system, with little pragmatic capability. One of the visitors asked 'Who is Giotto?' It took a long while for the system to reply, you could perceive how much effort it was putting into it and how difficult the computation must have been. The answer finally came: 'Giotto'.

But, beside this kind of unintentional humour, I argue that humour in human-computer interfaces will have a part similar to the one it has in human-human communication: its absence will make life unbearable to a large part of us.

Larry Tesler of Apple Computers said the interface is the system. Technology has advanced, new concepts have emerged. And in fact Human-Computer Interaction has developed a lot. More important than anything else is the fact that the role of the computer in society has changed and will change much more. Computers are with us at the workplace, at home, for work, leisure, for being informed, for staying connected to other persons and for learning.

2. THE ART OF BUILDING INTERFACES
Interfaces must be seductive: they must attract and satisfy the user. This is particularly true of situations where the goal is not so much work, but a reality that is a mix of entertainment, information and education: so called edutainment constitutes
an increasing portion of what the computer can offer to our life. We know that a major challenge for our society will be to find ways to improve education, not only institutionally, but at all levels of activity. And we know that learning can go well together with active entertainment. At least in certain situations, this is possible and productive also for work. Interfaces can offer temporary mental disengagement from the real world, by presenting users with alternate engaging activity that brings up challenges at the right level for their skill, culture or capacities. Seductive interfaces are good for entertainment and potentially conducive to learning.

The human mind likes engagement (and emotions). In a conceptually simple extension of noninteractive multimedia - such as animation in films - so far a lot of effort has been put in improving perceptual aspects of output, especially improved rendering (resolution in sound and graphics). Perception is certainly important for instance in interactive games, and also for favouring immersion. But the real challenge goes much beyond that.

New interfaces include characters, agents with a personality [Maes et al. 1995], a mix of real world and virtual world, extending what happens in the realm of predefined animated movies. At Microsoft, for instance, a new interface is being tested that is based on a graphical parrot that entertains a simple spoken dialogue with the user for retrieving music hits.

A good interface requires some of the characteristics we know of movies, or theater [see Laurel 1991], or novels. In the first place a story: for example a person with a problem, a progression as the person meets success and resistance trying to solve the problem, a resolution to the problem. The user must be hooked: the problem, the characters in the story and the exposition style must be established fast.

Then storytelling principles apply:
i) the user must be engaged with plot devices;
ii) the system must set and fulfill expectations;
iii) it must set up payoffs and reversals;
iv) it must build and release tension;
v) it must ensure symmetry, rhythm and order.

In all this a vital technique of animation plays a fundamental role to enhance the clarity of an impending action: anticipation. All significant actions should be anticipated and anticipation should be clearly seen. Just as a consistent action is preceded with preparatory moves that anticipate it, to create reversal the expectation created by an anticipation must be clearly violated.

While these considerations are assumed if we have in mind a multimedia environment, with some adaptation they can be considered valid for any interface design, even for operating systems interfaces.

But the future is simply to forget we have a computer system a component of which is the interface. And the interface itself will cease to have its current physical aspect. We shall interact with objects: we shall communicate and retrieve information through small, personalized flexible devices that will know the owner [see for instance Philips Corporate Design 1996], postcards will become active and speak to us and show videoclips of the sender, we shall interact with the office or home environment, that will know where we are and will be able to react to our words.

After introducing the theme of humorous interaction in the next section, I shall talk of useful technologies and some prospects for advanced interfaces.

3. HUMOROUS INTERACTION

Humour is an essential part of communication. The relaxation of inner censorship and the release of energy that derives from it produce an intense pleasure that tends to repeat itself. The first question is of course when is the situation favourable for giving rise to this phenomenon.

Freud [Freud 1905] states what are the most favourable conditions for the production of comic pleasure:

a) A generally cheerful mood in which one is ‘inclined to laugh’;
b) Expectation of the comic, by being attuned to comic pleasure;
c) Encouragement coming from any other pleasurable accompanying circumstance.

In contrast, major interference factors to humour are:

d) The kind of mental activity with which a person is occupied at the moment;
e) Attention is focused on the comparison from which the comic may emerge;
f) The situation gives rise at the same time to a release of strong affect.

The conditions have been discussed by many authors, and I have not the space to discuss them here from my specific point of view. Similarly I
will not talk about a deep psychoanalysis-inspired treatment of humour, such as the one of Minsky [Minsky 1980], derived from the theory of censors of Freud.

I would like instead to take as a reference the work of Victor Raskin [Raskin 1985], that positions itself at the linguistic level, with some elements that make it a viable tool for designing a limited but concrete system. It relies on the concept of script, a large chunk of information, typically commonsense stereotypical information, evoked by a word or word combination.

His Main Hypothesis states that a text can be recognized as a single joke-carrying text if both of the following conditions are satisfied:

(i) the text is compatible, fully or in part, with two different scripts (i.e. partially overlapping scripts).
(ii) the two scripts which the text is compatible are opposite.

Joke-telling is then seen as subject to maxims of non bona fide communication (established with reference to the well-known Gricean maxims of bona fide communication):

- Raskinian Maxims of non bona fide communication:
  - Maxim of Quantity: Give exactly as much information as is necessary for the joke
  - Maxim of Quality: Say only what is compatible with the world of the joke
  - Maxim of Relation: Say only what is relevant to the joke
  - Maxim of Manner: Tell the joke efficiently

To include joke understanding in a more general communication processing, a general heuristics is defined: if the hearer establishes beyond reasonable doubt that the speaker violates the cooperative principle for bona fide communication, the next hypothesis, in our culture, is that the speaker is engaged in humour.

The semantic mechanisms of humour build on combinatorial rules of script-based semantic theory, according to Raskin. Essential is the relation of script opposition, that holds in a kind of script ontology. To give rise to the humorous effect, a trigger, obvious or implied, is there to release the oppositeness relation among scripts: fundamentally it is based either on ambiguity or on contradiction. Clearly, the punch line of a joke implies or contains the trigger.

Many other aspects can contribute to humour. As Bergson noticed [Bergson 1899], repetition is a typical aspect of humour. We can consider it as a form of metacommunication. Firm script establishment is often performed with serials before the punch line disestablishes it; also in this case the rhetorics of the communication per se causes anticipation of some triggering utterance.

4. STATE OF THE ART AND APPLIED HUMOROUS PROSPECTS

In one extraordinary slip of the pen in Raskin's bibliography, the well known book by J. Weizenbaum, 'Computer Power and Human Reason' [Weizenbaum 1977], is quoted as 'Computer Problem and Human Reason' [Bibliography at pag 266 of Raskin 1985]. This inadvertent, magnificent contradiction is really the heart of what is there to say. The fuzzy and often wrong perception of what is the computer power, in particular of its cognitive and communication capabilities, results in a fundamental computer problem. This is particularly clear when one comes to see that the so called Eliza Effect, emphasized in its danger by Weizenbaum himself, is still going on, if modern events like the Turing Test contest (at the basis of the popular Loebner Prize), are considered. A practical consequence is that whatever achievement is reached on the long way to understanding communication and cognitive processes, its significance is not popularly appreciated, as, at the superficial level, it will compare poorly with the performance of programs based on tricks [see also Hofstadter 1995]. On the other hand, it is a difficult problem to make understandable the limits of the computer one is interacting with, if it has to exhibit some intelligent behaviour, but it is restricted to what the state of the art of the technology supports.

There is some slight progress in that direction with the development of simulation techniques (e.g. so called Wizard of Oz techniques), that can help us see the behaviour of users with systems that do not exist yet.

Coming to the specific case of humour, the attitude of the few people active in computational humour in the intelligent interfaces and natural language processing community is relatively cautious. Bimsted [1995] considers situations for which humour may help in current interfaces, basically to alleviate tension (often NL systems can be seen as patronizing). She proposes the use of humour for self-deprecation, for humorous observation (e.g. about other parts of the system), or for referring expressions. In particular she has
developed a simple pun mechanism that can be well integrated into a natural language-based interface.

In general I claim that natural language processing as a research field [for an introduction see Allen 1995] is in a good position to yield results for humour processing. Let us leave apart questions of representation, such as representations of knowledge sources (grammar, lexicon, concepts etc.) and various representations of utterance interpretations. As far as the NL understanding area is concerned, the field can basically be considered addressing two main objectives (stated separately just for the sake of presentation): be capable to produce efficiently all different interpretations of linguistic expressions, and be capable to choose efficiently the appropriate one in the given context (and take it to its consequences).

This avenue offers exactly the necessary material for a part of verbal humour, if the latter is based substantially on ambiguity and contradiction. There are systems for yielding efficiently all possible interpretations at various levels of analysis (though with a decreasing power as this list is followed): morphological, lexical, syntactic, semantic, pragmatic, discourse, intentional. Disambiguating and taking into account the context is a more difficult task, as are all tasks that involve complex reasoning, but with limitations and restrictions it can be accomplished for specific applications.

Speech recognition is also improving and spoken dialogue opens the way to greater possibilities for the interface in general and for the introduction of humour in particular, beginning with the ambiguities and contradictions that appear at the phonetic level. It is worth noting that often recognizers yield unexpected, wrong interpretations of input that was inadvertently ambiguous or nearly ambiguous.

On the text generation side, the field has developed quite a lot, integrating communication act planning, rhetorical structure planning, lexical choice and sentence generation. Systems have been designed also for multilingual text generation and for including a model of the reader, so as to act differently with different users.

One further aspect worth mentioning here is a tendency in the natural language processing field to integrate various modalities of communication into the same environment. It is not just a question of involving different media, but also of coordinating the overall communication and taking into account the larger potential that a computer can offer for exploration and navigation, combined with 'normal' linguistic communication [Stock 1995, Stock et al. to appear]. Humour can certainly exploit the larger 'communication bandwidth': the minimal way is through the integration of images and deictic references. Yet the whole cognitive potential of the situation has to be explored and understood.

Coming to the semantic-pragmatic level of communication, one should take into account concepts such as presupposition sharing among speaker/hearer, implicature, possible worlds and speech acts. This in general is an area that has fewer results available for applications. As Raskin points out, stereotypical knowledge may be useful as a solution for humour modelling.

On the other side the field of user modelling (with its specialization in student modelling within flexible educational systems) is increasingly recognized as important. A user model can have several levels, from the static aspects, as in a profile, up to dynamic aspects, as in the mental state where the belief's, intentions and goals of the user are represented and updated when some event occurs. This has a lot in common with the flourishing field of 'intelligent agents' and with so called Distributed Artificial Intelligence. Communication is an essential component in intelligent agent behaviour and this brings together this part of artificial intelligence with dialogue systems and the semantic/pragmatic problems mentioned above. For flexible humour it is a fundamental prospect.

Agents in electronic entertainment and interfaces, even in quasi absence of deep language capabilities, are beginning to be experimented [Hayes-Roth 1995], [Maes et al. 1995]. In Maes's work, for instance, one important aspect is that an effort is made to separate the interface from the underlying system so as not to make the user perceive the interface is taking responsibility for the whole system. This is not in contradiction with the view that the interface is the system you are interacting with. Similarly, at the post office your (task-oriented) conversation is with a clerk, not with the postal system.

With a longer term perspective, I believe that this approach will become more and more the way to proceed: one will possibly interact with more agents, each with its social role [see also Conte and Castelfranchi 1995], with its own character and attitude, and each with some degree of knowledge of the user. The user will decide whether to pay attention to what each of them has to say, take into account their suggestions, enjoy the interaction, explore the possibilities.
Again, I believe that while humour without restrictions is certainly ‘AI-complete’ [for the term see Raymond 1991], it will not be possible to leave humour out and guarantee our survival. If we do not propose its introduction, as also Binsted notes, it will be the world of advertisement and commerce that will push for an adoption of flexible humour, just as it does now within fixed commercials.

Interactive and individual-oriented humour, beyond entertainment, per se important, will help all kinds of concept promotion.

I am sure, among application areas, dynamic humour will constitute the key to a number of children activity and games: think of the children ability in finding ambiguities and absurd meanings and the match they could find in a computer; simple humour on the part of the system can be a great resource as it helps memorizing errors and corrections; it will help develop a social behaviour, etc.

Let me conclude this section with an abstract schema for a variety of types of humourous interventions of agent X, based on the latest occurring utterance in an exchange:

• get all interpretations of input from agent Y (normally the user)

• find the intended interpretation in context

• find other (possibly fragmental, and of any level of analysis) interpretations that have interesting properties: semantically very different, based on contrasting, attractive concepts, disruptive of conventions etc.

• of those try to find an interpretation in line with the character of the subject (agent X), or coherent with its previous expressions

• plan an output expression based on this latter interpretation: define a communicative act that makes Y understand, with some necessary abduction, what is the reasoning X followed. In defining the output, some specific reasoning can have a major role

• refine the plan taking into account MX(Y), a model that agent X has of agent Y, and BMBX(A, world): what X thinks the mutual beliefs of X and A about the world (including their relation) are. Here A is the audience, normally coinciding with Y.

• develop the rhetorical structure and generate the output expressions, with particular attention to lexical choice

Heuristics can include: try to stick to the context of previous funny interpretations in the interaction, repeat as much as possible of the previous situation, comment at the metalevel etc. Certainly the use of stereotypical knowledge and behaviour is here very useful, as emphasized by Raskin.

The third point involves many different levels of analysis. In a complex version it can also include problems similar to what is involved in metaphor understanding. While metaphors are at the heart of cognition [Lakoff and Johnson 1980], it is a fact, though, that only limited processing capabilities have been implemented so far.

Creative analogies [see for instance Hofstadter 1985] combined with stereotypical knowledge may be at the basis of the type of reasoning best suited for the initial phase of output planning.

Finally, the character of X, the relation between X and Y and in particular MX(Y) and BMBX(A, world), can help allow producing dynamically humour expressions of different nature, mainly based on puns: irony, sarcasm, satire etc.

A variation of the above schema is for humourous discourse by a single speaker: in this case the first point must be substituted by the following:

• get all interpretations of generated sentences, by means of a parser analyzing X’s own output

while the second point can be removed: the original intended meaning is well known to the system.

Of course not all elements in the planning part can currently be very sophisticated, but in certain cases parts can be based on preorganized or fixed, instead of dynamically constructed, material.

An overall strategy can then put in balance the Gricean part and the ‘Raskinian’ part of the communication. Just think as an example of the great result for instructional text, an important area of application for natural language generation.

In the next section I will limit myself to check if the Marx Brothers can be helpful and if some aspects of their kind of humour is potentially adaptable in interfaces.
5. MARXISM AND INTERACTIVE INTERFACES

The old question is still with us: 'Why a duck?' Why is it that this silly serious expression of Chico's in Cocoyanis, when he goes on and on misunderstanding the word 'viaduct' of Groucho in the context of an urbanistic project, helps us to cross into the future of interfaces? He is Italian and he speak no good English... An ethnical joke, based on repetition of a simple pun and on absolute absurdity. A symbol of pure humour.

As an enjoyable gedanken exercise [see Raymond 1991], I will try to see if some of the mechanisms underlying some kind of Marx Brothers' humour could be automatically producible in the context of the schema indicated earlier, provided that the system has a story, has a character (or more characters) and in general is a good seductive interface. Of course I do not mean that the phenomenal creativity of the Marxes is to be reproduced in all details.

I refer to two sources of transcriptions of Marx Brothers movies: the already quoted [Anobile 1971] and [Marx 1976], while my account owes much to the work of analysis in [Tiersma 1985].

The example at the beginning of this section is centered on the simplest of puns, based on phonetic similarity of word or phrases, (homophones and near-homophones) in the context of ethnic characterization.

In a similar pun, 'multimodal' this time, the great Neapolitan actor Toto explains to an American tourist that the coin he is trying to sell is a real sesterzo (the Roman coin) and to emphasize its authenticity makes the act of steering (the Italian word is sterzo).

Puns of this kind can include context reversal as in the following exchanges in Duck Soup:

Bob (Zeppo): General Smith reports a gas attack
Firefly (Groucho): Tell him to take teaspoon of bicarbonate and a half a glass of water

Mrs Teasdale: This is a gala day for you.
Firefly: Well, a gal a day is enough for me. I don't think I could handle any more

The latter example can be seen also as decomposition, an often adopted resource, based on dividing in parts established units: words, idioms, sayings, names. Here is a clear case of decomposition:

Firefly: I suggest you give him ten years in Leavenworth or eleven years in Twelvelworth (Duck Soup)

While all the previous jokes are based on quasi ambiguity at the phonetic-lexical level, syntactic ambiguity is also a very powerful tool adopted by the Marx brothers. These cases as the ones above fall nicely in the described framework.

Captain Spaulding (Groucho): One morning I shot an elephant in my pajamas. How he got in my pajamas I'll never know (Animal Crackers)

Mrs. Rittenhouse: Captain Spaulding, you stand before me as one of the bravest men of all time.
Captain Spaulding: All right, I'll do that (Animal Crackers)

A similar case is in the ambiguity between idiomatic and literal interpretation:

Prof. Wagstaff: Baravelli, you've got the brain of a four year old boy, and I'll bet he was glad to get rid of it (Horsefeathers)

The following case can also be considered along the same line or just as a lexical slip.
Mrs. Teasdale: Oh your excellency
Firefly: you are not so bad yourself (Duck Soup)

Ambiguity between idiomatic and literal interpretation can give rise to second level ambiguity. The following is a case in which the range of values of the spatial pronoun is very different in the two interpretations. The simple apparition of the other interpretation, where the value of the spatial variable is obvious, has a humorous effect, emphasized by the precise anwer.

Quale (Groucho): Say where 'd I see your face before?
Joe (Chico): Right where it is now. (Go West)

Inversion is similar to ambiguity, but adds some creativity. Instead of exposing a secondary meaning, it creates a new meaning out of the available material. The previous section schema would need some slight alteration: instead of ambiguity recognition, we would have parts permutate in view of maximizing contrast, and a presentation of the new material in output, without any abduction meant on the part of the audience.
Firefly: I could dance with you till the cows come home. On second thoughts, I'd rather dance with the cows till you come home. (Duck Soup)

Groucho (to Lucille): Oh why can’t we break away from all this, just you and I, and lodge with my fleas in the hills. (she looks at him doubtfully)... I mean flee to my lodge in the hills. (Monkey Business)

The following kind of exchange can be inserted in the model only if one introduces a substantial amount of stereotypical representation and reasoning. Yet it can be reconciled with the same basic schema.

The concept is the violation of pragmatic conditions, in particular presuppositions or real world knowledge.

Driftwood (Groucho): Two beers, bartender
Forelo (Chico): I’ll take two beers too. (A Night at the Opera)

Driftwood: Do they allow tipping on this boat?
Steward: Oh yes sir!
Driftwood: Have you got two fives?
Steward: Yes sir!
Driftwood: Well then you won’t need the ten cents I was going to give you. (A Night at the Opera)

Prof. Wagstaff (to bartender): Can you cash a check for 15 dollars and twenty-two cents?

Thanks. As soon as I have a check for fifteen dollars and twenty-two cents I’ll send it to you.
(Horsefeathers)

Finally, in the previous section I mentioned multilinguality (in particular multilingual generation) as an area of active development in NLP. Interlingual jokes can add to the possibilities of making puns, on the basis of quasi homophones. So, while in the ‘why a duck?’-exchanges the false interpretation is made possible by the scarce competence and the phonetic bias of the foreign immigrant Chico, often ‘false friends’ can provide hilarious situations even in monologues. I’ll mention another ‘interlingual’ scene (that includes also multimodality) with Toto’, in the same spirit of the Marx Brothers references, although with a different cultural background. Toto’ is in disguise as an Arab prince, and warned to be tough with his men, while inspecting the guard of honour he hits some black soldiers splitting his sides with laughter. When his aide asks him why is he doing so, he explains ‘Castigo ridendo mores’, a well-known classical expression in Latin (meaning while laughing I chasen the mores), but where Toto’ plays on the assonance between the Italian word mori, i.e. blacks, and mores, i.e. mores.

CONCLUSIONS

Interfaces will have to be seductive and be designed to include a good story and well-developed characters. A restrictive view of interfaces, as seen today, will eventually yield to a more advanced view, in which flexible interaction with the user will be accomplished through an assistant, a critic or through a group of characters that may comment, suggest or react otherwise to what is happening in the communication. The agents will know of the user. Interfaces of this type will accompany us in the various aspects of our life, but to mention one fundamental area, I would like to emphasize edutainment, the field that is at the crossroads of education, entertainment, interactivity and multimedia. Humour has to play an essential role in that. I have tried to indicate some aspects of the art of interface design and hinted at some ideas for the introduction of some type of computational humour. In doing that I have taken various references, including, as examples, sparkling exchanges of the Marx Brothers. I would like to conclude with the consideration that we need to understand more of what an intelligent interface can be, what this medium we are trying to define really is. For certain aspects reminiscent of what we find in nature, for other aspects different and even amplifying our communicative capabilities. The characteristics of this medium will determine its potential on education and on society in general. It is worth reminding what Groucho said of TV: 'I find TV very educational. The minute somebody turns it on, I go to the library and read a good book.'

REFERENCES


Binsted, Kim Using Humour to Make Language Interfaces More Friendly. Proceedings of the AI,
Artificial Life and Entertainment Workshop, IJCAI-95, Montreal, 1995


Freud, Sigmund Der Witz und Seine Beziehung zum Unbewussten. Deutike, Leipzig and Vienna, 1905

Hayes-Roth, Barbara Agents on stage: Advancing the State of the Art of AI. Proceedings of IJCAI-95. Montreal, 1995


Laurel, Brenda Computer as Theater. Addison-Wesley, Reading, Mass. 1991

Maes, Pattie, Blumberg, Bruce, Darrell, Trevor, Pentland, Alex and Wexelblat, Alan Modeling Interactive Agents in ALIVE Proceedings of IJCAI-95, Montreal, 1995.


Tiersma, Peter Language-Based Humour in the Marx Brothers Films. Indiana University Linguistics Club. Bloomington, Indiana, 1985

COMPUTER IMPLEMENTATION OF THE GENERAL THEORY OF VERBAL HUMOR

Victor Raskin
Interdepartmental Program in Linguistics, Purdue University,
and HUMOR: International Journal of Humor Research
West Lafayette, Indiana 47907-1356 U.S.A.
vraskin@purdue.edu

Abstract
This paper belongs in the series of publications exploring the possibilities of humor computation in general and of computer implementations of the general theory of verbal humor (GTVH: the revised and extended script-based semantic theory of humor, or SSTH) in particular. It focuses on three major issues. First, a full-fledged theory of humor is postulated and explored as the necessary foundation for defining and underlying the overall architecture as well as the static and dynamic resources of any computational system for humor analysis and/or generation. Second, the goals of humor computation and its implementation, ranging from the “bag of tricks” approach to full understanding, are discussed. Third, the ontological semantic approach to machine translation is examined from the point of view of the suitability and adaptability of its lexical resources for computing humor. The three issues may appear intimately connected to each other only in the author’s twisted mind.

Introduction
The paper is based on a number of premises, which are all more or less open to debate and which are not going to be defended much here. The first premise is that no computational work on humor should proceed independently of a reasonably comprehensive theory of humor. Such a theory will determine, to the extent it itself is defined, the architecture of the computational system and the composition and contents of each of its static and dynamic resources. The second premise is that any meaningful computation of humor should occur on top of a full-fledged meaning-based NLP system, and because this is not really feasible at this time, any partial system should justify its existence in terms of a hypothetical and/or partially realized non-toy NLP system. This may be a tall order, but anything short of that is based on a bag, often barely filled, of tricks, and having done that himself, the author is no longer much interested in tricks of this nature. The third premise is that a full-fledged meaning-based NLP system falls within the fold of the ontological-semantic approach developed in the MikroKosmos environment, first at the Carnegie Mellon’s Center for Machine Translation and later at the Computing Research Laboratory, New Mexico State University, both under Sergei Nirenburg and with the author’s active participation.

1. From SSTH to GTVH (and Back?)
The early SSTH (Raskin 1979, 1985, 1987) was a combination of two claims about a joke-bearing text, namely that:

- such a text is compatible in part or in full with two different scripts;
- these two scripts are opposite in a specially defined sense

Thus, a primitive Soviet joke of some 10 years ago (1) opposes the apparently intended political-support script with the unintended but not impossible script of providing physical support for a person who has trouble walking on his own:

(1) Who supports Gorbachev?
   Oh, nobody does. He can still walk by himself.

   The concept of script is used here rather loosely. A more careful treatment of the first SSTH claim goes like this, and it is pure linguistic semantics: the meaning of a word is a part of a structured chunk of semantic information, out of which the word picks its allotted part of the chunk. Each sentence obtains its meaning in a largely combinatorial process called, for instance, ‘composition,’ in which the meanings of the words making up the sentence are checked for their compatibility and somehow combined together. The process, outlined first by Katz and Fodor (1963), who called it “amalgamation,” or—more confusingly—“projection,” follows the syntactic structure of the sentence and it weeds out the incompatiable combinations. Thus, in (2), the bird-anatomy sense of bill is ruled out by the occurrence of payable, which strongly and practically exclusively reinforces the invoice sense of bill.

(2) The bill is large but payable in installments.
In a felicitous case, which is (non-)surprisingly common in casual language, only one sense of each word is reinforced by the others, and this disambiguated combination of the compatible scripts adds up to the unambiguous semantic interpretation of the sentence. The first STH claim loosely refers to such an interpretation as the script compatible with the text.

As any linguist will immediately recognize, there are several gross simplifications, overstatements, short cuts, and outright lies in this happy scenario. Structured chunks of information constituting lexical meaning are hard to delimit. Lexical meaning is often fuzzy, slippery, and plastic, i.e., taking a new shape in each environment. Many words have several, often unrelated meanings. Compositionality is not the only factor in semantic interpretation: much is contributed to it from various non-compositional resources. Besides, compositionality is not always straightforward. Compatibility between scripts have never been defined. There are many ifs and buts, but worst of all, there are many gaps and unanswered questions in linguistic semantics in general as well as in sentential, grammatical, and lexical semantics in particular.

STH cheerfully ignored all of these problems, which, the author cynically believes, contributed greatly to its success in the humor-research community and beyond. It conveyed the sense of involving a crucially important process of semantic interpretation, which represents how this is really done in linguistics, and showed how the same process is applicable to humor research. "Forget about jokes for a moment!" the author seems to remember himself preaching. "This is how linguistics analyzes texts semantically. This is what I do on my day job. And, miracle of all miracles, it works in humor research as well!"

A former colleague claimed years ago, not entirely benevolently, that the author's inimitable accent made a native speaker listen to his presentations more carefully than they deserved in order not to lose the thread and thus let that said author command more attention than he was due and sound more persuasive than he was entitled to. That colleague was, of course, completely wrong: the author's accent is perfectly imitable: he does it himself on a daily basis and with admirable ease.

The STH claim actually amounted to this. One needs to interpret a text semantically. One wants to apply the strongest methodology available for this purpose. It happens to be one's own script-based semantic theory. One goes ahead and applies the methodology. The result is that the text is deliberately ambiguous, at least in part, and, as an unintended but welcome side product of regular semantic analysis, one recognizes the text as a potentially joke-bearing text.

This seemed to make STH a rigorous application of linguistics to humor research, and the theoretical and methodological might of linguistics, arguably the most rigorous and well-defined discipline in the humanities, social sciences, or even sciences (nobody knows where linguistics belongs, except that everybody agrees that it is not a part of agricultural economics—at least, not yet, because the mad cows may change all that), is thus lent to the otherwise very "soft" field of humor research.

Well, not so fast, mister! We have left the second STH claim out so far. The two scripts discovered in the process of script analysis to be compatible with the text must be also opposed in a certain predefined way. Thus, both (1) and (3) meet the conditions stipulated by the first STH claim, but only (1) is a joke pretender.

(3) John went to the bank (river bank, money bank).

STH, again somewhat loosely, defined the major types of the specially predefined script oppositions as normal/abnormal, real/unreal, and probable/improbable: one of those in the left parts was what the text was about while the other was deliberately imputed for a calculated humorous effect. When a man sentenced to die on a Monday morning remarks on his way to the gallows that that was not a very good start for his week, the imputed unreal script is that he has a week ahead of him, which could start badly or otherwise, while in reality he does not.

STH specified a few popular manifestations of those types, such as the opposition of sex/no-sex, rich/poor, life/death (which is manifested in this example of (literally) gallows humor—the joke, apparently, contributed the name to this form of humour noir).

But the list of oppositions required by the second STH claim was not part of linguistic analysis. Instead, it was needed specifically for humor research. In fact, therefore, STH was not really a rigorous application of linguistics to humor research. It was instead an extension of linguistics into humor research—this is not necessarily bad, but it is important to discriminate against rigorous and loose applications, of which extensions are the most popular form, because a rigorous application borrows all of its conceptual and intellectual arsenal from the source field (linguistics), leaving nothing for the target field (humor research) to do, except watching that the application pursue the issues of the target field, while an extension leaves some theoretical and methodological work to be done to extrapolate, adapt, indeed extend the conceptual and methodological apparatus of the source field to the target field.

STH also simplified matters. The text (1) would be
recognized by it as a joke-bearing potential, but unless considerably augmented, STH would leave out an important allusion to the fact that a couple of Gorbachev’s predecessors had competed in the speed with which they relinquished the top position in the Soviet government to death and that both they and their long-time predecessor had to appear in public looking very ill and being supported by their bodyguards in the literal, physical sense of support.

An even deeper allusion, also easily and universally accessible to the competent native audience, was the stereotype of a Soviet leader as a strong and energetic leader of men, a head taller, literally and figuratively, than his associates. Thus, Stalin, who was two inches shorter than the author (the latter lying on all of his official papers that he is 5’6”) and in the same height bracket as Lenin, Hitler, and Charles Aznavour, had been always photographed a few feet ahead of his associates, so he could look a lot taller.

This is not to claim that STH is unable to handle these allusions or that it cannot, in principle, clean its act as to the clarity and definability of its concepts. In fact, since the late 1970s, when the underlying script-based semantic theory was conceived, a tremendous amount of progress has been made in semantic theory, and if reformulated today, STH would make much better defined claims. Most progress has been made with regard to the concept of script, and we will address it briefly in Section 3.

The reason for the widespread use of STH in humor research in the last decade is that it has been the best defined and seemingly universal theory of humor. This perception has been wrong, but the theory has played a useful role in making scholars focus on a number of categories that have been conveniently used to compartmentalize the important elements of the joke text. Thus, in (1), both of the two scripts and the switch trigger, the polysemous word support, are reasonably clearly distinguishable within the confines of the theory.

In 1991, however, STH underwent a major revision, subsequently resisted by most of its adherents. At an earlier stage of his own productive work in humor research, Attardo came up with a five-level representation of the joke text. Based on STH, this hierarchical representation did not seek any improvement or clarification of STH. Instead, like many efforts of the late 1980s, the hierarchy assumed the accuracy and efficacy of the STH. What it did instead was add several more components to the convenient dominant theory. Curiously, it was not a broadening revision because the domain of the theory remained the same; nor was it a deepening of the theory because the level of analysis remained largely the same as well or even became coarser-grain in some new components. The revision was mostly along the lines of uncovering something that had purportedly been blocked from view before.

Attardo’s unpublished paper on the hierarchy was incorporated into STH, thus making it into GTVH (Attardo and Raskin 1991; see also Attardo 1994 for further elaborations), in the following way:

- the text of the joke and the script opposition were redefined as two of the six knowledge resources (KR’s) constituting a joke, Language (LA) and Script Opposition (SO), respectively;
- four more knowledge resources were introduced, namely, Narrative Strategy (NS), Target (TA), Situation (SI), and Logical Mechanism (LM);
- a considerable effort was spent on the theoretical justification of the new components and, especially, on their hierarchy.

Each joke was presented, as a result of this revision, as an entity whose LA, the KR which is closest to the directly observable surface of the joke, is manifested as its text. The next KR, underlying LA and second closest to the surface, is NS, which contains information on the expository genre of the text. Thus, NS distinguishes between, say, a question-answer format of the joke text and a simple exposition. Its finer-grain distinctions will include such opening gambits as Have you heard the one about...? and triplet-based structures, popular in (especially American versions of) jokes, such as The first man... the second man... and the third man... or On the first day... on the second day but on the third day...

The next KR down is TA, which identifies the targeted group, individual, institution, etc., who or which is made the butt of the joke. Thus, the Soviet geriatric leadership is the butt of (1) and, therefore, its TA.

TA is followed down by SI, which defines the actual situation(s) presented in the text. Thus, in (1), the situation is ambiguous between helping an invalid move and lending somebody political support. In a typical light-bulb joke, SI is changing a light bulb.

LM has emerged as the most problematic of all KR’s and the most intriguing one. It is supposed to provide the logical structure, as it were, to the joke by using such strategies as figure-ground reversal (as in the original light-bulb joke, where one person holds on to the bulb while four others turn the table he stands on), garden path (as in the joke in which the reader is insistently led to believe that what Gorbachev has that is long and Bush has short and Madonna does not have and the Pope does have but does not use is a penis while, in fact, it is the last name), or numerous restored-missing-link jokes.
SO is, then, the deepest and most abstract level, and the hierarchy has been convincingly confirmed in a psychological experiment, brilliantly designed by Willibald Ruch (Ruch et al. 1993) and based on the GTVH prediction that similar jokes will be perceived the more similar the more superficial their differences are. In other words, if two jokes share all of their KR values, except for some differences in LA, they are perceived as mere paraphrases of each other, while SO differences make the affected jokes much less similar. The hierarchy was confirmed in its entirety except for the elusive LM, which has led to such doubts about its very existence as Raskin's heretical (1995).

With all the expressed and unexpressed reservations about SSTH and GTVH, they are reasonably full-fledged theories of humor, and they define what a computational system of humor needs to account for.

In the simpler case of SSTH, a computational semantic analysis of a text must identify the two scripts which are compatible, in part or in full, with the text and then check if these two scripts fit one of the 20 or so script oppositions from a predefined list. GTVH adds to this task by requiring that the system:

- recognize the narrative strategy utilized in the text;
- determine the butt of the joke;
- capture the situation of the joke;
- and, worst of all, identify the LM strategy.

The computational task of generation is even harder: the system must:

- start with a permittec SO;
- match it with an appropriate LM;
- match this combination with a suitable SI;
- appoint the TA;
- select an effective NS; and, finally
- generate the text (LA) of the joke.

Presented like this, both tasks look scary enough but the latter, that of generation, is especially prohibitive. So what are we doing here, discussing computational humor?!

2. Quo Vadis in Computational Humor?

SSTH/GTVH has always insisted on its analysis/generation neutrality. At the same time, its claim that humor is a universal human faculty and that humor competence is part of human intelligence (augmented by some scholars to mean that humor competence is part of communicative competence, next to and along with language competence) is not really neutral. Computational humor, whatever else it might be, is part and parcel of the artificial-intelligence enterprise (or ‘racket,’ in former U.S. Senator Proxmire’s terms, he of the Golden Fleece Award fame). AI models human intelligence. Now, the ability to generate jokes is not part of human intelligence because most humans do not have this ability and because any attempt to establish a correlation between the sense of humor (defined loosely and measured in most suspect terms) and intelligence (undefined and measured in heavily compromised terms) (see, for instance, Mindess et al. 1985) has been unsuccessful. (The fact that the author is both extremely witty and highly intelligent—handsome too!—should not be really seen as such a proof—but thanks anyway!)

Understanding jokes is, however, accurately viewed in humor research as a universal human faculty, so humor analysis is part of human intelligence and, therefore, a legitimate target for AI and NLP. What can a computational system of humor analysis aspire to do then?

First, it can indeed analyze a text and determine whether it is a joke or not. Since, premised on SSTH/GTVH, it will seek the two opposing scripts and the opposition they satisfy in order to determine that, why not ask the system to identify the scripts and their opposition? This, actually, does not sound too scary. In fact, in the extent that semantic script analysis works, it seems quite doable. What we need for the implementation is a lexicon, in which the entries contain script-related information, and a compositional mechanism which negotiates the script compatibilities and incompatibilities and comes up with a script-like label for the whole sentence or a considerable part of it. The rest seems easy—just checking the pair of scripts against the opposition list.

There are, of course, difficulties every step of the way. Neither the lexicon nor the compositional mechanism is available at this stage of NLP research, but the field is getting there. The successfully identified scripts may be labelled differently from the SO list, and establishing the necessary equivalencies is likely to take more than a good dictionary of synonymy (please e-mail the author if you find one!).

These difficulties pale away, however, when compared to the computation of the additional KR’s of GTVH. No established inventory of narrative strategies, modes of exposition, or even genres is available, and, in fact, a theory for each of these phenomena and for the entire phenomenon of narration is both controversial and extremely conceptually soft. When an individual life is treated as a narrative as in I am leaving you because you disrupt the narrative of my life, the question becomes what is not a narrative, and one certainly does not want to develop a theory of life to account for this KR. A practical solution will, of course, distinguish a
few clearly marked strategies, and continue to add to this list as our knowledge increases, and to mark every failure to distinguish such a strategy as a successful identification of the default, simply expository strategy. It should be noted, in this connection, that a text involving a sequential temporal arrangement of events will establish the chronological expository strategy positively and not as a default, but an unrecognized strategy may also turn out to be a slightly disturbed chronological exposition (for instance, if a chronological sequence is interrupted by the statement of a panchronic truth).

TA can be just about any object and many events, even though it can be rather convincingly argued that most joke buts are human in various combinations and, therefore, it is the agents of the events which constitute TA's. If true, the claim may simplify matters somewhat. Simple techniques such as hypothesizing that the first instantiated individual or group is the TA of the joke is very risky, even though it may work out for a statistically impressive subset of jokes. But if two people meet on a street, frequently only one of them will be the TA, so how do we determine which one? Practically again, one will probably set up a list of easily recognizable TA's and set up a default for all the unrecognized ones. The degree of success here strikes one as even lower than with NS. Is it possible, then, that the computational task of accounting for each KR becomes more complicated as we go down from the surface of the joke.

At first glance, this does not seem to be the case with SI. Identifying the main event(s) in a (short) text does not strike one as a particularly difficult task. One does, however, need the concept of a situation, perhaps no less vague than, for instance, in Barwise and Perry (1983), and a way to match individual events or brief series of them with a situation, which will thus gain the status of a "superevent."

The main problem withLM's is, of course, that we do not know what they are. In fact, we are not entirely convinced that they exist. That, by itself, has never been much of an obstacle for NLP or AI, but it does render a practical solution outlined for the earlier KR's still more problematic for this one because the search even for a well defined LM is quite complicated. How, for instance, do we establish that the "Polish" way of changing the light bulb is an exact reversal of the common way to accomplish the task?

One additional scary thought is that, by attempting to account for and to compute those additional KR's, we abandon casual language and cross into the impossible waters of non-casual language, language for special purposes, such as to convince, to impress esthetically, to shock, to attract, and so on. A joke becomes then a work of art, which it actually is, both from the point of view of language craftsmanship and of copyright law. And NLP cannot handle fiction or poetry--not, at least, in machine translation--and probably should not.

All of this sounds pretty defestist, but this is not the end of the story. Analysis sounds like rather an unexciting task to aspire to through considerable difficulties--not even if we attempt to jazz it up by attempting to translate jokes into non-joke, bona-fide texts, explaining why the joke works. This would not add any labor to the identification subtasks of computational analysis of humor identified above.

In fact, this translation task would introduce a medicum of empirical verification to the computational analysis of humor that the component-identification tasks clearly lack. What if the system succeeds in identifying correctly each and every KR of GTVH? Does it confirm that a view of humor based on the theory is correct? No, it will not. It will confirm only that we can build a system which returns to us what we want it to return, but we cannot validate these returns in terms of reality. No matter how little respect, let alone admiration, one has for reality (unless it is virtual), this would be bothersome.

Now, computational generation of humor is a totally different beast. If a system, built on certain theoretical premises, can produce texts which are perceived by human native speakers of the language to be funny, this would provide a strong empirical verification of the underlying theory. But joke generation is indeed hoary. So, what to do? Quo, indeed, vadis?

Why do we compute texts, in general, though? There are two major games going on in AI in general and NLP in particular. On the one hand, there are serious needs in NLP systems emulating human intelligence to an ever growing degree. These tasks include machine translation, information retrieval, automatic abstracting, and perhaps a few others. The other challenge is to build a system on the concepts, rules, and principles of a linguistic theory and have it validated if such a system accomplishes the prescribed task at the level of accuracy typically achieved by a successful human operator.

The first goal seems to be absent from computational humor. People diagnosed with compulsive humorism (as the author officially was, at the 1995 Birmingham, England, Annual Meeting of the International Society of Humor Studies) produce enough jokes, good and bad, and they do not really want an artificial humor generator for a competitor. Add to this that a successful machine translation system will save the labors of thousands, if not millions of people for any pair of languages, but jokes are sold by joke writers to performers at $50 a
piece, a price which does not seem to have risen with inflation for several decades.

This leaves computational humor with one goal, that of theory verification. And it looks like computational humor should focus on joke generation in order to capture that ultimate moment of empirical verification. This, in turn, signifies that computational humor aims to model the human intelligence of joke writers, comedians, humorists, and those few members of the public who generate jokes for free (for deeply personal reasons, the author will not use the term 'freaks of nature' for this last category).

Raskin and Attardo (1994) reported efforts in this direction, and that was a stark illustration of the bag-of-tricks approach, coming—somewhat surprisingly—from the proponents of a seemingly powerful theory of humor. In fact, it is amazing to what extent the actual computation in LIBJOG is theory-free. The authors applied the theory in their migration from the early releases of the system to the later ones, when they expanded the range of each release compared to a previous one by manipulating the KR's, but the computational principle was essentially that of hardwiring, which requires and produces a near-zero level of intelligence. In this respect, LIBJOG was typical of other current approaches to computational generation of humor.

The main trick in the bag was the compilation of a static, lexicon-type resource, in which each lexical entry, always an English word indicating a group of humans sharing a geographical, professional, ethnic, sexual, or some other trait, was characterized by a character trait and a structured event, as shown in (4)

(4) Polish Americans DUMB (activity_1 hold light bulb) (number_1 5) (activity_2 turn table) (number_2 4)

The contents of the entry were used to fill the slots in the single template for a 1 releases (3):

(5) How many [lexical entry head] does it take to change a light bulb? [number_1], [number_1-number2] to [activity_1] and [number_2] to [activity_2].

The point here is that each such lexicon entry is already a ready-made joke, and while, in a degenerate way, the system does generate the text of the joke, it does not assemble and manifest any features of the joke nor does it do any other significant linguistic work either. Intelligence begins at least with some choices to make. The moment the system is allowed, for instance, to choose or—better still—to propose and assemble a structured event which appropriately illustrates the necessary character trait, the intelligence is in and the task becomes difficult—unless events are marked for this trait in the lexicon (see Section 3). It is important to note, however, that this intelligence bit, an ability to match a character trait with an illustrating structured event, is not really a specifically or exclusively humor intelligence.

In general, the moment any KR is opened up for the system to express, the system is intelligent and also incredibly complex. Each KR-related component of such a system also deals with non-specifically-humor intelligence—it is only all of them together which are humor-related, and successful generation would indeed confirm that these KR's constitute the set of sufficient conditions for a text to be a joke.

None of that was allowed to happen in LIBJOG, and every humor generation system should be called upon to examine itself critically as to what intelligence is modeled in it. Apparently, what is implementable in computational humor at this time is relatively small and not exclusively humor-related bits of intelligence that an independent theory postulates as necessary components of a joke. A typical NLP system would then fix or hardwire all the other components, with the help, for instance, of appropriate partially filled templates, and test each such bit of intelligence separately.

This current level of feasibility renders computational humor a somewhat more hazardous enterprise than even NLP. The native speaker analyzes humor on top of the full-fledged native language competence. To model this type of intelligence, as Raskin and Attardo (1994) argue, computational humor would have to implement a full-fledged meaning-based NLP system and add a humor module to it. Now, the existing NLP systems still fall short of complete understanding of the meaning of the processed text and have to take various shortcuts around this problem— even though this situation is rapidly improving. Computational humor "inherits" this feasibility level from NLP and adds a similar situation with regard to humor theory: because we have only a partial, even if ever growing understanding of humor we have to postulate some parts of it as given and test some selected patched of it. In both cases, one is hoping for an incremental success, but in the meantime, computational humor suffers from double jeopardy to NLP's single one.

In this precarious situation, the importance of the underlying full-fledged theory of humor for identifying each component and subcomponent of the joke, both individually and in terms of its relations to every other component, cannot be exaggerated. But the theory will not be validated or falsified until all the components are implemented. Those attempts at computational humor which do not base themselves on an explicit theory sneak in an implicit theory and thus subject themselves to a much larger degree of uncertainty as well as consid-
erably complicating the task of the verification and evaluation of their results.

The next and, mercifully, last section sketches out what seems to be a promising environment for a theory-based modular implementation of computational humor.

3. Ontological Semantic Approach to Computation of Humor

It appears that the greatest depth and richness of meaning analysis has been achieved and implemented in the ontological-semantic environment of the MikroKosmos multilingual knowledge-based machine translation system (Nirenburg et al. 1992; Onyshkevych and Nirenburg 1994; Beale et al. 1995; Raskin and Nirenburg 1995; Nirenburg and Raskin 1996; Nirenburg et al. 1997). It is the vast resources, both static and dynamic, developed in MikroKosmos for industrial strength, multi-domain, spectacularly non-toy NLP that make MikroKosmos the best candidate for a computational humor "extension."

MikroKosmos combines findings from a variety of quasi-autonomous microtheories of language phenomena, world knowledge organization, and procedural knowledge at the level of computer system architecture. The basic motivation for this organization is the continued inability of the fields of linguistics and NLP to produce a general-coverage, unified theory of treatment of language phenomena, a failure especially pronounced in areas beyond computational syntax.

The purpose and result of the MikroKosmos analysis process is the derivation of an interlingual representation for natural language inputs. The language in which these representations are expressed is called the "text meaning representation" (TMR) language, and "texts" in this language are called, simply, TMRs. TMR is a frame-based language, where frame names typically refer to instances of ontological concepts, slot names are derived from a set of ontological properties and slot fillers are either elements of property value sets or pointers to concept instances.

An ontology, a world model containing information about types of things, events and properties in the world, is a necessary prerequisite for a TMR language. "An ontology for NLP purposes is a body of knowledge about the world (or a domain) that a) is a repository of primitive symbols used in meaning representation; b) organizes these symbols in a tangled subsumption hierarchy; and c) further interconnects these symbols using a rich system of semantic and discourse-pragmatic relations defined among the concepts" (Mahesh and Nirenburg 1995: 1). The function of the ontology is to supply "world knowledge to lexical, syntactic, and semantic processes" (ibid; see also Nirenburg et al. 1995).

The lexicon in MikroKosmos "mediates between the TMR and ontology" (Onyshkevych and Nirenburg 1994: 2). Lexicon entries for most open-class lexical items represent word and phrase senses, which can be either directly mapped into ontological concepts or derived by locally (that is, in the lexicon entry itself) modifying constraints on property values of concepts used to specify the meaning of the given lexical item. Lexical-semantic information as well as clues for contextual semantic and pragmatic processing are typically located in the lexicon.

It may be useful to illustrate the MikroKosmos treatment of lexical meaning, the most important static resource, on the example of adjectives. For a variety of reasons, the microtheory of adjectival meaning became the first relatively complete microtheory in MikroKosmos. One of these reasons was that adjectives are anchored in the ontology in a more complicated way than typical nouns or verbs.

A simple, prototypical case of adjectival modification is a scalar adjective, which modifies a noun both syntactically and semantically. The microtheory associates its meaning with a region on a scale which is defined as the range of an ontological property (cf. Carlson and Nirenburg 1990). The contribution that the adjective makes to the construction of a semantic dependency structure (TMR) typically consists of inserting its meaning (a property-value pair) as a slot filler in a frame representing the meaning of the noun, which this adjective syntactically modifies.

Thus, in big house, big will assign a high value as the filler of the property slot SIZE of the frame for the meaning of house. The range of the ontological property concept SIZE is a numerical and continuous scale. Each numerical scale can be measured in an absolute manner (e.g., linear-size in feet, yards, or millimeters, or time in seconds). But often natural language expressions do not refer to absolute magnitudes but rather to abstract relative ones, as in the case of big. We assume a 0 to 1 numerical range for such abstract scales. For abstract references to SIZE, the fillers in English can be:

```
| 0 | minuscale | small | medium-size | big | enormous | gigantic | 1 |
```

*Big* will, then, be assigned a value of ' > 0.75' value on the SIZE scale. These values are a crucial part of the lexical mapping (LEX-MAP) from language units to TMR units included in the semantics (SEM-STRUC) "zone" of
their lexical entries. Equally crucial is the syntactic-semantic dependency mapping (linking) between the syntactic-structure (SYN-STRUC) and SEM-STRUC zones, which in MikroKosmos is carried out with the help of special variables. The syntactico-semantic information in the lexicon entry for big is as follows (6):

(6) (big
  (big-Adj1 ;the first adjectival sense of BIG
    (CAT adj)
    (SYN-STRUC
      (1 ((root $var1)
        (cat n)
        (mods ((root $var0)))) ;bound to the noun
        ;the adjective modifies;
        $var0$ is bound to the adjective itself
      (2 ((root $var0)
        (cat adj)
        ;2 (predicative
        (subj ((root $var1)
          ;this standard Adj
          ;SYN-STRUC is
          (cat n))))) ;omitted from the other
        ;examples
      (SEM-STRUC
        (LEX-MAP ;the syntax-semantics mapping
          ((1 2) (size-attribute ;ping valid for both
            (domain (value "$var1") ;patterns: "$"
              ;means "the meaning of"
            (sem physical-object)) ;selectional
            ;restriction
            (range (value (0.75)) ;the value is
              ;in the top 25 percentile
            ;of the scale
            (relaxable-to (value (0.6)))))));
            ;laxed values are for pro-
            ;cessing non-literal usages
        (SEM-STRUC
          (LEX-MAP
            ;communicative-event
            (agent (value "$var1") ;sem human)
            (benef (value "$var2") ;sem human)
            (theme (value refsem1))
            (attitude1
              (type evaluative)
              (attitude-value (value (< 0.25)))
              (scope refsem1)
              (attributed-to (OR ("$var2" speaker))))
            (attitude2
              (type evaluative)
              (attitude-value (value (< 0.25)))
              (scope "$var2")
              (attributed-to "$var1")))
    (SEM-STRUC
      (LEX-MAP
        ;private-home
        (size-attribute (value > 0.75))
    (8) (private-home
      (size-attribute (value > 0.75))

Many lexical entries for verbs are also related directly to the corresponding event concepts in the ontology, but most are anchored in larger concepts, and the specific meaning is arrived at through appropriate constraints on various slots of the concept as well as with the help of other devices, such as attitudes. Thus, the abuse-verbally sense of the verb is defined in the lexical entry (9) as the communicative-event, in which the message, i.e., what is said by the agent, is considered negative and the beneficiary (i.e., patient or recipient) is viewed negatively by the agent.

(9) (abuse
  (abuse-V1
    (CAT v)
    (SYN-STRUC
      ((root $var0)
        (cat v)
        (subj ((root $var1)
          (cat n))
        (obj ((root $var2)
          (cat n)))))
    (SEM-STRUC
      (LEX-MAP
        ;communicative-event
        (agent (value "$var1") ;sem human)
        (benef (value "$var2") ;sem human)
        (theme (value refsem1))
        (attitude1
          (type evaluative)
          (attitude-value (value (< 0.25)))
          (scope refsem1)
          (attributed-to (OR ("$var2" speaker))))
        (attitude2
          (type evaluative)
          (attitude-value (value (< 0.25)))
          (scope "$var2")
          (attributed-to "$var1")))

The lexical meaning of the adjective is grounded in the ontology through the property concept SIZE, whose frame contains slots for DOMAIN and RANGE as well as for selectional restriction information. The lexical entry in the lexicon fills the ontological slots and adds more constraints, if necessary.

The standard procedure for representing adjectival modification in TMRs is to insert the scale name and scale value for an adjective as a property-value pair in the frame describing the meaning of the noun the adjective modifies. For a noun like house, whose appropriate sense (7) is directly mapped into an ontological concept, the meaning of big house will be represented as a TMR fragment shown in (8):

(7) (house
  (house-N1
    (CAT n)
    (SYN-STRUC
      (1 ((root $var0)
        (cat n)))))
It also seems clear that a reasonable approach to humor would have to start with the lexicon, in which the Mikrokosmos lexical entries will have to be extended in the direction of humor. A few steps in this direction were outlined in Raskin and Attardo (1994), which suggested that:

- certain groups, e.g., Polish Americans, be marked for their humorous association with a certain trait and a certain (ridiculous) way of doing things;
- a homonym slot be added to the lexical entries to mark the potential puns as the humorous switch triggers in jokes;
- humorous stereotypes be added to entries, wherever possible.

This paper has no quarrel with these proposals. They were associated, however, with the SMEARR lexicons (see Raskin 1990, Raskin et al. 1994a,b) developed in 1988-91. The Mikrokosmos lexicons incorporated the SMEARR philosophy and results and improved on them considerably. Even more significantly, the ontological basis for the lexicon was assumed but not really implemented in SMEARR, and it has been in Mikrokosmos. One can now see even more possibilities for a prettier seamless humorous extension of the Mikrokosmos lexical entries—and the ontological concepts—underlying them to provide humor material.

One can see, for instance, how one can provide both humorous associations and homonymy information to generate such serial jokes as the American “do it” bumper stickers of the 1980s, almost all of them defined by the template (10):

(10) [group name] do it Adv-Mod.

Typical examples are (11) and (12):

(11) Linguists do it with the tongue.

(12) Hackers do it all night.

The generation of (11) is conditional on relating linguists with tongue, which is already implementable in Mikrokosmos as well as marking tongue, already treated as homonymous, for its sexual use, which is not yet reflected in the lexicon. A permanent imperative for the search of possible sexual connotations, a force motivating many human humorists, is rather easy to build into the process. It is somewhat more troublesome to relate hackers with all night. The entry for the former does include the idea of spending a great deal of time at the computer, but the idea of all night, with night independently marked for its sexual connotation, has to come from something like a script, and this brings up the last point of the paper.

The most challenging and promising of the extensions mentioned in Raskin and Attardo (1994) seems to be related to humorous stereotypes. And the way to incorporate them in the Mikrokosmos environment appears to reside in the acquisition of ontological scripts, a process currently underway in the system for humor-unrelated purposes.

The Mikrokosmos ontology already includes concepts for complex events. Thus, Carlson and Nirenburg (1990: 15-18) propose a decomposition of the ontological event concept TEACH into three subevents, roughly, LECTURE, ASK-QUESTIONS, and ANSWER-QUESTIONS. Work on the bilingual English and Spanish merger and acquisition corpora including The Wall Street Journal corpus later on has led the Mikrokosmos team to the realization that the concept of MERGER, for instance, is even more complex than a complex event, and the solution was the development of an ontological script for MERGER, which consists of a chronological and logical sequence of events, some of them complex events. These include, of course, bids, negotiations, stockholders' votes, government permissions, etc.

The introduction of ontological script concepts makes it possible to implement all the sexual, ethnic, political, and other humorous scripts mentioned, for instance, in Raskin (1985). It also opens the much welcomed possibility of incorporating the elusive LMMR in such a script—if, for instance, we learn how to reverse the figure and ground of any applicable process automatically.

The participation of an individual concept in an ontological script will also automatically mark for humorous use all the lexical entries grounded in the concept, thus achieving the other lexicon extensions for humor proposed earlier. It is, of course, trivially true that the practical aspects of the ontological scripts, such as their identification, delimitation, and implementation, are complex tasks. What is more significant, however, is that these concepts fit the task of humor computation exactly right.

Conclusion

It is clear, as it has been all along, that only the most complex linguistic structures can serve any formal and/or computational treatment of humor well, even if one has to make do with the simpler one before the more complex ones are developed. Ontological script concepts in Mikrokosmos forcefully bring back the notion of script from the original STH. In general, scripts as used lexically in the script-based theory of semantics (Raskin 1986) have been implemented as a form of the NLP's semantic-network approach (see, for instance, Nirenburg and Raskin 1987 and references there). A
logical continuation of this approach has now led to the independent introduction and justification of the notion of the sentential script as an "amalgamation" of lexical scripts, on which the STH Main Hypothesis rests. It is gratifying that this development was not motivated by computational humor, but it is the existence of a humor theory, no matter how imperfect, as the basis of computational humor, that has always sought the implementation of such a concept.

Viva STH and GTVH! Down with whatever!

Acknowledgments

The research reported in this paper was supported in part by the NSF Grant #IRI-8803733 to the Purdue Natural Language Processing Laboratory in 1988-91 and by the CRL/NMSU Contract #MDA904-92-C-3189 with the U.S. Department of Defense. The author is grateful to Purdue University for permitting him to consult CRL/NMSU on the MikroKosmos project. He is greatly indebted to the other members of the MikroKosmos team, especially to its leader, Sergei Nirenburg. The author's fruitful collaboration with Salvatore Attardo in the early 1990s is obvious from the references.

References


Why do people use irony?
– The pragmatics of irony usage –

Akira Ito and Osamu Takizawa
Kansai Advanced Research Center
Communications Research Laboratory
Iwaoka, Nishiku, Kobe 651-24, Japan
{ai, taki}@crl.go.jp

ABSTRACT

The mechanism of rhetoric irony is explained with the help of a dialogue model. According to this model, the essence of irony is that it allows the transmission of the speaker's intention without having to register it in the shared beliefs, consequently inhibiting orexcusing the hearer from counterarguing. How irony is used for achieving various purposes is investigated. Several usage types of irony are introduced, and an algorithm for determining the usage type is proposed.

1 INTRODUCTION

Irony has attracted the interest of those researching linguistics and natural language processing for its peculiar feature – it allows a speaker to convey his true intention by using an expression seemingly opposite to his intended meaning. To fully understand rhetoric irony, the following questions must be answered.

Q1 What are the necessary and sufficient conditions for irony? (Definition of irony)
Q2 How can a hearer determine whether an utterance is ironic? (Detection of irony)
Q3 What is the speaker's intention in using an ironic expression rather than a straightforward expression? (Intention of irony usage)
Q4 How can a hearer determine the true intention of the speaker? (Inference of speaker's intention)

For Q1 and Q2, the mention theory proposed by Sperber and Wilson (Sperber & Wilson 1981) and extended by various authors (Littman & Mey 1991; Wilson & Sperber 1992) may give a plausible answer. According to this theory, irony is an utterance that mentions or alludes to an “ironic situation” which existed prior to the time of the ironic utterance. In spite of the differences in opinion as to what is really an “ironic situation,” these authors share the common view that an ironic utterance is a mention or allusion to the ironic situation. We also agree that an ironic situation must precede an ironic utterance. The weak point of this theory is, however, that it cannot explain why NOT every utterance mentioning or alluding to an ironic situation is ironical.

To overcome this problem, Utsumi (Utsumi 1995) proposed supplementary conditions for an utterance to be ironical:
- The utterance violates the principle of pragmatics in some way.
- The utterance suggests the speaker's mental attitude and/or feeling.

However, every utterance suggests in some way the speaker's mental attitude and/or feeling. The important point is what kind of mental attitude and/or feeling does the usage of irony suggest. Moreover, another question remains unanswered. If the function of irony
is simply to mention a prior ironic situation, why must irony take a complex form, such as the “violation of the principle of pragmatics?” Can the mention theory explain all the facets of irony?

Take the following dialogue. What does a mother intend to convey with the use of irony?

**D1:** To a child coming home with a dirty face,
Mother: “What a beautiful face you have!”

Of course she wants to inform her child that he has a dirty face, but if that is all she intends to convey, why not simply say “What a dirty face you have!”? There is no reason why she has to use the complexity of irony. Similarly, how can we judge the boss’s intention in the following remark?

**D2:** To the new employee who submitted a poor hand-written report,
Boss: “What a beautiful report you have prepared for me!”

If the intention of irony is simply to mention or allude to an ironic situation, use of such sophisticated rhetoric goes against Grice’s Principle of Manner: “Be brief”.

Thus, the mention theory cannot explain why people use ironic expressions instead of saying things straight. Our answer is as follows. Irony is used to achieve a goal unattainable without the use of rhetoric irony. The speaker’s intention is defined as the difference between the information conveyed by the straight utterance “What a dirty face you have!” and that conveyed by the ironic utterance “What a beautiful face you have!” Actually this difference is what the mother wants to convey with the use of an ironic expression.

Considering why people use irony in various situations, questions Q3 and Q4 require a more detailed analysis than has been done using the mention theory. As it is difficult to construct a theory covering all aspects of irony, in this paper we focus on questions Q3 and Q4. We take a bottom-up or case-based approach, i.e., by looking at “typical” usages of irony, we investigate how irony is used to convey a particular attitude or feeling of the speaker, and why it can be used for that purpose.

2 MECHANISM OF IRONY INTERPRETATION

We introduce a “typical irony” and then go on to explain the mechanism of irony interpretation with the use of a dialogue model (Ito & Takizawa 1995).

**Dialogue model**

In a dialogue, speaker A utters $U_A(X)$ to inform hearer B, his intended meaning (proposition on X) I. Utterance $U_A$ may contain ambiguities, anaphora, or ellipses and does not correspond uniquely to the proposition I describing the actual situation. Hearer B must therefore infer A’s intended meaning by using interpretative function $E$: $I = E(U_A, Bel_B(X))$. We assume that E, which represents the linguistic knowledge, is shared by the dialogue participants. Note that B’s belief, $Bel_B(X)$, is used to interpret the utterance.

If A wants to ensure that B interprets A’s utterance correctly, A must first guess B’s belief $Bel_B(X)$, then choose $U_A$ appropriately so that $E(U_A, Bel_A(Bel_B(X)))$ equals A’s intended meaning. From B’s standpoint, B might think that A’s intention can be more precisely inferred if $Bel_B(Bel_A(Bel_B(X)))$ is used in place of $Bel_B(X)$. To avoid these endless reflections of beliefs, the dialogue participants maintain a set of shared beliefs among themselves. That is, each participant maintains shared beliefs $Bel_S(X)$, and promise that his beliefs $Bel_{self}$ consistently include the shared beliefs $(Bel_S(X) \subseteq Bel_{self}(X))$.

Accordingly, if the shared beliefs contra-
dict his own, the hearer must either change his belief or negotiate a change in the shared beliefs. So long as this rule is observed, the speaker can safely generate $U_A$ using $Bel_S(X)$ in place of $Bel_A(Bel_B(X))$. An utterance of $U_A$ ($I = E(U_A, Bel_B(X))$) is transmitted to $B$, and is interpreted to obtain $A$'s intended meaning $I$. At the same time, $I$ is registered in shared beliefs $Bel_S(X)$. If $B$ does not want $I$ to be registered in $Bel_S(X)$, he has to inform $A$ by objecting to $A$'s utterance.

**Mechanism for irony interpretation**

Irony is an utterance whose propositional content $I$ is the speaker's evaluation of $X$. Suppose that the speaker's intention is to inform the hearer that his evaluation of $X$ is unfavorable ($X$ is bad). The simplest way to do so is to explicitly say "$X$ is bad." However, another alternative exists if there is a shared belief that the evaluation of $X$ is bad, or at least not good.

We assume that every utterance expressing an evaluation has, in itself, with probability $p_a$, the possibility that it really means the opposite of the surface meaning. Of course, in ordinary situations, probability $p_a$ is small compared to the probability of the straightforward meaning ($p_a = 1 - p_{-a}$). The probability distribution function for the evaluation of $X$ can be approximated using the sum of the delta functions, as in Fig. 1a.

An utterance of $A$ is interpreted by $B$ with the use of $Bel_S$. Note that we assume, as a precondition for using irony, that the shared belief about the evaluation of $X$ is not good. The probability distribution for the shared belief is thus described as a function where peak deviates to the negative side, as in Fig. 1b.

With this background, hearer $B$ must determine speaker $A$'s intention as either $X = a$ or $X = \neg a$. Of course, an exact inference is impossible, so we must be content with calculating the probability distribution for $A$'s evaluation of $X$. We use, as a first-order approximation, the following algorithm:

$$P_I(X) = P_U(X) * P_S(X).$$
confirmed, as $Bel_{imp}$, while publicly confirmed shared beliefs are represented by $Bel_S$. As was mentioned previously, the information communicated through the dialogue, $I = E(U_A, Bel_S(X))$, should be included in $Bel_S(X)$, for the information thus communicated is what A intended to convey, and at the same time, what is generated as the result of B's interpretation process.

However, the situation is different when irony is used in the dialogue. Suppose that when generating $U_A$ on A's side, a probability distribution function $P_5$, like the one in Fig. 1b, is used as a shared belief. However, A can deny the fact that such a distribution function is used in generating utterance $U_A$. Of course this is possible only when $P_5$ is in $Bel_{imp}(X)$, but is not in $Bel_S(X)$.

In this case, the generation of a probability distribution $P_t$, like the one in Fig. 1c, is entirely B's responsibility. Formally, A need not commit to B's interpretation. Accordingly, the information communicated by irony cannot be registered in the shared beliefs $Bel_S(X)$, even if it is possible to register it in $Bel_{imp}(X)$. Moreover, the information thus communicated is exempt from the various axioms and rules of the dialogue.

In summary, irony can be used to achieve the apparently conflicting goals of: 1) informing the hearer of the speaker's unfavorable evaluation, and 2) making the information thus communicated exempt from the axioms and rules of the dialogue. Actually, this function is the essence of the rhetoric irony, and the function to convey the opposite to the surface meaning is merely a means to achieve the above goals.

4 Classification of Speaker's Intention in Irony Usage

In Sec. 3, we asserted that the main function of irony is to inform the hearer of the speaker's unfavorable evaluation, and, at the same time, not to register the communicated evaluation in the shared beliefs. By using...
this function, irony is used to achieve various goals of the speaker. In the following, we classify speaker's intentions in using irony.

Type 1 A coercive expression that inhibits counterargument

Formally speaking, the speaker is not responsible for the hearer's interpretation of the irony. It is thus difficult to argue against the speaker's evaluation implied by the irony. By using irony, a speaker can convey an unfavorable evaluation without the fear of objection by the hearer. This usage is especially powerful when the speaker's evaluation is unfair (such as blaming the hearer's personality for his single mistake). Dialogue D2 is an example of type 2. Another example of this type is:

D3: Taking important guests to a usually quiet restaurant, which happens to be noisy on that day,
One of the guests: "What a nice atmosphere this restaurant has!"

When expressed this way, the host has no way of making an excuse for this accidental situation.

Type 2 A sympathetic expression that respects the hearer's standpoint

As stated previously, the contents of an utterance is regarded as automatically confirmed unless they are objected to by the hearer. Irony can thus be used, in some situations, to relieve the hearer of confirming an unfavorable evaluation. Dialogue D1 is classified into this usage type.

D4: To a husband who returned very late at night,
Wife: "Why, you've come home so early today!"

If the wife knows that her husband cannot explain the reason for being so late, the above irony may be better than saying, "Where have you been until this late hour?".

Type 3 An expression that conveys a very irresponsible evaluation

In irony, the speaker does not need to commit to the hearer's interpretation. He can thus convey an extremely irresponsible evaluation by using irony.

D5: To someone who appears for a meeting dressed in a bright and cheerful manner,
"Oh, you're wearing a nice dress today."

With this utterance, the speaker can express disapproval about the way she is dressed, and at the same time be free from any kind of commitment. If she is a true friend, she might better advice her to dress in a more conservative manner by saying "I'm afraid your dress is a little too bright for today's meeting."

Type 4 Irony about the evaluation of the speaker himself

Use of irony is not restricted to the evaluation of the hearer. By using irony, speaker can express a kind of thoughtfulness when he has to inform the hearer of a bad evaluation of himself.

D6: When asked to be a partner in a tennis tournament, (suppose that both know the skill level of the asked person) "You know, I am a good player. Is that OK?".

In the above example, if the asked said "You know, I am a bad player. Is that OK?", the asker must make a corrective utterance. Otherwise, he would be regarded as accepting the evaluation of the asked, which would be impolite. By using irony the asked can free the asker making a diplomatic utterance like, "Oh, you're a good player."

5 Machine processing of ironic dialogue

A machine processing of ironic dialogue must include:

- Detecting discrepancies between the shared beliefs and the surface meaning of the utterance.
- Maintaining (both implicitly and publicly
confirmed) shared beliefs.
- Inferring speaker’s intention for using irony.
- Managing dialogues including ironic utterances.

Detecting discrepancies
First, an utterance is checked to see if it is ironical as follows.
1) Check if the utterance concerns the evaluation of some event X.
2) Check if the evaluation of X=a deviates to the good side compared to that expected from the shared beliefs.
3) Calculate probabilities \( P_U, P_S, \) and \( P_I \) as explained in Sec. 4.
4) If \( P_I(\neg a) \gg P_I(a) \) holds, then the utterance is regarded as ironic.

Maintaining shared beliefs
Once ironic utterance is detected, the shared beliefs on X are separated into publicly confirmed shared beliefs \( Bel_S(X) \) and implicitly shared beliefs \( Bel_{Imp}(X) \), and maintained separately. The surface meaning of the utterance, \( X=a, \) is registered in \( Bel_S(X) \), while the intended meaning, \( X=\neg a \) is registered in \( Bel_{Imp}(X) \).

Inferring speaker’s intention
Next, why the speaker used the ironic expression must be inferred. This is a difficult problem even for humans – misunderstanding of it often leads to a disastrous results. Here we propose an approximate method for inferring the intention only from the observable facts of the situation

This method tries to determine which of the usage types the irony utterance belongs to by using the questionnaire.

Questionnaire
Q0) Is the evaluation about the speaker?
Q1) Is the evaluation fair?
Q2) Is the evaluation acceptable to the hearer?
Q3) Is the information conveyed by irony useful to the hearer?
Q4) Does responsibility arise on the speaker’s side due to making the evaluative utterance?
Q5) Does the speaker need to make the evaluative utterance?
Q6) Is the speaker in the position to bestow a favor on the hearer?
Q7) Is a straightforward expression of the evaluation embarrassing to the hearer?
Q8) Is the situation serious?

Question Q0 distinguishes Type 4 from Types 1-3. Types 1-3 are classified as follows: First, Questionnaire Q1-Q8 is checked. Next, the point is calculated using the decision table in Table 1. For each usage type, the value in the table is added if the answer is Yes, and subtracted if it is NO to obtain the total points for each type. The one which earned the highest point is the usage type we are searching. The results for the sample dialogue of D1-D5 (D6 is excluded as it is of type 4) is given in Table 2, where the bold numbers are the type with the highest point.

Table 1. Decision table for irony usage type

<table>
<thead>
<tr>
<th></th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>-1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Q2</td>
<td>-2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Q3</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Q4</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Q5</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Q6</td>
<td>-1</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>Q7</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Q8</td>
<td>0</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2. Classification of ironies for example dialogues.

<table>
<thead>
<tr>
<th>Dialogue</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>-5</td>
<td>1</td>
<td>5</td>
<td>-3</td>
<td>1</td>
</tr>
<tr>
<td>Type 2</td>
<td>8</td>
<td>0</td>
<td>-4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Type 3</td>
<td>-2</td>
<td>6</td>
<td>-4</td>
<td>-6</td>
<td>6</td>
</tr>
</tbody>
</table>

The intention thus inferred is not exactly the intention of the speaker, but what is ex-
pected to be inferred by the hearer. However, sensible speakers are assumed to consider how their utterances are understood by the hearer. Hence, if the speaker makes an utterance knowing its effect on the hearer, we can say that he intended its effect.

Managing dialogues for ironic utterances

Lastly, an appropriate response should be generated to the ironic utterances. Possible strategies are as follows.

1) The hearer accepts the evaluation and responds accordingly. In Dialogue D1, the child could wash his face and/or explain the reason for his dirty face. This is a practical strategy if the irony is of type 2.

2) The hearer behaves as if he failed to understand the irony. This is a possible strategy in a situation where the hearer does not want to accept the speaker's evaluation. Of course, this strategy endangers the continuation of the dialogue.

3) The hearer tries to refute the speaker's evaluation. However, this is very difficult because the speaker can deny saying what the hearer is trying to refute. As we pointed out, making the refutation difficult is one of the main functions of the irony usage.

These strategies are just example strategies – it should be determined according to the purpose of the dialogue system being developed. Our aim here is to demonstrate how important the inference of the intention of the irony usage is to the management of ironic dialogues.

6 Summary and Discussion

We explained the mechanism of irony through the help of a dialogue model. According to this model, the essence of irony lies in allowing the transmission of the speaker's intention without having to register it in the shared beliefs, consequently inhibiting or excusing the hearer from counterarguing. We analyzed how irony is used for achieving various purposes and classified them according to their usage type.

Next we proposed an approximate method for determining the speaker's intention for using irony. To achieve a mechanical understanding of irony, this is very important. Without understanding the speaker's intention for using irony, we cannot cope with the situation appropriately, i.e., either to repel, or to accept obediently, etc. We are experimentally refining algorithms for inferring the speaker's intention from the situation of the ironic utterance.

As is easily observed, the heuristics we proposed for inferring the speaker's intention can be used for machine generation of irony. Suppose that the computer system wants to inform something without committing to what it says, it may want to use irony. Of course, it must first calculate the effect the usage of irony causes, and if the effect matches to what the system intends to attain, then the usage of irony can be considered. Possible intention for using irony includes the creation of informal atmosphere, soften the otherwise harsh error messages.

Probability \( P_{\neg a} \), on which our formulation is based, can be understood in the framework of the mention theory. Suppose that a speaker 1) wants to mention to the evaluation of \( X \), and at the same time 2) does not want to express it straightforwardly. The simplest solution to satisfy both is to reverse the proposition, which leads to our assumption for \( P_{\neg a} \). The mention theory focuses on how irony is generated by a speaker, while our analysis focuses on how it can be understood by the hearer. Therefore, our analysis is seen as the extension of the mention theory in the direction of pragmatics.

Our analysis is based on the use of irony by Japanese people. Even if irony can be understood beyond cultural differences, there may be some differences in how irony is used and accepted. This is also an interesting research theme, and involves separating the logic from idioms in irony.
References


Implicit Display Theory of Verbal Irony: Towards A Computational Model of Irony

Akira Utsumi
Department of Computational Intelligence and Systems Science
Tokyo Institute of Technology
4259 Nagatsuta, Midori-ku, Yokohama 226, Japan
utsumi@utm.dis.titech.ac.jp

Abstract
This paper proposes the implicit display theory of verbal irony that overcomes several difficulties of previous irony theories, and then describes a computational model of irony interpretation and generation based on the theory. The theory claims that irony implicitly communicates the fact that its utterance situation is surrounded by ironic environment which has three properties, but hearers can assume an utterance to be ironic even when they do not see all the three properties implicitly displayed by the utterance. Implicit communication of three properties is accomplished in such a way that an utterance alludes to the speaker's expectation, violates pragmatic principles, and is accompanied by several cues for implying the speaker's emotional attitude.

1 INTRODUCTION

Irony is one of the most ubiquitous ingredients of nonliteral (or humorous) language. Imagine the following situation and an utterance (1a):

[Situation 1] A mother asked her son to clean up his messy room, but he did a sloppy, half-hearted job. After a while, she discovered that his room is still messy, and said to her son:

(1) a. This room is very clean!

This utterance is a typical example of irony, and the speaker, a mother, intends to convey something other than what the words explicitly mean.

Then what do ironic utterances convey? One notorious answer provided by traditional accounts of irony is that irony communicates the opposite of the literal meaning. This account is problematic for several reasons, and what is worse, it leads to the misconception that irony is governed only by a simple inversion mechanism. Recent studies denying the meaning-inversion assumption have revealed that irony is a more complicated pragmatic phenomenon governed by several mental processes (e.g., Wilson and Sperber 1992; Kumon-Nakamura et al. 1995). However, they still do not give plausible answers to the following three essential questions: (Q1) what properties distinguish irony from non-ironic utterances?; (Q2) how do hearers recognize utterances to be ironic?; and (Q3) what do ironic utterances convey to hearers?

In the domain of computational linguistics, surprisingly little attention has been given to ironic uses of language, although other nonliteral language such as metaphor and implicature has been a popular topic (Fass et al. 1991). However, it is important to construct a computational model of irony for at least two reasons: irony, which is heavily dependent upon context, can throw new light on computational studies of many pragmatic phenomena, and it offers an effective way of accomplishing various communication goals that are difficult to convey literally (Roberts and Kreuz 1994).

The author has been exploring the mechanism of recognizing and interpreting irony and trying to construct a computational model of irony. This paper presents the implicit display theory, a unified theory of irony that answers the three questions (Q1)~(Q3), and illustrates how irony is interpreted and generated by computer. The theory provides a computationally feasible framework of irony as the first step toward a full-fledged computational model of irony, and it can account for several empirical findings from psycholinguistics. The essential idea underlying the theory is that an ironic utterance implicitly displays ironic environment, a special situation which has three properties for being ironic — expectation, incongruity, emotional attitude —, but the hearer does not have to see all the three properties implicitly communicated in order to recognize the utterance to be ironic.

This paper focuses only on verbal irony which can be distinguished from situational irony (or irony of fate). Situations are ironic when an expectation is violated in specific ways and such
nonverbal irony appears even without verbal communication. Note that situational irony can be indicated by metareferential expressions like "it is ironic that...", but those expressions themselves are not ironic.

This paper is organized as follows: Section 2 discusses the problems of previous irony theories and then Section 3 explains, in detail, implicit display theory of irony that overcomes the difficulties. Section 4 gives a computational formalization of the theory and a basic framework for irony interpretation and generation.

2 Previous Approaches to Irony

A number of studies have so far tackled the three questions (Q1)–(Q3). They are classified into two essentially different approaches: the pragmatic approach and the cognitive approach.

The traditional pragmatic approach (Grice 1975; Searle 1979) assumes that an utterance is recognized to be ironic when the hearer becomes aware of an apparent violation of some pragmatic principles (e.g., the maxim of quality or sincerity conditions for speech acts), and as a result it conveys the opposite of the literal meaning. This view applies well to typical ironies such as (1a), but it completely fails to explain what irony is, how irony is recognized and what irony communicates. First, irony can be communicated by various expressions that do not include such violation. For example, the following utterances — a true assertion (1b), a understatement (1c), and an overpolite request (1d) — are interpreted as ironic under Situation 1 though they violate neither the maxim of quality nor any felicity conditions.

(1) b. I love children who keep their rooms clean.
   c. This room may be slightly messy.
   d. Would you mind if I asked you to clean up your room, please?

Also, Peter’s echoic reply (2a) of the following exchange is also one of ironies without such violation.

[Situation 2] Just after his colleague Jesse said to him
   "I’d be promoted before you", Peter replied:

(2) a. You’d be promoted before me, huh?

Moreover, in some cases, ironic intent can be communicated under the situation where hearers are not aware of any violation. For example, Mary can perceive Peter’s utterance (3) in Situation 3 as ironic even when she does not at all think that she is not a good cook. All these facts imply that violation of such principles is not a necessary condition of irony and the traditional approach cannot give an answer to (Q2).

[Situation 3] Eating an unsavory fruit cake made by Mary, who is proud of her skill in cooking, Peter says:

(3) You are a good cook indeed.

Although Haverkate (1990) extended a target of intentional violation to felicity conditions for five speech act classes, it is still incomplete for the same reasons. Secondly, the traditional approach cannot discriminate irony from other non-literal utterances (e.g., a lie) in which the maxim of quality is flouted, and therefore does not provide a plausible answer to (Q1). Finally, the notion of “the opposite of the literal meaning” is problematic because it is applicable only to declarative assertions but many ironic utterances can take non-declarative forms: questions such as (4c); requests such as (1d); offerings such as (4e); and expressives such as (5a).

[Situation 4] Candy baked a pizza to satisfy her hunger. When she was dishing it up, her husband entered the kitchen and gobbled up the whole pizza. Candy said to her husband:

(4) a. I’m not hungry at all.
   b. I’m really satisfied to eat the pizza.
   c. Have you seen my pizza on the table?
   d. I’ll get to sleep.
   e. How about another small slice of pizza?

[Situation 5] Peter broke his wife’s favorite teacup when he washed the dishes awkwardly. Looking at the broken cup, his wife said:

(5) a. Thank you for washing my cup carefully.
   b. Thank you for crashing my treasure.

Mention theory proposed by Sperber and Wilson (Sperber and Wilson 1981, 1986; Wilson and Sperber 1992) is the first theory to focus on the allusive nature of irony. They have argued that verbal irony is a variety of echoic mention of an attributed thought or utterance. By mentioning or alluding to someone’s thought, utterance, expectation or cultural norm, irony communicates a speaker’s attitude toward a discrepancy between what actually is and what has been expected. This view of irony is shared by other cognitive approaches such as Kreuz and Glucksberg’s (1989) echoic reminder theory, and it has been supported by psychological experiments (Jorgensen et al. 1984). These theories, however, are still incomplete as a comprehensive framework for irony for at least three reasons. First, their concepts of mention/allusion — Sperber and Wilson’s echoic interpretation and Kreuz and Glucksberg’s echoic reminder — are too narrow to capture the allusive nature of irony. For example, Nancy’s utterance (6a) is an echoic interpretation of Nancy’s expectation of the fine weather, but (6b) does not interpretively echo any states of affairs: (6b) is an
implication derived from the failed expectation.

[Situation 6] Nancy and Jane were planning a trip to the beach, but that day was a cold and stormy one. As she looked out the window, Nancy said:

(6) a. Oh, the weather is really nice.
   b. Maybe the beach is crowded with people.

Likewise their theories cannot explain what (4c) and (4e) allude to. Thus their notions are not a necessary condition of irony. Second, these theories provide no plausible explanation of how irony is discriminated from non-ironic echolalic utterances, and therefore do not provide a plausible answer to (Q1). Finally, they implicitly assume that properties that characterize irony can be applied to recognition of ironic utterances as they stand or they do not focus on how hearers recognize utterances to be ironic. Thus they cannot also explain a certain kind of ironic utterances in which hearers are not aware of any pragmatic violation.

Allusional pretense theory (Kumon-Nakamura et al. 1995) integrated these two approaches and claimed that all ironic utterances allude to a failed expectation and violate one of the felicity conditions for well-formed speech acts. This theory has a powerful ability to explain more ironic utterances than previous studies, resolving some difficulties of previous studies. However, allusional pretense theory still suffers from the same disadvantage as other theories: 1) their notion of allusion is not clear enough; 2) some ironies — e.g., (1b), (1c), (2a) — do not include any violation of the felicity conditions as we described above; 3) they do not explain how hearers recognize utterances to be ironic; and 4) the role of ironic cues (e.g., ironic tone of voice) is not addressed.

Clark and Gerrig's (1984) pretense theory takes a rather different approach to irony: "in being ironic, the theory goes, a speaker is pretending to be an injudicious person speaking to an uninhibited audience; the speaker intends the addressees of the irony to discover the pretense and thereby see his or her attitude toward the speaker, the audience and the utterance" (ibid., p.121). According to pretense theory, each and every time the speaker says irony he/she is pretending to be an injudicious person, which corresponds to the victim of irony. For example, this theory explains ironies in Situation 4 in a way that Candy, the speaker of '4a)~(4e), is pretending to her husband who ate Candy's pizza, exaggerating how ridiculous his behavior is. However, that there are ironies that do not have victims such as (6a) makes pretense theory less convincing. Furthermore, as Kreuz and Glucksberg (1989) pointed out, the notion of pretense is too powerful for a comprehensive theory of irony in that it can be applied to all indirect speech acts. Though pre-
tense theory claims to be superior to the mention theory, there is a negative suggestion, which we agree with, that both theories may not differ significantly from one another (Williams 1984).

From the above discussion, we have made the following observations on a comprehensive theory of irony which will be presented in Section 3.

- The notions of allusion and pragmatic violation are essential to irony. However, previous theories are too general in that each of their proposed properties for being ironic also covers a part of non-ironic utterances, and at the same time, too specific in that each of them does not cover all ironic utterances. To make an adequate theory that explains both what irony is and what irony is not, we must integrate allusion, pragmatic violation, and other properties of irony into the theory.
- All the previous theories make the same mistake in that they confuse the two different questions (Q1) and (Q2), since there are several cases like Situation 3 in which after recognizing a given utterance to be ironic the hearer becomes aware that one of properties is satisfied. Thus the theory of irony must distinguish conditions for perceiving an utterance as ironic from those for an utterance being ironic.

3 Implicit Display Theory of Irony

3.1 What Is Irony?

The implicit display theory claims as an answer to (Q1) that irony is an utterance which implicitly displays to the hearer the situation surrounded by ironic environment. It can be divided into two parts: ironic environment as a situational property and implicit display as a linguistic property.

3.1.1 Ironic Environment

In order for an utterance to be ironic, a speaker must utter in a situation surrounded by ironic environment. Given two temporal locations $t_0$ and $t_1$ such that $t_0$ temporally equals or precedes $t_1$, a situation in which an utterance is given is surrounded by ironic environment if and only if it satisfies the following three conditions:

1. The speaker has an expectation $E$ at $t_0$.
2. The speaker's expectation $E$ fails (i.e., $E$ is incongruous with the reality) at $t_1$.
3. As a result, the speaker has a negative emotional attitude toward the incongruity between what is expected and what actually is.

Note that the notion of expectations here subsumes culturally expected norms and conven-
Table 1: Allusion of ironic utterances

<table>
<thead>
<tr>
<th>Conditions for allusion</th>
<th>Utterances satisfying the condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. $U \sim Q$ and $U \not\sim P_B$</td>
<td>(1a) (1b) (2b) (3) (4a) (6a)</td>
</tr>
<tr>
<td>2. $U \sim A$ where $A$ is an action that brings about $Q$ (type-1, type-2)</td>
<td>(1d) (4b)</td>
</tr>
<tr>
<td>3. $U \sim X$ where $X$ is a premise of $Q$ or a precondition of $A$</td>
<td>(1c) (4c)</td>
</tr>
<tr>
<td>4. $U \sim Y$ where $Y$ is an implication derived from $Q$</td>
<td>(4d) (6b)</td>
</tr>
<tr>
<td>5. $U \sim B, W$ or $Z$ where $B$ is an action that precludes $A, W$ is a premise of $B$ and $Z$ is an effect of $B$ (type-1, type-3)</td>
<td>(2a) (2c) (4e) (5a) (5b)</td>
</tr>
</tbody>
</table>

Ironic environment can be classified into the following four types.

- a speaker’s expectation $E$ can be caused by an action $A$ performed by intentional agents
  - $E$ failed because $A$ failed or cannot be performed by another action $B$ (type-1)
  - $E$ failed because $A$ was not performed (type-2)

- a speaker’s expectation $E$ is not normally caused by any intentional actions
  - $E$ failed by an action $B$ (type-3)
  - $E$ accidentally failed (type-4)

For example, ironic environment of Situation 4 falls in type-1: Candy’s expectation can be realized by an action of eating a pizza, but her husband’s action of eating the whole pizza hindered her expected action. In the same way, ironic environment of Situation 1 falls in type-2, those of Situations 2, 3 and 5 fall in type-3, and that of Situation 6 falls in type-4.

3.1.2 IMPLICIT DISPLAY

An utterance implicitly displays all the three conditions for ironic environment when it
1. alludes to the speaker’s expectation $E$.
2. includes pragmatic insincerity by violating one of pragmatic principles, and
3. implies the speaker’s emotional attitude toward the failure of $E$.

For example, utterances (1e) and (1f) for Situation 1 are not ironic even when they are given in the situation surrounded by ironic environment:

(1e) I’ve expected a clean room.
(1f) I’m disappointed with the messy room.

(1e) and (1f) directly express the speaker’s expectation and the speaker’s emotional attitude, respectively, and both do not include pragmatic insincerity. On the other hand, all ironic utterances implicitly express the three components of ironic environment, as we will see below.

3.1.3 ALLUSION

We give a formal definition of allusion to $E$. Given $P$ expressing the propositional content of $U$, $Q$ expressing the speaker’s expected event/state of affairs, and $P_B$ expressing the state of affairs that the speaker expects $E$, an utterance $U \text{ alludes to}$ the expectation $E$ if and only if it satisfies one of the conditions shown in Table 1. The relation $U \sim X$ in Table 1 holds if $U$’s propositional content $P$ and $X$ are conceptually/semantically identical or unifiable (i.e., $P = X$) or if $U$ includes a referring expression $U_r$ (i.e., phrases or words) whose referent is $X$. For example, the following ironic utterances for Situation 1 show a variety of the relation $U \sim Q$:

(1a) This room is very clean! ($P = Q$)
(1b) I love children who keep their rooms clean.
(g) How do you feel in such a comfortable place?

The notion of allusion here is wider than Sperber and Wilson’s echoic mention/interpretation and clearer than that of allusional pretense theory. The definition allows ironic utterances to allude speaker’s expectations, but it does not allow (1e) to allude to it because of the condition $U \not\sim P_B$. Table 1 also shows which condition each of ironies presented in this paper satisfies. For Example 4, the irony (4c) that mention theory cannot explain alludes to Candy’s expectation by the phrase “my pizza on the table” referring to one of the conditions $X = “Candy’s pizza was on the table” for an action $A = “eat Candy’s pizza”$. Other four utterances for Situation 4, (4a), (4b), (4d) and (4e), also refer to $Q, A, Y, B$, respectively.

3.1.4 PRAGMATIC INSINCERITY

Table 2 lists the pragmatic principles violated by the ironic utterances in this paper. In many cases an ironic utterance is pragmatically insincere in the sense that it intentionally violates one of the preconditions (i.e., sincerity, preparatory and propositional conditions) that need to hold before its illocutionary act is accomplished, but pragmatic insincerity also occurs when an ut-
Table 2: Pragmatic principles violated by ironic utterances

<table>
<thead>
<tr>
<th>Violated pragmatic principles</th>
<th>Utterances violating the principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sincerity condition for Inform (S believes P)</td>
<td>(1a) (2a) (2b) (3) (4a) (4b) (4d) (6a) (6b)</td>
</tr>
<tr>
<td>for Question (S does not want to know P)</td>
<td>(4c)</td>
</tr>
<tr>
<td>for Offer (S wants to do an action P for H)</td>
<td>(4e)</td>
</tr>
<tr>
<td>for Thank (S feels grateful for an action P)</td>
<td>(2c) (5b)</td>
</tr>
<tr>
<td>Propositional content condition for Thank (P is a past action done by H)</td>
<td>(5a)</td>
</tr>
<tr>
<td>Preparatory condition for Question (S does not know P)</td>
<td>(4c)</td>
</tr>
<tr>
<td>for Offer (S is able to do an action P)</td>
<td>(4e)</td>
</tr>
<tr>
<td>Maxim of relevance (P is relevant in Sperber and Wilson’s (1986) sense)</td>
<td>(1b)</td>
</tr>
<tr>
<td>Politeness principle (U should be made at an appropriate level of politeness)</td>
<td>(1d)</td>
</tr>
<tr>
<td>Maxim of quantity (P is as informative as required)</td>
<td>(1c)</td>
</tr>
</tbody>
</table>

Notes: S, H and P denote the speaker, the hearer and the propositional content, respectively.

Table 3: Examples of cues for implicitly displaying emotional attitudes

| Verbal cues | 1. hyperbole, exaggeration — adjectives (e.g., amazing, splendid), adverbs (e.g., certainly, really, absolutely), metaphors 2. interjection — “Oh!”, “ah!”, “O!”, “Dear me!”, “Oh dear!”, “huh” 3. prosody(paralinguistic cues) — accent, intonation contour, exaggerated stress, slow speaking rate, tone of voice, nasalization 4. speaker’s counterfactual pleased emotions — thank, compliment |
| Nonverbal cues | 1. facial expression — quizzical, sneering, deadpan 2. gesture |

terance violates other pragmatic principles. Requests often become insincere when they are overpolite like (1d) since they violate the politeness principle (although (1d) also becomes insincere when a mother no longer intends her son to clean up his room). Understatements like (1c) are also insincere since they do not provide as much information as required. The true assertion (1b) violates the principle of relevance in that it does not yield any contextual implication. As mentioned earlier, the last three cases have been problematic for all the previous theories of irony because none of these theories recognized that a wide variety of principles is violated by ironic utterances.

3.1.5 EMOTIONAL ATTITUDE

Speakers can use a variety of cues/signals listed in Table 3 for implicitly communicating their emotional attitude. The use of the interjection “Oh” with a special tone of voice in (6a) offers one typical example of this. Implicit communication can also be accomplished by utterances explicitly referring to the pleased emotion that speakers would experience if their failed expectation became true, which corresponds to verbal cue 4 of Table 3. For example, the utterance (5a) is accompanied by this cue since it explicitly expresses the speaker’s counterfactual emotion.

3.2 HOW IS IRONY RECOGNIZED?

In many cases, all the three components for implicit communication of ironic environment are easily recognized by the hearer. As we mentioned in Section 2, however, there are also many cases such as Situation 3 that an utterance can be ironically interpreted even though all the three components cannot be recognized by the hearer because the hearer’s mental situation differs from the speaker’s one. Furthermore, in the case of (2a), Jesse may not become aware of pragmatic insincerity of (2a) until he perceives the utterance as ironic.

Hence the implicit display theory adopts the following condition as an answer to (Q2):

Hearers can assume an utterance to be ironic when they recognize that the utterance implicitly displays at least two of the three components for ironic environment, and if the utterance situation does not rule out the possibility of including the unrecognized components, if any.

This “2-of-3” criterion1 for recognizing irony explains the above phenomenon, and makes it possible that hearers can recognize utterances as ironic even though speakers do not intend their utterances to be understood as irony. It provides empirical evidence of the theory since such uninten-

1 Practically speaking, this classificatory definition is too strong to distinguish irony from non-ironic utterances, since whether an utterance is ironic is a matter of degree. As a better criterion for recognizing irony, a quantitative method for evaluating the degree of ironicalness will be proposed in Section 4.3.
tional irony has been found in a number of psychological experiments (Gibbs and O'Brien 1991).

3.3 **What Does Irony Convey?**

By recognizing an utterance to be ironic, the hearer understands an intended illocutionary act that the speaker performs in saying irony: the act of informing the fact that the utterance situation is surrounded by ironic environment (i.e., all the three components for ironic environment hold in a current situation). That is an answer to (Q3). By understanding the illocutionary act, the hearer turns out to know that the informed fact is shared by both the hearer and the speaker.

In many cases, however, since the hearer already knows that the three components hold in the situation, interpretation of irony results in confirmation of the most uncertain information, that is, the speaker's emotional attitude. That is why previous irony theories regard the speaker's negative attitude as what irony communicates. Therefore, when the hearer does not recognize all components, he/she also obtains new information that the unrecognized component holds in a current situation. For example, in the case of (2a), after recognizing Peter's utterance (2a) to be ironic, Jesse turns out to know that Peter thinks Jesse's preceding utterance is absurd, and tries to confirm Peter's emotional attitude.

Furthermore, together with that illocutionary act, irony achieves various communication goals intended by the speaker as perlocutionary acts, some of which are as follows (Roberts and Kreuz 1994):

- to give pain
- to emphasize a point
- to get attention
- to be humorous
- to clarify
- to dissemble

3.4 **Implications of The Theory**

3.4.1 **Echoic Irony and Expectation**

It appears that some particular kind of echoic irony like (2a), which Cutler (1974) called "provoked irony", does not allude to speaker's expectation, since its propositional contents must be assumed to be desirable less from the point of view of the speaker than from that of the hearer. Therefore, in order to explain provoked irony, previous theories (e.g., mention theory and echoic reminder theory) assume that what such irony alludes to is not a speaker's expectation but other person's thought or utterance. However, what might be the echoed materials of the following utterances that can be interpreted as ironic under the same situation as (2a)?

(2b) You are very smart, so you already think I'd be promoted before you.
(2c) Thank you for informing your valuable opinion.

To overcome this difficulty, implicit display theory argues that provoked irony also alludes to the speaker's expectation: expectation that "the hearer knows his/her own utterance or thought is false". Such expectation is often brought up by the hearer's preceding utterance, and at the same time (i.e., t0 = t1), the ironist becomes aware that the expectation is incongruous with the reality. Hence we may safely say that the act of saying the hearer's thought or opinion corresponds to B, an action that makes E incongruous.

In the case of Situation 2, by hearing Jesse's utterance of "I'd be promoted before you", Peter has an expectation of convincing Jesse that Jesse would not be promoted before Peter, knows that Jesse does not think his own utterance is false, and consequently, the situation is surrounded by ironic environment. Provoked irony (2a) alludes to that expectation by referring to one of the effects Z = "Peter believes that Jesse would be promoted before Peter", which would normally be brought about by an action B of "informing that Jesse would be promoted before Peter", and at the same time, it violates the sincerity condition of Inform in that Z is false. Also, (2b) and (2c), which do not allude to one's utterance/thought, can be explained within implicit display theory: both allude to that expectation, because the former refers to Peter's expectation Q and the latter refers to Jesse's action B.

You may doubt the appropriateness of the above explanation, but Kaufer (1981) mentioned the similar view in the discussion of what assumptions about the context ironize clearly false utterances like "Columbus discovered America in 1900". He argued that, in order to be perceived as irony, such utterance must be given in the contextual setting in which "the ironist knows the utterance is false (and thus rejects it), knows that the addressee does not know this, and (most importantly) also believes that the latter should know it" (ibid., p.503). The last assumption clearly corresponds to the speaker's expectation of provoked irony explained above.

3.4.2 **Ironic Cues**

Several theories assume that irony can be identified by a number of cues specific to irony, in particular, ironic tone of voice and intonational cues. Although there is little doubt that such intonational cues may accompany ironic utterances, the empirical finding in psychology (Gibbs and
O'Brien (1991) has shown that people can interpret ironic statements without any special intonational cues. Furthermore, both ironic and non-ironic interpretations are derived from the use of the same intonational contour in different contexts (Ward and Hirschberg 1985). These observations imply that such cues themselves are neither sufficient nor necessary conditions for irony. Implicit display theory is consistent with this implication since these cues are only a part of Component 3 for implicit display.

Kreuz and Roberts (1995) have addressed another cue for irony, hyperbole, and they claimed that ironic tone of voice is nothing more than the use of hyperbole. However, the same discussion can be done about the hyperbolic features of irony: there are ironies that do not include hyperbole such as (1c). Implicit display theory also covers hyperbolic features of irony and explains non-hyperbolic irony.

3.4.3 Victims of Irony

Several studies (e.g., Clark and Gerrig 1984) have pointed out that most (but not all) ironies have victims and thus that the notion of victims may be an important property for irony. Implicit display theory implies that ironic utterances have potential victims when their ironic environments fall in one of types 1, 2, 3: in the case of type 1 or type 3 an agent of B becomes a victim, and in the case of type 2 an agent of A becomes a victim.

3.4.4 Sarcasm and Irony

Sarcasm is a figure of speech designed to cut or give pain and often expressed by verbal irony. Thus explicit victims of irony often become the target of sarcasm, and by displaying the speaker’s counterfactual pleased emotion, which is one of cues for emotional attitudes, sarcasm has the effect of giving the target pain. We argue that both are distinctive properties of sarcasm. For example, (5a) and (5b) are sarcastic irony because they have an explicit victim, Peter, and they refer to the wife’s counterfactual pleased emotion.

4 Towards a Computational Model of Irony

4.1 Representation

In order to represent ironic utterances and ironic environment, this paper uses situation theory (Barwise 1989) and situation calculus. As an example, Figure 1 illustrates the representation of ironic environment of Situation 4. In the figure, all events states of affairs are represented as pairs $s \models \sigma$ of an infon $\sigma$ and a situation $s$. For example, given that the situation $t_0$ expresses a spatiotemporal location and $x$ denotes “Candy”, the state of affairs that Candy is hungry is represented as $t_0 \models \langle \text{hungry}, x \rangle$ and its negation as $t_0 \models \langle \text{hungry}, x \rangle$. Actions are expressed by predicates: for example, an action of eating the pizza preformed by Candy’s husband is expressed by the predicate $\text{eat}(y, a)$. The state of affairs that an action $A$ is performed is expressed by $\langle \text{did}, A \rangle$. Furthermore, a proposition $p$ expressing the claim that $s \models \sigma$ is written as $(s \models \sigma)$. The proposition $p$ is true if $s$ supports $\sigma$, and otherwise false. Infons and actions can include parameters denoted by capital letters. Parameters can be restricted by infons: for example, $\langle \text{precedes}, a,T \rangle$ is a parameter for temporal situations which temporally succeed $t_0$. A causal relation between two events $s_1 \models \sigma_1$ and $s_2 \models \sigma_2$ is expressed by $s_1 \models \sigma_1 : [A] \Rightarrow s_2 \models \sigma_2$. This relation means that if an action $A$ is executed in a situation $s_1$ supporting the infon $\sigma_1$, then it causes the infon $\sigma_2$ to be true in the resulting situation $s_2$. When we omit an action $A$ from a causal relation, that relation becomes a constraint in situation theory, denoted by $s_1 \models \sigma_1 \Rightarrow s_2 \models \sigma_2$.

A conversational participant such as a hearer and a speaker has a finite number of beliefs in his/her mental situation that are manifest, then perceptible and made available to him/her. An agent X’s mental situation is represented as a situation $u_X$ and his/her beliefs as support relations between $u_X$ and infons. For example, the fact
Speech Act: Inform(S, H, P)
Preconditions: $\langle \text{proposition}, P \rangle$, $u_S = P$
Effects: $u_H \models \langle \text{intend}, S, \text{Convince}(S, H, P) \rangle$

Speech Act: Request(S, H, P)
Preconditions: $u_S \models \langle \text{want}, S, \text{Know}(S, P) \rangle$, $\langle \text{proposition}, P \rangle$, $\neg \text{Know}(S, P)$
Effects: $u_H \models \langle \text{intend}, S, \text{Inform}(H, S, P) \rangle$

Speech Act: Thank(S, H, P)
Preconditions: $\langle \text{action}, P \rangle$, $\langle \text{agent}, H, P \rangle$, $T \models \langle \text{did}, P \rangle$, $\langle \text{proceeds}, T, T_0 \rangle$
Effects: $u_H \models T \models \langle \text{grateful}, S, H, P \rangle$

Notes: $S$ and $H$ denote the speaker and the hearer, and thus $u_S$ and $u_H$ denote the speaker's and hearer's mental situations. Also, $T_0$ denotes the utterance situation, $\text{Know}(S, P) = u_S \models P \lor \neg P$, and $\neg \text{Know}(S, P) = u_S \not\models P \land \neg P$.

Figure 2: Action schemes for speech acts

that Jim believes/knows Candy is hungry is represented as $u_{int} \models t_0 \models \langle \text{hungry}, x \rangle$. Although Figure 1 does not include any mental situations (i.e., ironic environment is represented from god's eye view), when Candy intends the utterance to be ironic her mental situation must support all states of affairs, events and causal relations in this figure. These manifest beliefs constitute a context, which corresponds to what Sperber and Wilson (1986) call a cognitive environment.

An utterance $U$ is characterized by its propositional content $P$ and the illocutionary force $F$ that the speaker performs in saying $U$. For example, (4c) consists of the propositional content $P = \langle t_1 \models \langle \text{see}, y, T_0\langle \text{proceeds}, T, T_1 \rangle \models \langle \text{loc}, a, h_1 \rangle \rangle \rangle$ and the illocutionary act $F = \text{Request}$. These illocutionary acts are defined as action schemes as in (Litman and Allen 1987), some of which are shown in Figure 2. However, in ordinary conversations, the intended illocutionary act of utterances often differs from such literal illocutionary acts. Thus in this paper literal illocutionary acts are denoted by $F_1$ and the intended illocutionary acts of utterances by $F_i$. In the case of irony, as we mentioned in Section 3.3, $F_i$ is to inform the fact that the utterance situation is surrounded by ironic environment.

Perlocutionary communication goals that ironists usually intend to achieve by irony may affect the decision of how implicit display is accomplished (i.e., which proposition should be referred to, which pragmatic principle should be violated, and what cues should be used to convey the attitude). Conversely the hearer can infer what com-

---

2Prosodic and nonverbal features such as in Table 3 must also be taken into account for spoken ironic language, but computational formalization of these features is beyond the scope of this paper.

InterpretIrony($U, W$)
0. $G \rightarrow \phi$
1. Identify the propositional content $P$ of $U$ and its surface speech act $F_1$.
2. Identify the three components for implicit display of ironic environment as follows:
   (a) Allusion — If the speaker's expectation $E$ is included in $W$, find out the referring expression $U_0$ in $U$ and the referent $R$. If $E$ is not included, assume $U_0 = U$.
   (b) pragmatic insincerity — Find out what pragmatic principle is violated by $U$.
   (c) emotional attitude — Detect verbal/nonverbal expressions that implicitly display the speaker's attitude.
3. Calculate the degree of irony $d(U)$ of $U$.
4. If $d(U) >$ a certain threshold $C_{irony}$, then
   (a) Infer the speaker's emotional attitude.
   (b) Infer the expectation $E$, if necessary.
   (c) Add $F_1$ (to inform that $W$ includes ironic environment) to $G$.
5. Recognize communication goals achieved by irony, and add them to $G$.

Figure 3: Irony interpretation algorithm

communication goals are intended by the speaker by examining the way of implicit display. For example, to achieve the goal give pain by irony, the following strategies can be applied:

- $U$ refers to the victim's action, its premises, or its effects.
- $U$ has a surface speech act Thank.
- $U$ expresses the counterfactual pleased emotion toward the victim's action.

4.2 ALGORITHMS FOR INTERPRETING AND GENERATING IRONY

A rough sketch of an interpretation algorithm is given in Figure 3. The algorithm takes as input an utterance $U$ under interpretation and a hearer's context $W$, and it produces a set of goals $G$ the speaker has intended to achieve by $U$. It must be noted that we do not argue here that irony interpretation is an optional process done before other (literal) interpretation process (Usumu 1995). The interpretation algorithm checks whether each of the three conditions for implicit display is satisfied by $U$, and then calculates the degree of irony $d(U)$ by the method described in Section 4.3. At line 4, the algorithm judges that $U$ is ironic when $d(U)$ exceeds a given constant $C_{irony}$. The degree of irony takes a real value between 0 and 1, and thus this criterion can be seen as a comparative version of the recognition condition described in Section 3.2. If $U$ is judged ironic, the algorithm then rea-
\[
\langle \text{hope}, X, (S \models I) \rangle \iff \langle \text{want}, X, (S \models I) \rangle \\
\text{a predict}, X, (S \models I) \rangle
\]
\[
S_1 \models \langle \text{disappointed}, X, (S_1 \models I) \rangle \iff \\
S_0 \models \langle \text{hope}, X, (S_0 \models \overline{I}) \rangle \land S_1 \models I \land \\
S_1 \models S \land \langle \text{precedes}, S_0, S_1 \rangle \land \\
\langle \text{precedes}, S_0, S \rangle \land \langle \text{agent}, A, X_2 \rangle \land \\
S_0 \models \langle \overline{A} \rangle \Rightarrow S_1 \models I \land \\
S_1 \models \langle \text{blameworthy}, A \rangle
\]

Figure 4: Emotion-eliciting rules

\[
\text{GenerateIronic}(G, W)
\]

\[
0. \ G_P \rightarrow ^*
\]

1. Select possible subgoals (plans) for achieving \( G \), and add them to \( G_P \).

2. Construct pairs \((P, F)\) that implicitly display the ironic environment in \( W \) so that they satisfy \( G_P \).

3. Select the pair \((P^*, F^*)\) with the highest degree of ironicalness among them.

4. Construct \( U \) from \((P^*, F^*)\).

Figure 5: Irony generation algorithm

sons about the speaker’s emotional attitude using emotion-eliciting rules, some of which are shown in Figure 4. Moreover, the algorithm tries to recognize communication goals using relations between the way of implicit display and the goals.

On the other hand, the generation algorithm shown in Figure 5 takes as input the set of the speaker’s communication goals \( G \) (including the ironic illocutionary goal \( P_1 \)) and a context \( W \), and then it generates an ironic utterance \( U \) appropriate to the communication goals and to the ironic environment. More concretely, the algorithm first selects more specific goals that serve for many decisions of what should be said ironically and of how it should be said. To take a simple example, a communication goal “to give pain” is decomposed into several subgoals such as “to refer to actions or events ascribed to the victim” and “to directly express the counterfactual pleased emotion.” The next step is that the algorithm decides what to say and how to say ironically. The propositional content is mainly concerned with the decision of what should be referred to by irony, while the literal illocutionary act is mainly concerned with the decision of which pragmatic principle should be violated. Decision of what cues should be used to convey the attitude mainly affects stylistic and prosodic features. These decisions are done so that the selected subgoals are satisfied and that their degrees of ironicalness is much more than a threshold.

4.3 A METHOD FOR ASSESSING THE DEGREE OF IRONICALNESS

To evaluate the degree of ironicalness of \( U \), we use the following measures whose values are real numbers ranging from 0 to 1:

1. For the allusiveness of \( U \):
   (a) context-independent desirability of the referring expression \( U_R: \ d_1 \)
   (b) degree of similarity between \( Q \) and \( R: \ d_2 \)
   (c) expectedness of \( E: \ d_3 \)
   (d) indirectness of expressing the fact that the speaker expects \( E: \ d_4 \)

2. For pragmatic insincerity of \( U \):
   (e) degree of pragmatic insincerity of \( U: \ d_5 \)

3. For emotional attitudes in \( U \):
   (f) degree to which \( U \) implies the speaker’s attitude: \( d_6 \)
   (g) indirectness of expressing the attitude: \( d_7 \)

The desirability \( d_1 \) reflects the asymmetry of irony that many studies have pointed out: more positive utterances are more likely to be ironic. On the other hand, some studies (Kreuz and Glucksberg 1989; Kumon-Nakamura et al. 1995) suggested that a negative utterance can also convey irony if the speaker’s negative expectation is incongruous with a positive event and if that is obvious to both the speaker and the hearer. The measure \( d_3 \) reflects this fact: whether they are positive or negative, utterances which allude to stronger or more obvious expectations are more likely to be ironic. Hence, although \( d_1 \) of the following irony (7) is much less than that of (1a), \( d_3 \) of (7) is nearly equal to that of (1a).

[Situation 7] A wife asked her husband to scold their son for leaving his room messy. They entered into his room and found that it was sparkling clean.

As he looked around the room, her husband whispered to her:

(7) This room is certainly dirty!

Furthermore, the expectedness \( d_5 \) reflects an intuitively plausible hypothesis that personal expectations should be stronger than culturally/socially expected norms and conventions. The similarity \( d_2 \) serves to evaluate to what degree an utterance alludes to an expectation. It implies that an utterance referring to \( Q \) must have the greatest similarity \( d_2 = 1 \), and that the similarity of an utterance referring to the action \( A \) or \( B \) might be greater than that of an utterance referring to other propositions. When the interpretation algorithm does not decide a speaker’s expectation \( E \), it can be assumed that \( d_3 = d_7 = 0 \). The last measure for allusion, the indirectness \( d_4 \), is necessary for ruling out
non-ironic utterances that directly express the speaker's expectation. Thus the value of $d_4$ for (1e) is 0.

The measure $d_5$ is equal to 0 when utterances do not violate any pragmatic principles, and otherwise it is greater than 0. For example, $d_5$ of provoked irony (2a) is 0 on the hearer Jesse's side, while it is much greater on the speaker Peter's side. In particular, $d_5$ should be 1 when utterances violate the sincerity condition like (1a), since it is the most frequently violated principle by irony. Likewise, the measure $d_5$ is 0 when utterances do not include any cues listed in Table 3, and otherwise it is greater than 0. The last measure $d_7$ is also necessary for ruling out non-ironic utterances like (1f) that directly express the speaker's attitude.

Using these seven measures, we define the degree of ironicalness $d(U)$ of an utterance $U$ as follows:

$$d(U) = d_4 \cdot d_7 \cdot \left( \frac{d_1 + d_2 + d_3 + d_4 + d_5 + d_6}{3} \right)$$

This equation means that direct expressions of expectations and of emotional attitudes cannot be ironic even if they implicitly display other components. Three measures $d_1 \sim d_3$ for allusion are averaged so that the three components for implicit display can equally contribute to the degree of ironicalness.

5 Concluding Remarks

This paper has presented implicit display theory of verbal irony and illustrated a rough sketch of a computational model of irony. Through the discussion about many features of irony, we have confirmed the superiority of the theory over existing irony theories and also shown that irony is a more complicated pragmatic phenomenon than was supposed. To deal with irony by computer, several computational techniques in NLP, especially methods for treating pragmatic phenomena — e.g., referring expressions, speech acts, plan recognition and inference, conversational implications, politeness — are required to be cooperatively implemented in a computer program.

REFERENCES


ON COMPUTATIONAL PROCESSING OF RHETORICAL EXPRESSIONS
-PUHS, IRONIES AND TAUTOLOGIES-

Osamu TAKIZAWA*, Masuzo YANAGIDA**, Akira ITO* and Hitoshi ISAHARA*
tekii@crl.go.jp myanagida@doshisha.ac.jp aii@crl.go.jp isahara@crl.go.jp

* Kansai Advanced Research Center, Communications Research Laboratory, Ministry of Posts and Telecommunications
588-2, Iwaoka, Nishi-ku, Kobe 651-24, JAPAN
http://www-karc.crl.go.jp/index.html

** Doshisha University
1-3, Tatara-Miyakodani, Tanabe-cho, Tsuzuki-gun, Kyoto 610-03, JAPAN
http://www.doshisha.ac.jp/index.e.html

ABSTRACT

The computational processing of rhetorical expressions is applicable to many advanced natural language processing systems, e.g., advanced machine translators, intelligent dialog systems, and so forth. Since 1987, the Communications Research Laboratory (CRL) of Japan has performed research on the computational processing of rhetorical, and humorous, expressions. This paper introduces research conducted on the computational processing of rhetorical expressions at CRL. Puns, ironies and tautologies have been our research targets, and this paper mainly introduces our research on computational processing of puns.

1. INTRODUCTION

Computational processing of metaphor is one of the major research themes in the fields of artificial intelligence and natural language processing. As mentioned in the following, our target expressions -- puns, ironies and tautologies, can be treated as variations of metaphor. Therefore, we have paid attention to the three rhetorical expressions.

The general metaphor uses the similarity of the meaning of concepts, whereas all puns use similarity of the pronunciation of concepts [Yanagida88]. Moreover, several types of puns use the relation of the meaning of concept, too. Therefore, pun is, so to speak, more complicated metaphor than general metaphor. In sections 2, 3 and 4, our approach to the computational processing of Japanese puns is introduced.

A function of general metaphor is to describe a concept by highlighting some features of the concept by indicating another concept which has similar features. On the contrary, it is irony to highlight some features of a concept by indicating an opposite concept which has opposite features. Irony is more contextual and situation-dependent than general metaphor. Furthermore, tautology, such as "A boy is a boy", can also be regarded as a type of metaphor. As source and target concepts of a tautology coincide, tautology is also more complicated than general metaphor. In section 5, our studies on computational processing of ironies and tautologies are briefly introduced.

It is indispensable to use many Japanese words and their pronunciations for explanation in this paper. In this paper, Japanese words are notated as shown in the following:

"### I&&&/

Here, "###" denotes a Japanese word written in Japanese characters, and "&&&" denotes a sequence of phonemic symbols. If needed, the English translation of the word is attached just after the notation.

2. OVERVIEW OF PUNS

2.1 Importance of Pun Processing Research
"Pun" is a text or speech fragment in which linguistic ambiguity is used aimed at humorous effect. Research on computational processing of puns is important for the following reasons:

1. Pun process.ng is a main theme of computational processing of humor -- humor is one of the excellent human intelligent functions, and hence research on computational processing of puns is an important issue in artificial intelligence [Binstead94].

2. In most metaphors, "the source concept" and "the target concept" have semantic similarity. In contrast, the source concept and the target concept of a pun generally have phonemic similarity, though, some types of puns can have semantic relations, too. These facts mean that a pun can be considered as a type of complicated metaphor. Research on computational processing of puns can contribute to metaphor processing which is one of the hot issues in Natural Language Processing.

2.2 Component Words in Puns

An interpretation of a pun is based on assuming a single or a set of appropriate words. These words are called "foundation words". Additional meanings come from words with possibly skewed, distorted, or perverse interpretations. The words from which the additional meanings are derived are called "material words". For example, consider the puns "Frankfurt", distorted from "Frankfurt" (a city in Germany), and "Japan Trouble Bureau", distorted from "Japan Travel Bureau". In these examples, "Frankfurt" and "Japan Travel Bureau" are foundation words, and "Krank" (illness) and "trouble" compose a part of material words, respectively [Takizawa89b].

2.3 Overview of Japanese Puns

Puns appear in Japanese language quite frequently in daily conversation in particular. There are mainly the following two reasons why puns appear in Japanese language so often:

1. Japanese language has many homonyms. Besides word homonyms, there can be multiple possibilities of morphological boundaries. For example, "汚職事件 /oshokujikeN/" (a corruption scandal) and "お食事券 /oshokujikeN/" (a meal ticket) have completely the same pronunciation. The former word is decomposed morphologically into "汚職 /oshoku/" (corruption) and "事件 /jikeN/" (scandal), and the latter, into "お /o/" (a prefix for politeness), "食事 /shokujii/" (meal) and "券 /keN/" (ticket).

2. As Japanese Kanji characters are ideograms, even a random sequence of Kanji characters is often suggestive. For example, the pronunciation of a random sequence "強行妻決 /kyookoosaiketsu/" is completely the same as that of the idiom "強行採决 /kyookoosaiketsu/" (ramming a bill). Underlined "妻 /sai/" means "wife", therefore the random sequence suggests "My wife ignores my opinion and decides everything by herself!"

For these characteristics of Japanese language, puns are frequently used as brand names or advertising words and phrases in Japan. For example, "强行革命 /saNgyookakumee/" is a catch phrase of a new electric printer which can print three lines simultaneously. Here, "强者 /saNgyoo/" means "three lines" and "革命 /kakumee/" means "revolution". Its pronunciation is completely the same as "産業革命 /saNgyookakumee/" which means "The Industrial Revolution".

2.4 Classification of Japanese puns

A Japanese pun is precisely defined as follows [Takizawa89a]:

"An expression in which hidden meanings are added onto surface meanings using a word or words whose phoneme sequence is (a) the same, or (b) similar, where the words may (I) overlap, or (II) be separately-located on the phoneme sequence."

An example of type (a) is “勤続疲労
input phoneme sequence and corresponding outputs by DUJAL.

```
input:
/tsuujozaNgyooshoo/
(a sequence of phonemic symbols)

<table>
<thead>
<tr>
<th>DUJAL</th>
</tr>
</thead>
</table>
| output:
| 通商産業省 0.2873 |
| 通商産業庁 0.2789 |
| 通商産業病 0.2740 |
| 通商産業省 0.2667 |
| ... |
| (Possible interpretations represented by Japanese character sequences and their pun-likelihoods) |
```

Fig.1 Processing by DUJAL

DUJAL can be regarded as a "pun detection system" when the input is a sequence of phonemic symbols intended as a pun. DUJAL can be regarded also as a "pun generation system" when the input is a random sequence of phonemic symbols.

The original DUJAL [Takizawa91] however has several faults to be solved. (In the following, the original DUJAL is denoted as "DUJAL-1"). There have been the following two main problems:

1. Semantic functions of a pun as humorous expression have not been well-defined.
2. The calculation method of pun-likelihood has not been properly designed: the magnitude of the degree in particular has been inappropriate.

In order to solve these problems, "DUJAL-2", a revised version of DUJAL-1, has been designed under the following concepts:

1. Humor is often produced from "difference of worth" [Takizawa93]. Based on this hypothesis, DUJAL-2 checks whether the input sequence contains words of "low-worth" or not, and the pun-likelihood is revised to be calculated taking the result into the evaluation process.
2. The new pun-likelihood is kept within the range of 0 to 1.

---

As a matter of fact, "DUJAL" is a pun of "駄洒落 /dajare/" (a cheap joke).
The following section introduces our approach to machine analysis of Japanese puns by describing an overview of DUJAL-2.

3. Overview of the Pun Processing System "DUJAL-2"

3.1 Restriction of the Process Target of DUJAL-2

The present DUJAL-2 is for restricted OLT-puns in which material words express the feature of foundation words. For example, the pun “Norm Securities Co.” consists of a foundation word “Nemura Securities Co.” (the foremost security company in Japan) and a material word “norm” which is semantically related to “Nemura Securities Co.”, as the company is often abused to claim that it compels its employees to work hard imposing strict norms. Another pun “Maneshita Electric Industrial Co.” is phonemically skewed from “Matsushita Electric Industrial Co.”, which is a famous Japanese electrical products firm whose brand “Panasonic” is well known. In Japanese, “maneshita” means “to imitate”. Therefore, the pun is an abuse to claim that the products of Panasonic lack original technology.

However in the real situation, there are some puns whose material words do not directly express the feature of foundation words. For example, in case of the above-mentioned “3行革命/saNgyookakumee”, the foundation word “産業革命/saNgyookakumee” (The Industrial Revolution) and the material word “3行/saNgyoo” (three lines) have no direct relation in meanings. In order to understand this pun, it is indispensable to interpret “3行革命/saNgyookakumee” as an advertising phrase for an electric printer which can print three lines simultaneously, and this fact is related to the word “3行/saNgyoo” (three lines), and such an excellent product is suitable to be called “The Industrial Revolution”. In order to perform such reasoning, complicated processing, such as context dependent understanding employing common knowledge, is indispensable. The recent DUJAL-2 can not deal with such contextualized puns, but can deal only with OLT-puns in which only reasoning of relations between foundation words and material words is required for understanding.

3.2 Design Concept of DUJAL-2

DUJAL-2 is a system to identify both foundation words and material words. As shown in Figure 1, DUJAL-1 has a structure to obtain a set of possible material words. However in real puns, candidates for material words depend on each candidate for foundation words. For example, let us consider the above-mentioned input /maneshitadeNki/. If “松下電器/matsushitadeNki/” (Matsushita Electric Industrial Co.) is assigned as a candidate for foundation words for the input, “/maneshita/” (to imitate), which is often said as an abuse on the company, will be adopted as a candidate for the material words. On the contrary, if “松下冷機/matsushitareki/” (Matsushita Refrigeration Co.), which also has similar pronunciation to the input, is assigned as a candidate for foundation words, and if “/maneshita/” (to imitate) is NOT said as an abuse on the company, “/maneshita/” (to imitate) should not be adopted as a candidate for material words. That is, candidates for material words and each pun-likelihoods are dependent on the candidates adopted for foundation words. Considering this feature, DUJAL-2 is designed to output candidates for material words and their pun-likelihood for each candidate for foundation words.

The target OLT-puns have a surface structure in which the foundation word is long, and some of the material words are overlapping with the foundation word where the material words are correctly pronounced with the sounds similar to the pronunciation for the foundation word. Hence at the beginning stage of processing in DUJAL-2, candidates for foundation words are assigned to the input phoneme sequence based on the longest-match principle. And in parallel with that, candidates for material words are assigned to the input by a best-phonemic-match principle. In the next procedure, pun-likelihoods are calculated for each candidate for material words. The pun-likelihoods are defined as multiplications of the following three degrees: the degree of phonemic-match between the input
Fig. 2 Process flowchart of DUJAL-2
phoneme sequence and the assigned candidates for material words, the degree of mutual relation between the candidates for material words and the foundation word, and the degree of "low-worthiness" of the candidates for material words. Figure 2 shows the process flowchart of DUJAL-2. Notations of "obtained data" in Figure 2 will be explained in Figures 3, 4 and 6.

3.3 Procedure of the Pun Processing System DUJAL-2

In the following, each procedure of DUJAL-2 is explained.

3.3.1 The "Longest-match" Procedure

The longest-match procedure of DUJAL-2 looks for the candidates for foundation words based on the longest-match principle on phonemes. The "degree of longest-match" is defined as in Formula (1).

\[
\text{Degree of longest-match} = (\sum (\Pi p)) \times \frac{2^y}{2^x} \times \frac{1}{2^{z-1}} \quad \cdots (1)
\]

Where, "x" is the number of phonemes in the input phoneme sequence, "y" is the number of phonemes in the interpreted word, and "z" is the number of words in the sequence of interpreted words assigned to the input phoneme sequence. For example, "x" of /tsuujozaNgoyooshoo/ is 15. Here, a contracted consonant, e.g., /ts/, /gy/ and /sh/, is treated as one phoneme, and a long vowel, e.g., /u/ and /oo/, is treated as two phonemes. "y" of /tsuujozaNgoyooshoo/ is also 15. For example of counting "z", if a sequence of three words, "tsuujoa", "2azaNgoyo" and "shoof", are assigned to the input phoneme sequence /tsuujozaNgoyooshoo/, "z" is 3.

"\( \Pi p \)" represents a multiplication of degrees of phonemic similarities, and "p" denotes the degree of phonemic similarity between a phoneme in an input phoneme sequence and a phoneme of the interpreted word. Phoneme pairs of the same phonemes (e.g., phonemic symbol /a/ in input phoneme sequence and sound /a/ in a interpreted word) have the largest "p" value. Meanwhile, phonemes /b/ and /d/ are similar as they belong to the same phoneme group (voiced plosives), thus the "p" value for that phoneme pair is set to be high. On the other hand, the "p" value between phonemes /a/ and /i/ is set to be small, because these phonemes can be clearly distinguished. Table 1 shows actual values for "p" in DUJAL-2. Values in Table 1 are within the range from 0 to 1 and are decided empirically referring to the phoneme identification errors of phoneme identification systems. For example, in the case of the interpreted word "tsuujozaNgoyooshoo/ compared with the input phoneme sequence /tsuujozaNgoyooshoo/, there are two discordant phoneme pairs, i.e., /sh/ & /j/ and /s/ & /z/ whose similarities are both 0.8 as shown in Table 1, and the other phonemes exactly coincide, i.e., "p" values are 1.0. Therefore in this case, \( \Pi p \) is 0.8\(^2\times 1.0^{15} = 0.64. \)

<table>
<thead>
<tr>
<th>phoneme pair</th>
<th>similarity degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>same phonemes</td>
<td>1.0</td>
</tr>
<tr>
<td>/b/ /d/ /g/</td>
<td>0.9</td>
</tr>
<tr>
<td>/v/ /t/ /k/</td>
<td>0.9</td>
</tr>
<tr>
<td>/b/ /p/</td>
<td>0.8</td>
</tr>
<tr>
<td>/d/ /t/</td>
<td>0.8</td>
</tr>
<tr>
<td>/s/ /k/</td>
<td>0.8</td>
</tr>
<tr>
<td>/m/ /n/ /b/</td>
<td>0.8</td>
</tr>
<tr>
<td>/n/ /n/</td>
<td>0.8</td>
</tr>
<tr>
<td>/r/ /m/</td>
<td>0.8</td>
</tr>
<tr>
<td>/u/ /sh/</td>
<td>0.8</td>
</tr>
<tr>
<td>/z/ /j/</td>
<td>0.8</td>
</tr>
<tr>
<td>/s/ /z/</td>
<td>0.8</td>
</tr>
<tr>
<td>/sh/ /j/</td>
<td>0.8</td>
</tr>
<tr>
<td>/t/ /s/</td>
<td>0.8</td>
</tr>
<tr>
<td>/t/ /sh/</td>
<td>0.8</td>
</tr>
<tr>
<td>/a/ /a/</td>
<td>0.7</td>
</tr>
<tr>
<td>/e/ /i/</td>
<td>0.7</td>
</tr>
<tr>
<td>/o/ /o/</td>
<td>0.7</td>
</tr>
</tbody>
</table>

| between           | other vowels      | 0.5               |
| other consonants  | 0.5               |
| other phonemes    | 0.1               |
| insertion/        | deletion          | 0.1               |

Table 1 Values for "p"
The maximum value of the degree of longest-match is \(2 \times 1/2\) for "z" is 2, \(3 \times 1/4\) for "z" is 3, \(4 \times 1/8\) for "z" is 4, and so on. No matter how much "z" increases, the degree of longest-match never exceeds 1. Thus, the range of the degree of longest-match remains in [0, 1].

The degree of longest-match has a feature that the longer the phoneme sequence is common to input sequence and that of the interpreted word, the larger the degree becomes. Moreover, the less number of segmentation of the interpreted words is, the larger the degree becomes. The degree becomes high for the case of roughly consistent in a long sequence of phonemes than the case of tightly consistent in a short sequence of phonemes. Figure 3 shows examples of the interpreted concepts obtained by the longest-match procedure. As shown in Figure 3, a sequence of three words "通常 /tsuuzyour", "残業 /zaNgyoo", and "省 /shoo", whose phonemes are all exactly consistent with the input phoneme sequence, obtains less degree than a roughly consistent single word "通常産業省 /tsuushoosNgyooshoo".

In Figure 3, if the system true up the head of a phoneme sequence and compares a concept "通常産業庁 /chuushokigyooshoo" and an input /tsuuzjoozaNgyooshoo", "similarity degrees" are obtained as follows:

<table>
<thead>
<tr>
<th>Input</th>
<th>Pronunciation of &quot;通常産業庁&quot;</th>
<th>Similarity</th>
<th>Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ts/</td>
<td>/ch/</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>/u/</td>
<td>/u/</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>/u/</td>
<td>/u/</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>/i/</td>
<td>/sh/</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>/o/</td>
<td>/o/</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>/o/</td>
<td>/o/</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>/z/</td>
<td>/k/</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>/a/</td>
<td>/i/</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>/N/</td>
<td>/gy/</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>/gy/</td>
<td>/o/</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>/o/</td>
<td>/o/</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>/o/</td>
<td>/o/</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>/sh/</td>
<td>/o/</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>/sh/</td>
<td>/o/</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>/o/</td>
<td>(insertion error)</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

In contrast, if the phoneme /N/ of the input /tsuuzjoozaNgyooshoo/ is regarded as an insertion error, the following comparison leads to the largest degree as the degree of longest-match.

The system arranges suitable insertion and deletion errors making the degree largest employing DP.

3.3.2 The “Best-phonemic-match” Procedure

The "best-phonemic-match" procedure of DUJAL-2 identifies candidates for material words based on the "best-phonemic-match" principle. This procedure is performed in parallel with, but independently of the longest-match procedure.

The degree obtained by the best-phonemic-match procedure (denoted as "degree of phonemic-match" hereafter) is defined as \( \Pi p \) in Formula (1). Therefore, the length of a consistent phoneme sequence does not effect the calculation process of the degree.

Figure 4 shows the example of candidate words and degrees of phonemic-match obtained by the best-phonemic-match procedure for the input phoneme sequence /tsuuzjoozaNgyooshoo/.

3.3.3 Relation Retrieval Procedure

In DUJAL-1, all semantic relations between candidates for the foundation word and the material word have been treated as an "associative relation". DUJAL-1 has used the association mechanism whose details were described in our previous articles [Takizawa90a, 90b, 90c]. However, there can be other types of semantic relations between concepts as well as
input: /tsuujozaNgoyooshoo/

interpreted words:

\[
W_1 = \text{"通商産業省" (Ministry of International Trade and Industry)} \\
\sigma_W = 0.64 \\
(\sigma_W \text{ comes from } (0.8 \times 0.8) \times (2^{15}/2^{15}) \times (1/2^n))
\]

...

\[
W_2 = \text{"中小企業庁" (Small and Medium Enterprise Agency)} \\
\sigma_W = 0.0064 \\
(\sigma_W \text{ comes from } (0.8 \times 0.8 \times 0.5 \times 0.5 \times 0.1 \times 0.8) \times (2^{15}/2^{15}) \times (1/2^n))
\]

...

\[
W_{41} = \text{"通常" (always)} \\
W_{42} = \text{"残業" (overtime work)} \\
W_{43} = \text{"省" (ministry)} \\
\sigma_W = 0.00104 \\
(\sigma_W \text{ comes from } (1.0 \times (2^6/2^{15}) + 1.0 \times (2^6/2^{15}) + 1.0 \times (2^3/2^{15})) \times (1/2^n))
\]

...

Fig.3 Examples of the interpreted words "W" and the degrees of longest-match "\sigma_W" obtained by the longest-match procedure.

Figure 6 shows a part of the word network around the concept "通商産業省" (Ministry of International Trade and Industry: MITI). In the network, concept labels and relational degrees are assigned to nodes and links, respectively. In the case of concept pairs connected via multiple links in cascade, the relational degree is defined as the multiplication of degrees of the links. For example, in Figure 6, "通商産業省" (MITI) and "郵政省" (Ministry of Posts and Telecommunications; MPT) both have relational degrees of 0.9 with "省" (ministry). Then, the relational degree between MITI and MPT is calculated as 0.9 \times 0.9 = 0.81.

There are various relations in the word network as shown in Figure 5. It is somewhat unreasonable to treat the degree of these various relations in one-dimension. However, in order to compensate this unreasonableness, multi-dimensional degrees should be introduced. In that case, to decide the number and kind of dimen-
sions seems to be more difficult than to tolerate using the one-dimensional degree. Moreover, although the degree is treated as a static value in this network, the degree should be actually treated as a dynamic one, i.e., dependent on context. In the future, relational degrees in the word network will be assumed to be designed as dynamic numbers.

3.3.4 Low-worthiness Detection Procedure

H. Inoue, a famous scriptwriter of Japanese comedies, states that several types of laughter are provoked by a transition from high worth to low worth [Inoue79]. The larger the gap between the two, and the more sudden the transition, the greater the laughter becomes. For example, let us assume a noble gentleman slips on a banana skin, resulting in laughter. This laughter is caused by a sudden transition in the man involved from a noble gentleman (high worth) to a fool (low worth).
A pun is a language expression that provokes laughter. Therefore, the characteristics of laughter seem applicable to puns. Actually, Japanese puns usually express immoral behavior, and material words of Japanese puns often consist of words with "low worth". This fact seems to derive that the "low-worthiness" of material word effects on the appropriateness of the pun [Takizawa93]. Thus, DUJAL-2 is provided with a procedure to check whether candidates for the material words have low-worthiness or not.

In the "low-worthiness detection procedure", the low-worthiness of candidates of material words is detected by comparison with the "database of low-worth concepts". The database of low-worth concepts is a database accumulated by absolute low-worth or undesirable concepts. For example, "ごみ" (wastes), "惨状" (horrible scene) and "失恋" (a broken heart) are registered in the database. As a preliminary system, the database of low-worth concepts only has data of indices of low-worth concepts.

Now, the degree, obtained in the relation retrieval procedure, is then, multiplied by the degree of low-worthiness obtained here. The low-worthiness detection procedure multiplies the degree obtained in the relation retrieval procedure by 1.0 in the case where the candidate for the material word has been registered in the database of low-worth concepts, and by 0.1 in the case where the candidate has not been registered. Here, 0 is not applied to the unregistered concept, because all concepts may have possibility to get low-worthiness according to contexts, and there are puns whose material words have no low-worthiness.

### 3.4 An example of System Performance

Figure 7 shows an example of output results for a Japanese pun /tsuujoozaNgyooshoo/. For this pun, the most plausible candidate for the foundation word is "通商産業省/tsuushoosaNgyooshoo" (MITI) whose degree is 0.64 as shown in Figure 3. According to Figure 7, DUJAL-2 appropriately evaluates the possibility of interpreting the input phoneme sequence yielding large pun-likelihood to "通商産業省" (a pun which means "Ministry whose employees always work overtime"), the intuitively most plausible interpretation as a pun, as a candidate for the material word (underlined in the figure). "通商産業省/tsuushoosaNgyooshoo" and "通

![Fig.6](image-url)  

**Fig.6** A part of the word network around the concept "通商産業省" (MITI). (In the network, concept labels and relational degrees "R" are assigned to nodes and links, respectively.)
candidates for foundation and material word pairs and their pun-likelihoods

| 通商産業省 | 通商残業省 | 0.2873 |
| 通商産業省 | 通商残業庁 | 0.2769 |
| 通商産業省 | 通商残業局 | 0.2740 |
| 通商産業省 | 通商残業病 | 0.2667 |
| 通商産業省 | 通商残業者 | 0.2660 |
| 通商産業省 | 通商残業病 | 0.2657 |
| 通商産業省 | 通商残業商 | 0.2607 |
| 通商産業省 | 通商残業会 | 0.2607 |
| 通商産業省 | 通商残業課 | 0.2500 |
| 通商産業省 | 通商残業務 | 0.2563 |
| 通商産業省 | 通商残業事 | 0.2576 |
| 通商産業省 | 通商残業務 | 0.2573 |
| 通商産業省 | 通商残業病 | 0.2533 |
| 通商産業省 | 通商残業事情 | 0.2527 |
| 通商産業省 | 通商残業事 | 0.2524 |
| 通商産業省 | 通商残業事 | 0.2400 |
| 通商産業省 | 通商残業事 | 0.2400 |
| 通商産業省 | 通商残業事 | 0.2393 |

Fig. 7 Evaluation results of various interpretations for the input /tsuushoozaNgyooshoo/ by DUJAL-2

商業 /tsuushoozaNgyooshoo/, the first and the second candidates for material words in Figure 7 obtain larger pun-likelihoods, because these material words, i.e. “通商 /tsuushoo/” (trade), “残業 /zaNgyoo/” (overtime work), “省 /shoo/” (Ministry) and “庁 /choo/” (Agency) are all strongly related concepts with “通商産業省” (MITI), the candidate for the foundation word. Moreover, “通商残業病 /tsuushoozaNgyoobyoo/”, the third candidate for the material word in Figure 7 includes “通商 /tsuushoo/” (trade) and “残業 /zaNgyoo/” (overtime work) which are both strongly related concepts with “通商産業省” (MITI), and “病 /byoo/” (illness) which is a low-worth concept. These candidates can be interpreted as puns in suitable context, therefore the result of Figure 7 is regarded as appropriate.

3.5 Implementation of DUJAL-2

DUJAL-1 was implemented by ESP (Extended Self-contained Prolog) on MELCOM PSI-2; a prototype machine of the Fifth Generation Computer Project. DUJAL-2 is now under implementation by SICStus Prolog on a SUN SparceStation20.

4. DISCUSSION ON DUJAL-2

The performance of DUJAL-2 strongly depends on the structure of the word network and the database of low-worth concepts. Human intuitive understanding (or generation) of puns surely depends on semantic features of composing words of puns. So, it seems to be appropriate to adopt a mechanism of reasoning using this knowledge. This knowledge is accumulated manually in the present DUJAL-2 system. In the future, we have a plan to construct it semi-automatically and objectively based on dictionaries on market.

Our plans for future works are as follows:

1. Context-dependent puns, as "3行革命 /saNgyookakune/" mentioned above, should be handled.
2. Syntactic relations between material words should be considered. For example, an adverb “通常 /tsuujoo/” (always) and a noun “残業 /zaNgyoo/” (overtime work) are material words of the above-mentioned “通常残業省 /tsuushoozaNgyooshoo/”. These material words are often adjoined syntactically. This kind of syntactic feature should be considered.
3. DUJAL-2 should treat puns whose material words have no low-worthiness.
4. The recent DUJAL-2 was designed only for OLT-puns. However, according to our investigation, about 65% of puns in the Japanese language belong to the SLT-pun group [Takizawa93]. Thus, we have to develop a pun processing system applicable not only to OLT but also to SLT [Takizawa89c]. In the preparation to realize a processing system applicable to SLT-puns, we have recently analyzed the phonemic features of Japanese SLT-puns [Takizawa95a].
5. COMPUTATIONAL PROCESSING OF IRONIES AND TAUTOLOGIES

Besides puns, we have performed the studies of computational processing of ironies and tautologies. In this section, our studies on these rhetorical expressions are briefly introduced.

5.1 Machine Detection of Japanese Ironies

We have performed three research projects related on ironic expressions. They are a proposal to formulate irony based on a dialog model [Ito95], a system to detect Japanese ironies [Takizawa94], and an investigation of the prosodic features of ironies [Yanagida95]. In the following, we briefly introduce the second project.

Recently, several studies on computational models of irony have been performed ([Litman91], [Utsumi95]). However, most of them are only proposals of semantic models of irony, and have not well considered language-dependent features of irony. For example, in Japanese, the terminating particle “な /ne/” is an “ironic marker” which has an important role in deciding whether the expression is irony or not, nevertheless, previous models have not considered such language-dependent features. Such models are insufficient for actual natural language processing systems. Our system to detect ironies is designed with considering such language-dependent features.

Japanese irony, a pair made up of a situation and utterance, generally satisfies the following four conditions, (1) the situation has low “personal estimation value” (denoted as “PEV”), (2) the utterance has high PEV, (3) the situation and the utterance are opposite in meaning to each other, (4) the utterance includes an “ironic marker”. Here, PEV expresses the degree of human desirability. For example, the utterance “He is a high-achieving pupil!” assigns high PEV to the student. Our proposed system can detect ironies which satisfy the four above-mentioned conditions.

Inputs to the system are formalized situations and utterances. Figure 8 shows an example of the input. The format of the input is similar to the case-frame of the NLP system.

The system has four processing steps. Firstly, the system assigns PEV to the inputs by retrieving “the PEV and cause-effect database” (which is denoted by “the PEV database” in the following). Figure 9 shows an entry for the PEV database. Secondly, the system calculates “fulfillment scores” for the above-mentioned conditions for the inputs. Calculation is operated between the situation and the utterance of input by retrieving “pev” fields for each entry on the PEV database. Thirdly, the system detects the relationship between cause and effect in the inputs by retrieving “cause” fields for each entry on the PEV database. And fourthly, the system inspects the terminating particle in the utterance. In Japanese irony, terminating particles play important roles in forming irony. For example, the terminating particle “な /ne/” often appears at the end of ironic utterances. Through the above-mentioned processing steps, the system calculates the “degree of irony”. The degree is treated as output.

The system has been implemented using Prolog and ESP on the MELCOM PSI-II.
5.2 Computational Processing of Rhetorical Tautologies

A rhetorical tautology (denoted as "RT") is a kind of figurative expression, such as "A boy is a boy". There have been many studies on understanding RTs in the fields of psychology or pragmatics ([Wierzbicka87], [Gibbs90]). However, we have not come across any studies on computational models to understand RTs except Utsumi's model [Utsumi96]. Recently, we have selectively investigated a technique for the computational processing of RTs.

We have studied a computational model for understanding RTs [Takizawa95b,96]. Our target RT is a type that emphasizes that a concept belongs to a conceptual group even though the concept is not typical of the group. We assume an RT to be an expression for highlighting some properties of a concept represented contextually by a repeated word. Understanding an RT is defined by the degree of "salience" [Iwayama90], which is a feature that is commonly used in computational models for understanding metaphors. We have proposed a method for calculating the degree of salience of an RT. Our model can express the feature that a concept that is represented contextually by an RT is the factor that decides the emphasized properties. This seems to be similar to the human intuitive understanding of RTs.

Moreover, we have also been constructing a machine detection algorithm to collect real RTs from editorials and columns in Japanese newspapers [Takizawa95c].

6. CONCLUSION

This paper introduced our research on the computational processing of rhetorical expressions, especially puns. In order to advance research on the computational processing of rhetorical expressions, interdisciplinary studies on psychology, linguistics and information science are needed. Our remaining task is to improve the proposed method in terms of breadth and depth, using other disciplines as aids.

Acknowledgments

The authors would like to thank the reviewers and committees of IWCH’96 for their useful comments.

References


[Takizawa89a] O.Takizawa and M.Yanagida: "Detection of Hidden Semantic Structures in


RELEVANCE THEORY AND HUMOROUS INTERPRETATIONS

Carmen Curcó
Department of Phonetics and Linguistics
University College London
Gower Street
London WC1E 6BT
carmen@linguistics.ucl.ac.uk

ABSTRACT

This paper argues that the study of verbal humour is closely linked to the study of how speakers express propositional attitudes implicitly and how hearers retrieve them. In the type of verbal humour on which I concentrate here, speakers lead hearers to entertain mental representations that are attributable to someone other than the speaker at the time of her utterance, while they simultaneously express towards the semantic content of such representations an attitude of disengagement. My analysis accommodates humorous instances of verbal irony as sub-cases of the generalization I propose, and separates in a principled way those which are not. I also show that two crucial aspects in the creation of a humorous effect are the relative mutual manifestness of the participants’ contextual assumptions and the hearer’s ability to recognize the speaker’s intentions and attitudes, and I argue that relevance theory provides a powerful framework for the study of how humorous effects are created and recognized.

1. PRAGMATICS AND THE STUDY OF VERBAL HUMOUR

There are a variety of ways in which information can be conveyed and communication achieved. For instance, as you enter my house you may find me sleeping, cooking, reading, or speaking on the phone. If so, you may perhaps notice the colour of the clothes I am wearing, the kind of dish I am preparing, the title of the book I am reading, to whom I am talking on the phone, or the mood in which I am. It is conceivable that you may get all this information without me intending you to do so. In cases like these, I will be providing you with direct evidence of the information conveyed. By contrast, when an utterance is aimed at an audience, hearers are provided with direct evidence of the speaker’s intention to convey certain information. This last form of communication whereby a communicator provides direct evidence of his intention to communicate a piece of information is generally called ostensive communication, and verbal communication is a particular instance of it.

Intentional verbal humour is a result of ostensive communication. It arises from the processing of utterances, which are a kind of ostensive stimuli. It would therefore seem natural for a pragmatic theory of verbal humour to be able to predict what kind of utterances and texts will be humorous and why. Before addressing this question, two crucial aspects of utterance production and processing must be stressed.
First, the thoughts that a speaker communicates through his utterances are not fully encoded in the linguistic expressions he uses. In the second place, given that what is linguistically encoded by utterances underdetermines the speaker's intended meaning, hearers necessarily engage in some inferential process that leads them to recover the message the speaker intends to convey. Utterances are thus used to provide evidence for the hearer of the speaker's intentions.

In this paper I take the view that pragmatics is the study of how the type of inference that leads a hearer to retrieve the speaker's intended meaning takes place. My main purpose is to explore the extent to which, when humour occurs, there is something distinctive in the inferential process that leads to the recovery of the speaker's intended meaning which is characteristic of the creation and recognition of humorous effects.

Relevance theory, the approach to pragmatics which I adopt here, rests on the following four main premises (Sperber and Wilson 1986, 1995):

a) Utterances have several possible interpretations, all consistent with the information they linguistically encode.

b) Not all these interpretations are equally accessible to the hearer at the time of utterance.

c) Hearers have at their disposal a general criterion for choosing amongst all the possible interpretations of an utterance as they occur to them, and they use it consistently.

d) This criterion allows for only one interpretation to be accepted as the one intended by the speaker.

1.1 RELEVANCE THEORY

The pragmatic theory developed within the relevance-theoretic account of cognitive processing is unlike other approaches in that it does not assume the existence of communication maxims on which speakers and hearers rely. Rather, the theory suggests the existence of a single criterion which cannot be violated and which selects among possible interpretations. This criterion emerges from the following very general assumption about human cognition: we pay attention only to information that seems relevant to us.

Relevance is a technical notion defined in terms of the cognitive gains obtained from the processing of a given piece of information and the processing effort invested in deriving such cognitive gains. The greater the cognitive effects derived, the greater the relevance, the greater the processing effort invested in it, the lower the relevance.

There are three main ways in which new information can achieve cognitive effects in a given context:

a) by strengthening an existing assumption in that context,
b) by contradicting and weakening or eliminating an existing assumption,
c) by combining with existing assumptions to create new assumptions, called contextual implications.

Sperber and Wilson propose that humans automatically tend to balance effort and effects when processing information. Speakers know at least tacitly that their audience will allocate their cognitive resources to process the utterances they produce in a relevance-oriented manner. Audiences, on the other hand, will unconsciously expect any ostensive stimulus, such as an utterance, to be the most relevant one compatible with the communicator's abilities and preferences (Sperber and Wilson 1995). To use Sperber and Wilson's analogy, just as offers of food create expectations of edibility, offers of information create expectations of
relevance. Our general tendency to balance effect and effort is thus exploited in human verbal communication. Hearsers assume that in order to arrive at the interpretation intended by the speaker they will have to go to no unjustifiable effort. Additional effort should be offset by additional cognitive effects. So, irrespective of whether a speaker and a hearer are aware of it, every utterance communicates the presumption of its own optimal relevance. This is the central claim of relevance theory, known as the Principle of relevance, which applies not only to utterances, but to every ostensive stimulus:

*Principle of relevance:* Every act of ostensive communication communicates the presumption of its own optimal relevance. (Sperber and Wilson 1986: 158)

In order to be acceptable and comprehensible, an utterance does not have to be optimally relevant. The fact that some speakers are more competent and benevolent than others and they can still be understood shows this. All that is required is that the hearer should be able to see how the speaker might have intended it to be so. The pragmatic criterion that relevance theory suggests guides interpretation is the criterion of consistency with the principle of relevance.

*Criterion of consistency with the principle of relevance:* An utterance, on a given interpretation, is consistent with the principle of relevance if and only if the speaker might reasonably have expected it to be (or to seem) optimally relevant to the hearer on that interpretation.

The first interpretation tested by the hearer which meets this criterion is the one he selects as the interpretation intended by the speaker, in other words, the criterion enables hearers to exclude all other possible interpretations.

### 1.2 RELEVANCE THEORY: SOME

**BASIC DISTINCTIONS FOR THE STUDY OF VERBAL HUMOUR**

I have now sketched the main ideas underlying relevance theory. Before turning to the analysis of verbal humour within this framework, I will concentrate briefly on the aspects and distinctions of the theory that I find most central to the study of verbal humour.

*Cognitive environments and mutual manifestness*

An act of communication gives rise to shared information. Besides, if communication is to be achieved, some sharing of information is necessary beforehand. Any general theory of language use must address the issue of what one's interlocutor knows and the role this knowledge plays in interpretation.

In relevance theory the role of shared information in communication is accounted for by the notion of *mutual manifestness* (Sperber and Wilson 1986, 1995). For a fact or an assumption to be *manifest* to an individual it needs to be perceptible or inferable. Of course, as assumptions and facts can be more or less salient perceptually, and more or less difficult to infer, they differ in their degree of manifestness.

Sperber and Wilson's notion of what is manifest to an individual is clearly weaker than the notion of what the individual actually knows or assumes. For instance, there is a sense in which you know that Margaret Thatcher never went shopping with Cleopatra, although this very thought had probably never been entertained by you before reading these lines. In this weaker sense, something can be known without ever having been entertained. In the same way something can be manifest without being known or assumed. It need not be one of your
assumptions that Cleopatra was never jealous of Marilyn Monroe (or of Margaret Thatcher for that matter), but it is something you are capable of inferring from what you know and assume. Although the assumption that Cleopatra was never jealous of Marilyn Monroe is not part of your conscious and active knowledge, nor of what you assume, it is manifest to you because it can be inferred from what you know and assume.

Moreover, as mentioned above, something can be manifest simply by being perceptible. Imagine that you are so absorbed by a conversation you are having that you don’t notice the Concord flying over you. Although you have no knowledge or assumptions about it, because the fact that the Concord is flying over your head now is perceptible by you, it is manifest to you.

The set of facts and assumptions that are manifest to an individual at a given time constitutes his cognitive environment. Because individuals share physical surroundings and cognitive abilities, their cognitive environments intersect to some extent. Although we never share our total cognitive environments, we do share them to some extent.

When two individuals share a cognitive environment they do not necessarily make the same assumptions. Having a shared cognitive environment only means that the people who share it are capable of making the same assumptions, not that they actually do.

There is one further assumption about a cognitive environment that can be manifest to individuals: an assumption about the type of people who have access to it. If you and I are in the same room and no one else is or has ever been there, I know that we and only we have access to this cognitive environment. If I belong to a secret sect, I know that I share with the members of this sect a cognitive environment not shared by non-members of the sect. A mutual cognitive environment is any shared cognitive environment in which it is manifest to the individuals who share it which people share it. So, in a mutual cognitive environment, for each manifest assumption the fact that it is manifest to the people who share this environment is itself manifest. Another way of referring to this is to say that in a mutual cognitive environment every assumption is mutually manifest. I will return to this notion in my analysis of humorous interpretations and argue that it plays a crucial part in a theory of verbal humour.

**Explicit and implicit meaning**

Verbal communication comprises two processes: one based on coding and decoding and the other on ostension and inference.

An utterance has a logical form yielded by the automatic decoding of the language faculty which is inferentially completed by the hearer. This logical form is a well-formed formula, a structured set of constituents which can undergo formal logical operations determined by its structure (Sperber and Wilson 1986: 72). The hearer’s initial task upon hearing an utterance is to flesh out its logical form into a propositional form. This completion process involves doing three basic things: assigning reference, disambiguating and eliminating vagueness, and adjusting certain concepts encoded linguistically in the utterance. The result is the propositional form of the utterance, which in some cases, but not always, is one of the assumptions the speaker wanted to communicate. In such circumstances, Sperber and Wilson call the proposition expressed by the utterance the explication of the utterance.

However, the development of a logical form need not stop at this stage. If p is the proposition expressed by the utterance, it is possible for the hearer to
embed p in a higher-level description of the speaker's attitude towards the proposition he has expressed, such as

The speaker believes that p
The speaker regrets that p

Whenever such higher-level descriptions are part of what is overtly communicated, they are called higher level explicatures of the utterance.

So, any development of the logical form of an utterance which produces an assumption that the speaker overtly intended to make manifest by means of his utterance is an explicature of the utterance.

Assumptions which are communicated by the utterance but which are not a development of its logical form are implicatures. Assumptions can be implicitly communicated in two ways. A hearer may need to provide a contextual assumption in order to satisfy himself that the speaker's utterance is consistent with the principle of relevance. So, for instance, in the following exchange B needs to supply the assumption that Tommy is a musical.

(1) A: Have you seen Tommy?
   B: I don't go to musicals

Sperber and Wilson (1986, 1995) distinguish between assumptions conveyed implicitly in this way, which they call implicated premises, and assumptions that the hearer derives deductively, such as the assumption that B has not seen Tommy. These are accordingly called implicated conclusions. In later sections I will show that implicated premises play a key role in verbal humour, as a typical feature in the creation of a humorous effect is to lead a hearer to supply an implicated premise that contradicts a weakly mutually manifest contextual assumption.

Interpretations and descriptions
Thoughts are not fully encoded in utterances. The latter are merely public representations whose contents resemble that of the former. The fact that utterances can be used to convey thoughts which they do not reproduce exactly is a particular case of our more general ability to represent in virtue of resemblance.

We are constantly using objects in the world to represent other things they resemble. Although resemblance is a very vague notion, when the representations and objects involved in a relationship of resemblance have propositional content, it is possible to define a specific type of resemblance on the basis of how close such propositional contents are. So, the resemblance between a thought and the utterance that is used to represent it can be defined in terms of the amount of logical and contextual implications they share in a given context. The more shared implications they have, the greater their resemblance. Sperber and Wilson (1986) call this particular kind of resemblance interpretive resemblance and a representation that stands for some other in virtue of the interpretive resemblance between the two is called an interpretation.

Hence, two representations with a propositional content interpretively resemble one another in one context to the extent that they share logical and contextual implications in that context. Because of this, Sperber and Wilson argue that every utterance is an interpretation of a thought of the speaker's. In this framework, literal interpretations are simply a limiting case of interpretation by resemblance: those where the utterance and the thought is represents share all their logical and contextual implications in all contexts.

To sum up, there are two ways in which a representation with a propositional form, in particular thoughts and utterances, can be used to represent things. As I have just discussed, a representation can be used to represent some other representation with
a propositional form in virtue of a resemblance between their propositional forms. In this case, the representation is an interpretation. But also a representation can be used to represent some state of affairs in virtue of its propositional form being true of that state of affairs. If so, the representation is a description.

Every utterance then is an interpretation of a thought of the speaker's. The thought itself may be a description of a state of affairs or a representation of some further thought or utterance. In the first case, we will talk of descriptive use of language, in the second, of interpretive use. Sperber and Wilson (1986, 1995) have shown how this distinction allows for a characterisation of tropes and illocutionary force.

Propositional attitudes
I mentioned above that the development of a logical form is not confined to the recovery of the proposition expressed by an utterance, and that there are other propositions that the hearer can obtain by embedding the propositional content of the utterance into a higher-level description of the speaker's attitude towards the proposition he has expressed. Indeed, in some cases the main relevance of an utterance can be said to lie more in the speaker's attitude towards the proposition expressed than in the proposition itself, a point to which I will return later in my analysis of some humorous utterances and irony.

The fact that a speaker wishes to convey a particular attitude towards the proposition he has expressed may or may not be evident from the linguistic form of his utterance. There are linguistic and para-linguistic clues that may signal the attitude the speaker wishes to convey. Mood indicators, sentential adverbs, parentheticals and performative verbs are explicit markers of the speaker's attitude towards the proposition expressed. The speaker's tone of voice or the look on his face, for instance, may function as non-linguistic indicators of his attitude towards the proposition expressed.

As my treatment of humour will show below, a speaker can implicitly express his attitude not only towards the proposition expressed by his utterance, but also towards the assumptions it communicates implicitly. I will argue that humour exploits the entertainment of contradictory propositional forms as a mechanism that cues the hearer to the recovery of an implicitly expressed dissociative attitude.

Echoic use
There is a sub-type of representation by resemblance that is fundamental to the creation of humorous effects. It is possible for a speaker to use a representation that he attributes to someone other than himself at the time of his utterance, and simultaneously express one of a range of possible attitudes to it. In relevance theoretic terms, when a speaker uses a representation in this way, he uses it echoically.

In cases of echoic use, an utterance achieves relevance by informing the hearer that the speaker has in mind some representation which he is attributing to someone, and also that he has a certain attitude to it. The attributed representation may be an utterance (actual or potential) or a thought, and in fact, any representation, not necessarily one with a propositional form. The attitude expressed may vary from complete endorsement to total dissociation. Consider for instance the exchange below, where Peter echoes Mary's utterance with an attitude of incredulity.

Mary: Yes
Peter: Yes?

Echoic use is rather pervasive in
language, and this technical notion has successfully been put to use in accounting for phenomena as diverse as metalinguistic negation and irony.

It is possible for speakers to echo full representations, as in (2-3), or part of them, as in (4-5):

(2)  A: What did she say?
     B: (in anger) "I know nothing about it"

(3)  A: We are going
     B: (with delight) We are going!

(4)  We didn't see the HIPPOPOTAMUSES, we saw the hippopotami.

(5)  Around here we don't eat tom(æt uz), we eat tom(eiD uz)

Examples (4) and (5) are from Carston and Noh (1995), who have shown that it is possible to echo segments of a representation (e.g. part of the material falling within the scope of a negation operator), which, they have argued, is typical of instances of metalinguistic negation

The target of an echo may vary across a wide range of properties. As Carston and Noh mention (1995: 4) it can be one of a range of linguistic properties, for instance, phonetic, grammatical or lexical properties, aspects of dialect, register or style, as well as a number of paralinguistic features such as the tone of voice, pitch, and gestures. One can also have as a target the semantic content of a representation.

Also, utterances or thoughts can be echoed either implicitly, as in (6), where the range of attitudes that can be expressed varies from full endorsement to complete dissociation, or explicitly, as in (7).

(6)  The ban on British beef should be lifted

(7)  He said so cynically "the ban on British beef should be lifted"

Echoes can then be explicit or implicit, a diversity of fragments or of complete mental representations can be used echoically, a wide variety of attitudes can be expressed through such use and a number of linguistic and paralinguistic properties can be the targets of the echo.

In the next section I suggest that an important element in the process that leads to at least some humorous interpretations is the recognition of the speaker's implicit expression of his attitude of disengagement from an attributed implicit assumption. This assumption, made strongly mutually manifest by the utterance, functions as the target of the echo.

Main relevance of an utterance

An important characteristic of the contextual effects yielded by an ostensive stimulus is that they are not derivable either from the context or from the new information alone. Contextual effects result from an interaction between old assumptions and incoming information. So, a set of assumptions P contextually implies an assumption q in the context C if and only if the union of P and C non-trivially implies q, P does not non-trivially imply q and C does not non-trivially imply q either.

Suppose that an utterance \( u \), relevant in a context C, makes manifest a set of assumptions P. It is conceivable that not all the assumptions in P will contribute equally to the overall relevance of \( u \). For instance, a hearer may derive more contextual effects from the fact that a speaker has expressed his attitude towards the proposition expressed by his utterance than from the proposition expressed itself. Imagine that in uttering (8) a speaker also intends to communicate the higher level descriptions in (9)

(8)  I must go to Paris

(9)  (a) I believe that I must go to Paris
     (b) I regret that I must go to Paris
In certain contexts, the main relevance of the utterance may lie more in the speaker's attitude towards the proposition expressed than in the proposition expressed itself. For example, in a context where both the speaker and hearer know that the speaker must go to Paris, and he repeatedly utters (8) with a particular tone of voice, the main relevance of his utterance may lie in the higher level implicature in (9b).

Equally, the main relevance of an utterance may lie in one of its implicatures. So, consider the following exchange:

Peter: Will Piet be willing to spend all this on us?
Mary: He’s Dutch

Here, Mary’s utterance activates a set of stereotypical assumptions which include the assumption that the Dutch are mean. This implicated premise is used to derive the implicated conclusion that Piet will not be that happy to spend much money on us. The implicated premise contributes to the overall relevance of the utterance by giving immediate access to a context that will allow the derivation of adequate effects, thereby helping to reduce the effort needed to derive them. The implicated conclusion is responsible for triggering a number of further contextual effects. Assuming that the hearer already knows that Piet is Dutch, the main relevance of the utterance lies less in the proposition expressed than in its implicatures.

Whenever most contextual effects of an utterance are derived from some specific (segments of) the representations to which it gives access, we can say that the main relevance of the utterance lies in them. This means that if a hearer fails to gain access to such (segments cf) representations, his interpretation of the utterance will not be optimally relevant in the way intended by the speaker.

2. THE INTERPRETATION OF HUMOROUS EFFECTS

In the first section I mentioned that a pragmatic theory of humorous effects might be expected to predict which utterances are humorous and to explain why. My view however is that humorous utterances and texts are so only in a weak and indirect sense, and that what produces a humorous effect is rather the type of inferential pattern that their processing induces. Hence, rather than searching for textual constraints, I concentrate here on isolating inferential patterns that may turn out to be characteristic of humour. A humorous text or a humorous utterance will then be one whose processing in a particular context gives rise to such characteristic inferential patterns.

2.1 CHALLENGING MANIFEST ASSUMPTIONS: THE IMPLICIT EXPRESSION OF ATTITUDES

There is a variety of humour that makes use of the perception of the incongruous to cue the hearer into the recognition of the speaker’s implicitly expressed propositional attitude.

What is characteristic of such cases, which I illustrate in (10-13) below, is that in interpreting the utterance as consistent with the principle of relevance hearers are led to entertain two conflicting assumptions. One of these two assumptions is a strongly implicated premise that clashes with an accessible assumption in the current context of interpretation. For ease of reference, I will call such an implicated premise the key assumption. The assumption that gets contradicted is a so far weakly mutually manifest assumption in the context of interpretation which, as a result of the contradiction, becomes suddenly strongly mutually manifest. I will call this
assumption the target assumption. Let us consider the examples.

(10) Don't keep telling the lady you are 
    unworthy of her. Let it be a 
    complete surprise.

(Henry 1966, cited in Dolitsky 1992)

In the context of interpretation, the first part of the utterance makes (10.1) weakly mutually manifest (or at least the desire of the speaker to believe that (10.1)).

(10.1) (target assumption): The addressee 
    is worthy of the lady in question.

However, in order to assign pronoun 
    reference in accordance with the principle 
    of relevance, the hearer needs to supply 
(7.2) as an implicated premise.

(10.2) (key assumption): The addressee is 
    not worthy of the lady in question.

Consider (11):

(11) There is something tragic about the 
    enormous number of young men 
    there are in England at the present 
    moment who start life with a perfect 
    profile and end up by adopting 
    some useful profession

(Oscar Wilde)

As a social value, (11.1) is mutually 
    manifest in the context of interpretation.

(11.1) (target assumption): A useful 
    profession is a cause for 
    congratulation

But in order to interpret the second 
    part of the utterance as consistent with the 
    principle of relevance, it is the adoption of 
    a useful profession that must be taken as 
    the source of the tragic outcome, therefore, 
    the hearer is strongly encouraged to supply 
(11.2) as an implicated premise.

(11.2) (key assumption): A useful 
    profession is a cause for 
    commiseration

The mechanism is very similar for 
(12) and (13) below.

(12) There is no question that there is an 
    unseen world. The problem is how 
    far it is from midtown and how late 
    it is open.

(Woody Allen)

(12.1) (target assumption): An unseen 
    world has no physically determined 
    spatial-temporal location

(12.2) (key assumption): An unseen world 
    has some physically determined 
    spatial-temporal location

(13) It is perfectly monstrous the way 
    people go about nowadays saying 
    things against one behind one's 
    back that are absolutely and entirely 
    true.

(Oscar Wilde)

(13.1) (target assumption): The things 
    people tell against one behind one's 
    back are false.

(13.2) (key assumption): The things people 
    tell against one behind one's back 
    are not false.

My suggestion is that in all cases, 
    the speaker is implicitly expressing his 
    attitude of disengagement from the target 
    assumption. As an attitude, disengagement 
    is a variety of dissociation. It is possible to 
    dissociate oneself from a propositional 
    representation for a number of different 
    reasons, and with different accompanying 
    emotional overtones: anger, bitterness, 
    surprise, reproach, incredulity, mockery, 
    scorn, regret, etc. In cases of 
    disengagement, emotional overtones are 
    either not present, or not relevant. In cases 
    of humour, perhaps only mockery remains.

In all the examples above, before the 
    key assumption is recovered, the target 
    assumption is only weakly manifest in the 
    context of interpretation. When the key 
    assumption is supplied, the relative strength 
    in manifestness of the target assumption 
    undergoes a sudden and drastic change. 
    This is a crucial feature in the generation of 
    a humorous effect.

Target assumptions are usually 
    instances of downright taken for granted
assumptions, not relevant enough to be worth foregrounding. Besides, the target assumption typically represents an attributable thought. In (10), the thought represented by the target assumption (or at least the desire to believe it) is attributable to the hearer, while in (11) the thought is attributable to a group of people. Oscar Wilde is known for his incisive and scathing attack on Victorian values, which are echoed through his utterance.

Let us consider the Woody Allen case. The target assumption that the supernatural has no physically determined spatial-temporal location is an assumption shared by most human beings, moreover, one that would be difficult not to endorse. What can the point of implicitly challenging it be? What we have here is a case of recursive representation by resemblance. The target assumption is an interpretation of a complex thought, which we can simplify as the view that the supernatural and all we relate to it is of a radically different nature than the concrete mundane day-to-day. The speaker is strongly leading the hearer to entertain the clashing implicated assumption that the supernatural can be reduced to specific and material obsessions and doubts. Given that the two contradictory assumptions are now strongly mutually manifest to both speaker and hearer, the latter has to conclude that no rational communicator would simultaneously use both of them descriptively. This implicated contradictory assumption is put forward because it interpretively resembles the world view and system of beliefs of a specific type of people. By strongly implicating it, the speaker represents by resemblance holders of this particular world view. The speaker thus depicts himself as someone obsessively and neurotically concerned with properties, values and doubts of urban middle-class life. Only such people could conceivably question the diverging nature of mystic and temporal issues. It is this type of people - represented by resemblance by the comedian- who implicitly express an attitude of dissociation (not disengagement) from the target assumption. The disengagement emerges in us from our recognition of this relation of representation by resemblance between the actual speaker -the comedian- and the potential speakers -the targets of the mockery. The case of Woody Allen is besides a case of self-parody. Not only is he representing the neurotically obsessed, educated urban middle-class, but one can recognize in the comedian, the caricature of the New York Jew Woody Allen himself is. The possibility of caricaturing emerges as a consequence of the recursive interpretive use, as when representations are interpretively used what matters is not truth, but resemblance to other representations.

It is worth stressing that my purpose here is not to give a full account of the factors that are responsible for the experience of humour, but only to isolate a characteristic pattern of interpretation that may underlie it.

2.2 PRAGMATIC MECHANISMS OF HUMOUR: GENERATING THE OVERALL EFFECT

The mechanism that underlies the cases I have discussed can be generalized as follows:

a) An utterance is produced that makes strongly mutually manifest a key assumption. The key assumption is in overt contradiction with the target assumption, so far only very weakly mutually manifest.

b) The overt contradiction suddenly makes the target assumption strongly mutually manifest.

c) The entertainment of an incongruity produced by an ostensive stimulus signals
to the hearer the possibility that the speaker is implicitly expressing an attitude of dissociation towards one of the assumptions.

d) Because the key assumption is either strongly implicated or explicated, the expression of attitude is attached to the target assumption.

e) The dissociative attitude is extended to one of disengagement. There are presumably a variety of indicators that may cue the hearer in this direction, such as tone of voice, gestures, and the recognition of multiple and recursive interpretive use.

It is conceivable that this analysis may be extended to cover other types of humour, but I won’t pursue this here. In the next section I want to discuss how this mechanism subsumes irony.

2.3 VERBAL IRONY AND HUMOUR

Sperber and Wilson have rejected the view that irony is the figure of speech whereby a speaker says something and means or implicates the opposite. Instead, they have analysed irony as a particular case of implicit echoic use where the attitude expressed is one of dissociation (Wilson and Sperber 1992). The claim of relevance theory then is that verbal irony has three characteristic features (Wilson 1995):

(a) it is a variety of interpretive use in which the proposition expressed by the utterance (my emphasis) represents a belief implicitly attributed by the speaker to someone else or to herself at another time;
(b) it is echoic (i.e. it implicitly expresses the attitude to the beliefs being represented); and
(c) the attitude involved in one of dissociation from the opinions echoed.

It is important to note that echoes do not necessarily involve literal repetition of an attributed utterance. Often irony involves an element of exaggeration of the opinions being echoed. Also, what is echoed need not be what someone has said, but something they have assumed or implicated, or the thoughts and expectations of a group. Hence, the notion of echoic utterance is not restricted to the echoing of actual words.

One type of irony that has not been much discussed concerns cases where the speaker means what she says, (i.e. she endorses the content of the proposition expressed by her utterance), while still being ironic. Consider for instance (15) uttered in a context where the speaker is trying to say something to the hearer and the hearer’s attention is clearly somewhere else:

(15) I love it when you pay attention to me.

The utterance is ironic; however, it just does not seem to be the case that the speaker is implicitly dissociating herself from the belief represented by the proposition expressed by her utterance. In (15) the proposition expressed represents a belief the speaker endorses. Because of this, some have argued that similar cases cannot be accounted for by the echoic view "because the speaker does not dissociate herself from the opinion echoed" (Giora 1995: 247).

I don’t think that this type of example (also considered by Gibbs & O’Brien 1991 and Curcó 1995) proves the echoic view of irony wrong, but such cases certainly show that it needs some reformulation so that its claims are made clearer and more specific.

Imagine that it is a habit of the hearer to get distracted when the speaker addresses him, or to dismiss her opinions. Suppose as well that, to the speaker’s surprise, on this specific occasion the hearer has been particularly attentive to what the speaker had to say and has even replied ‘I think you’re right. Let’s do as you suggest’ In such a scenario, a speaker uttering (15)
endorses the proposition expressed by her utterance and also all the strong implicatures to which it gives rise, for instance, the implicated conclusion that she is now pleased that the hearer has paid attention to her, represented by (16).

(16) The speaker is now pleased by the attention the hearer has given her.

Normally, (16) would be derived from the utterance and the contextual assumption in (17).

(17) The hearer is paying attention to what the speaker is saying.

Imagine however a different scenario: one where the speaker is trying to get a point across and the hearer is overtly concentrated on something else. She utters (15). In this case, the utterance does not carry the implicature that she is now pleased that the hearer has paid attention to her. For one thing, (17) is not manifest in the context of interpretation.

The difference between situations where (15) is ironic and situations where it is not hinges on two related factors: a) whether the speaker dissociates herself not from the content of the proposition expressed by the utterance, but from the implicatures to which the utterance would normally give rise, and b) the kind of contextual assumptions that are at stake in interpreting the utterance.

The view of irony where the target of the echo is taken to be the opinion represented by the proposition expressed by the utterance is too narrow. But this is not what the first requirement on irony necessarily demands. The first requirement of irony is not that the target of the echo (that from which the speaker dissociates) should be the proposition expressed, but only that this proposition should interpretively resemble a thought of someone other than the speaker at the time of utterance. From this it does not follow that the target of the implicitly expressed dissociation characteristic of irony should be the content of the proposition expressed itself. As I will show, an ironic speaker who uses an utterance interpretively can be expressing an attitude of dissociation to any of the range of the assumptions that the utterance makes strongly mutually manifest. Whether a hearer will search for the implicit expression of such an attitude depends crucially on the type of assumptions that are mutually manifest to the speaker and the audience. In particular, whether he finds some contradiction like the ones I have described above.

In cases where (15) is not ironic, its main relevance lies in the proposition it expresses, given that adequate contextual effects can be derived from merely combining it with the context. When (15) is ironic, it does not carry the implicature that the speaker is now pleased that the hearer has paid attention to her. For one thing, (17), from which it is normally derived, is not mutually manifest in the context of interpretation. Not only this, but the context is not, as it were, neutral in this respect. The negation of (17), represented by (18), is (maybe only weakly) mutually manifest.

(18) The hearer is not paying attention to what the speaker is saying.

In producing (15), the speaker increases suddenly the relative mutual manifestness of (18). What an ironic speaker would be dissociating from in ironic cases is the assumption in (17) and, consequently, the otherwise potential implicature in (16).

It is worth remarking that without the retrieval of the dissociative attitude, the the hearer will not recover the optimally relevant interpretation intended by the speaker. But what indicators of the expression of such attitude are available to the hearer?

Speakers addressing an audience normally expect to receive attention. That is, (17) should standardly be mutually
manifest, if only weakly, as it is a generally taken for granted assumption. (17) functions as the equivalent of the target assumption in the humorous cases. By uttering (15), the speaker manages to increase the mutual manifestness of (18), which works as the equivalent of the key assumption in examples (10-13) above. The mutual manifestness of (18) before (15) is uttered is most likely not strong. A great deal of the relevance of the utterance lies in its suddenly increasing the relative mutual manifestness of (18) and communicating that it is relevant in its own right. Consequently, the retrieval of the attitude of dissociation towards both (17) and the associated implicate in (16) is made possible.

The main relevance of the utterance lies therefore in two aspects: the sudden increase in the manifestness of a contextual assumption, and the implicit expression of a dissociative attitude. The first feature provides a clue to retrieve the second.

So, we need to modify slightly the echoic view to acknowledge that the opinions echoed by the utterance need not be those represented by the content of the proposition it expresses. In fact, one of my main points in this paper is to argue that the target of an echoic utterance does not always lie within the content or the form of the proposition it expresses. I am arguing that an echoic utterance can have as its target the semantic content of one or more of its implicatures, and even the semantic content of an assumption that the utterance makes suddenly strongly mutually manifest, as the humorous cases I discussed above reveal.

It also seems that the issue of the degree of manifestness of contextual assumptions needs to be addressed in more detail, as it clearly plays a key role in interpretation, in particular, in the recovery of the implicitly expressed dissociative attitude.

I believe that the source of the main relevance of an utterance, as well as the sudden changes in the relative mutual manifestness of contextual assumptions of the kind I have discussed, are central features of verbal humour and wit. I will now try to show that they can explain some other apparent problems that the echoic approach to irony faces.

Giora (1995) has suggested that the requirement that an echoic utterance be accompanied by ridicule is not sufficient for it to be ironic. Her example is reproduced in (19) below:

(19) Dina: I missed the last news broadcast. What did the Prime Minister say about the Palestinians?
Mira: (with ridiculing aversion) That we should deport them.

She argues that Mira’s utterance in (19) is echoic in that it simultaneously reports a content and expresses the speaker attitude to what is reported, but nonetheless, it doesn’t come across as ironic. Her point is that the difference between (19) and (20) is that (20) conforms to her proposed marked informativeness condition of well-formed irony, while (19) does not.

(20) Dina: I missed the last news broadcast. What did the Prime Minister say about the Palestinians?
Mira: That we should host them in 5 star hotels in Lebanon.

In standard circumstances, a strongly manifest assumption in the context of interpretation of both (19) and (20) is that the comfort of Palestinians will not be one of the Israeli Prime Minister’s main concerns.

Adequate contextual effects can be derived from combining the proposition expressed by Mira’s utterance in (19) with current contextual assumptions, amongst which is (21).

(21) The comfort of Palestinians is not one of the Israeli Prime Minister’s
main concerns.

While the expression of the dissociative attitude of course adds to the overall relevance of the utterance by producing extra effects, its recovery is not essential for the utterance to achieve optimal relevance. In (19) Dina asks a question to which Mira’s utterance replies with adequate relevance. Notice that no mutually manifest assumptions in the context of interpretation are strongly challenged by the production of (19).

The interpretation of (20), by contrast, calls for an implicated premise that the comfort of Palestinians is a priority of the Israeli Prime Minister, thus contradicting (21). A naive hearer who does not hold (21) with enough strength might take its negation on board and miss the irony. Because of the strength with which (21) will probably be held by Dina, she will need to look for contextual effects that do not derive directly from the content of the proposition expressed by Mira’s utterance, given that it does not provide her with a relevant answer to her question. As the contradiction has arisen from an ostensive stimulus, and (21) is probably mutually manifest, the hearer of (20) will continue to search for contextual effects, and in particular, from contextual effects that can be derived from the encountered contradiction. A dissociative attitude implicitly expressed yields adequate contextual effects. What is crucial is that missing the dissociative attitude in interpreting (20) drastically detracts from an optimally relevant interpretation (if (21) is mutually manifest, though not necessarily otherwise). This contradiction of a mutually manifest assumption needs not be entertained in interpreting (19). The contradiction in (20) makes the possibility that the proposition expressed by the utterance does not correspond to the contents of the Prime Minister’s statement manifest to the hearer. This case matches the humorous examples discussed above in that the interpretation of (20) forces the introduction of a premise that contradicts a manifest assumption in the context of interpretation.

So, the main relevance of (19) lies in the propositional content it expresses, as it answers relevantly the question posed by the speaker. The expression of a dissociative attitude is, as it were, secondary. Most contextual effects are derived from the report of the content represented by the proposition expressed alone. The main relevance of (20), on the other hand, lies precisely in the expression of the dissociative attitude, as it should be clear to the hearer with the relevant knowledge of the world that the propositional content of the utterance will resemble very little that of the thought attributed to the Prime Minister.

Relevance, mutual manifestess and incongruity

Let me now consider briefly the relation of my analysis to incongruity theories of humour.

As Forabosco has suggested (1992: 54) a formal category of the incongruous should be able to capture instances of deviation from existing cognitive models of reference, where a cognitive model of reference is taken to be a mini-theory or a model that the subject constructs and uses as a result of the interaction of general cognitive principles with his experience. It seems to me that verbal humour and wit rely heavily on the perception of the incongruous in this way.

Relevance theory predicts that the inferential part of processing cannot stop with the entertainment of the incongruous in the form of contradictory propositional forms. An incongruity created by the means described above arises from an ostensive stimulus, hence a hearer will take it as a potential source of cognitive effects.
In particular, it enables hearers to recognize interpretive and echoic uses of language, as well as dissociative attitudes.

CONCLUDING REMARKS

I have argued that humour is to a large extent a matter of implicitly calling into question some taken for granted aspect in the world—often an attributed thought—and doing it with a certain attitude of disengagement. Also, I have proposed a mechanism through which an assumption representing the aspect in the world that is called into question becomes strongly mutually manifest through the implicit import of an utterance, and the attitude implicitly expressed gets recognized. I have shown how the entertainment of the incongruous plays an instrumental role in this. In doing so I have suggested that echoic utterances can also have as targets of their echo assumptions which they make strongly mutually manifest.

The reasons for dissociating oneself from a thought or an utterance, actual or potential, are numerous. The fact that one can express dissociation without disengagement explains why not all instances of verbal irony are humorously intended (or interpreted), though all remain instances of verbal wit.

Contrary to the cases of echoic use most commonly discussed, the examples above do not have as targets of their echo any of the assumptions explicitly conveyed by the utterance, nor any of its linguistic or paralinguistic features. Instead, the echo concerns assumptions that become strongly mutually manifest by the implicit import of the utterance. I think there are good reasons to extend the current notion of echoic use in this way. The analysis above suggests that therein may lie the route to a more unified picture of the workings of verbal humour and wit.

It may be that the enjoyment of incongruity is, after all, the essence of certain forms of humour. At a very early age children enjoy observing things that look or behave non-threateningly in an incongruous way. In attaining such amusement, a child has not had to recognize implicitly expressed attitudes or anything of the sort. However, what he has gone through in perceiving the incongruous is a drastic and sudden change in the relative manifestness of some of his assumptions. He has found a new way of looking at an aspect of the world he had already come to take for granted. The perception of the incongruity has unexpectedly made some so far weakly manifest assumptions become strongly manifest. This stage is probably preliminary to the complex abilities involved in the recognition of intentions and attitudes that we later develop.

References

Forabosco, G. (1992) Cognitive aspects of
the humor process: the concept of incongruity. *Humor* 5-1/2, 45-68


FROM ALIBI TO COLUMBUS.
THE LONG MARCH TO SELF-AWARE, REFLECTIVE COMPUTATIONAL MODELS OF HUMOR

EPHRAIM NISSAN

School of Computing and Mathematical Sciences,
The University of Greenwich, Wellington Street,
Woolwich, London SE18 6PF, U.K.
E-mail: E.Nissan@gre.ac.uk
URL: http://www.gre.ac.uk/~E.Nissan

ABSTRACT

One thing is (a) to develop a system that handles some task to one's satisfaction, and also has a universally recognized mythical side to its output. Another thing is (b) to provide an analysis of why you are getting such a byproduct. Yet another thing is (c) to develop a model that incorporates reflection about some phenomenon in humor for its own sake.

This paper selects for discussion especially ALIBI, going on to describe the preliminaries of COLUMBUS. The former, which fits in (a), is a planner with an explanatory capability. It invents pretexts. It's no legal defense, but it is relevant to evidential thinking in AI & Law. Some of the output pretext are mythical. Not in the sense they are silly: they are not. A key factor seems to be the very acularity at explaining out detail after detail of globally damning evidence. I attempt a reanalysis of ALIBI in respect of (b). As to COLUMBUS, it fits instead in (c). We introduce here the basics of this (unimplemented) model, developed to account for a sample text in parody.

BACKGROUND: A MENAGERIE

A number of programs or models I have been working at, at some time from the mid Eighties on, involve phenomena relevant to computational humor research. These projects are disparate by size, ambitions, and specific goals, as well as by whether they are specifically concerned with humor.

Dr. Elda Weizmann (Bar-Ilan University) and myself have been developing a model to account for irony. A separate paper, exclusively devoted to that model, is in preparation. Its paradigmatic background is Sperber and Wilson (1981) and the related literature.

From 1991, I have been working at another model, AKDENIZ. A formalism is proposed that captures a surprisingly large class of word-formation (derivation and compounding), of proper or even common names across languages. This class is characterized by what I term misantonymy, and it is discussed at length in a paper or book-chapter in preparation. Instances are listed and discussed,1 adopting an approach that is substantially prototype-theoretic. A sizeable portion is constituted by puns, whereas for several other cases, both origination and use apparently put on nothing else than a deadpan serious face.

Pragmathematics —not in humor—is a collection of (mildly) formalized discussions of a few cultural constructions (Nissan and Shimony 1996a,b). Through exemplification of the awkward, an attempt is made to reach insights on normalcy in common sense. A leading theme is human vs. animal roles. Within this framework, VegeDog analyzes a text on the fare of vegetarian diets for dogs, because of either the owner's ideological motives, or the animal's allergy to meat. CaroMemphis analyzes—in terms of triangles of power—an item of knowledge from Tuixidor (1985): namely, that some mummified holy baboons in Saqqara, Egypt, have been found in sarcophagi originally intended for humans from the immigrant Carian community.

At about the time I started work on AKDENIZ, my interest in puns had resulted in some implemented work. In 1990, Goded Shahaf, an undergraduate student of mine in Beer-Sheba, Israel, designed and implemented under my supervision a program. Called ARIK & BENZ, it is named after the two principals in Sesame Street...
broadcasts on Israeli television. ARIK & BENZ is a generator of Ernie/Bert equivocation dialogues in the style of Sesame Street. Equivocation is based on a generalization hierarchy, and on a notion of conceptual distance. Shahaf’s program was foreshadowed by an embryonic attempt, in 1986, to probe the principles of generating versions of a joke.2

Apart from equivocation on double senses, wrong generalizations and wrong analogies are also involved in humor. Whereas I have no model to show about this, the results of one small project, NAIVPOLITIKER, are conducive to such considerations.3 The semantics associated with the system’s logic calls for a kind of reflectiveness that belongs in commonsense reasoning, and also in computational humor. (By the way, the name, backformed from Realpolitikker, is an instance of misanthropy: for which, see above on AKDENIZ.)

The attitude towards the applicability of an analogy, however playfully proposed, may be ambivalent.4 Such an interplay among attitudinal layers belongs in several kinds of humorous communication. In a sense, this is syncategorematic —within a shared abstraction— with equivocation, puns and double entendre. Consider also ambivalent reference (e.g., by image) to both an instance and the very kind (or an archetype);5 which is not the same as jokes based on stereotypes. Moreover, distinguish between a conceptual mixup of the equivocation kind, and a contamination of repertoires, genres, or narratives.6 Expected belief is an issue anyway, for evaluating reception (including the reception of items of humor).7

The selection of an unexpected alternative is the key device, in some but not all of the above. In my automated inventor of pretexts, ALIBI, the selection of an explanation other than the one you would expect, is both the requested capability, and (part of) the reason its output is sometimes funny.

2 FOREGROUND: ‘ALIBI’

Calistri (1990), who proposes a probabilistic approach to misconception-detection in plan recognition, shortly relates the latter field to user models (Sec. 6.6.1):

“[E]ven perfect plan recognition is frequently ambiguous. As an observer traces the first few actions of an agent, or as a listener hears the first few sentences of an utterance, the complete plan that is being followed is often ambiguous [Goldman and Charniak (1988)]. For example, if we were to hear the sentence ‘Jack went to the supermarket,’ we could come up with many possible plans that Jack could be following: he could be going there to do some shopping, he could be going there to meet someone, he could be going there to rob the supermarket, or he may just be out for a stroll.

In most cases, however, we would immediately jump to the conclusion that Jack is going to the supermarket to do his shopping, and would not even consider the hundreds of other possible explanations. This sort of reasoning can be explained by using the information provided by a user model. If we aren’t given any evidence to the contrary, we would assign Jack to some generic user model, which tells us that most people go to the store to do shopping. If we have heard that Jack is a known fugitive on the FBI’s most wanted list, we might assign him to a “dangerous-felon” user model, which would prefer the explanation of robbing the store. Different user models will have different preferences for plans in the domain.”

For explanatory purposes, Calistri sought his example outside usual domains of user models. Calistri provides an example in crime to convey his point about user models more clearly. Instead, in my own ALIBI system, crime is the subject indeed (Kuflik et al. 1989; Fakhra-Elddeen et al. 1993).

ALIBI spoons imputed actions, specified in an input “police report”, of connotational or pragmatic interpretations (e.g., ‘stealing’ is reduced to ‘taking’ as in given objective circumstances). The program decomposes them into constitutive actions again and again. Then, the latter are differently reconstituted into alternative explanations of the imputed behavior. In 1990, I directed the development of one more version: ALIBI 3. Instead of just excluding incriminating elements from explanations, a mechanism is included, that explicitly computes a score of liability for plans, or parts thereof, being generated. This way, an
explanation in defense may obtain, that admits to a lesser guilt if total excuse is not possible
(Fakhri-Elden et al.). In the database of
total, knowledge in semantics and pragmatics
the latter, especially in respect of legality) is
stated about actions, their constitutive actions,
effects, and implements or settings.

Stereotypes are not explicitly involved, in
Alibi, other than in knowledge on the typical
effect: e.g., the reaction of people who perceive
a threat is typically fear, which may provoke a
weak reaction (unless it’s stronger a reaction than
the case warrants). Moreover, threats may be
indirect, and Alibi exploits this for exculpation.

Cf. Yamanaka (1995) and Gingiss (1986) on
indirect threats.

Instead, social stereotypes are central to Skill,
a companion project of Alibi, less thoroughly
developed than the latter (Fakhri-Elden et al.).

In Alibi, an accused person has to provide an
explanation (or pretext), and the system
does not rely on a model that seeks evidence
for intentions in the background of the person
indicted. Skill, instead, resorts to bias on
professionality levels as they would be expected
from given professional groups, age groups, or
(by prejudice) gender. This way, a little girl
is to be praised for baking a cake, even though
the cake may be a failure. A woman but not
a man (unless the latter is a cook) is expected
to be able to cook, because of gender prejudices,
but a career woman is an exception, because
of class prejudices (Fakhri-Elden et al.).

In Kufik et al. (Sec. 11), I discussed excuses
for incompetence, and, in particular, lines
of reasoning about guessing the truth. This is the
point in the following joke, analyzed in loco
and taken from an Italian magazine (Settimana
Enigmistica, 1973):

"In a cinema hall in the countryside,
placard announces the new program:
'A great color documentary film, The
Last Cannibals'. There is no trick.
When the image is shaky, it's because the
cameraman was afraid.'"

When it comes to skills, Basketball
appraises the professional assets of a basketball
team (Nissan, Simhon and Zigdon 1992). For
an impending match, the program conjectures
the plan of the adversary team, and proposes
a counterplan. It does, based only on
given quantified parameters —global, or of
the individual players— implicitly ignoring and
disregarding qualities other than those listed as
professional. However, the height of a player,
health conditions, and past performance, are
involved as professional parameters.

3 APOTHEOSIS INCOGNITO
(OR: HOW DID 'ALIBI' LEAK
INTO FOLKLORE)

Given an accusation, Alibi proposes an
alternative account. This account explains out
the facts according to an innocent, or at least
less reprehensible plan, for this to be ascribed
to the would-be offender to exculpate him.

Simpl(istic) text is generated as output, in
the first person, by manipulating canned text
fragments.

This section is concerned with the mutants of
items from the output of Alibi. Something
memorable seems to have taken place, in respect
of the propagation of humor or narratives. In
folklore studies, it is being realized that assuming
an ubiquitous progression from oral storytelling
to written narrative is oversimplistic. As it turns
out, some narratives have switched instead from
literacy to orality.

The following bears witness to a recession
presumably from oral deliveries about Alibi,
through orality, into popular written/graphic
media. The trigger may have been a talk I gave
in Urbino, in May 1993, to a heterogeneous Italian
audience. The sample set of inputs and outputs
I had devised for Alibi, has seemingly receded to
oral jokes, reappearing in written form, as jokes
or cartoons (and no mention of computing), in
Italy’s leading crossword and trivia magazine.

I had sporadic access to issues of Milan’s La
Settimana Enigmistica, from late 1994 and early
1995. Certain items in a few of them are relevant.

The Urbino lecture is the likeliest channel for
the reappearance of at least two of the examples
—respectively as a cartoon, and as a textual joke
with a variant (also from the lecture) transposed
into a cartoon.

As it’s in the very nature of jokes or riddles,
no mention is made of the source (the text
is anonymous, but in folklore, re-appearances
also occur). What is really important here,
is that the original context —the story having
been automatically (re)invented by a computer
program— is nowhere to be detected. The story
is unmistakable:
"In court, the defendant—an ex-con—explains to the judge: 'It’s quite false, that in the night of February 18 I plundered that shop! I just unwittingly knocked against the shop-window. the door opened, and I got inside to leave my name and address. If I was surprised by a cop with my hands inside the counter, it’s just because I was looking for paper and pen, to write down my data.' (Italian: "In tribunale, l’accusato, che è un vecchio pregiudicato, spiega al giudice: È del tutto falso, che la notte del 18 febbraio io abbia svaligiat to quel negozio! Ho semplicemente urtato la vetrina, senza volere: la porta si è aperta e così sono entrato, dicendomi che avei lasciato il mio nome e indirizzo. Se sono stato sorpreso da un agente con le mani in cassa è soltanto perché stavo cercando carta e matita per scrivere le mie generalità." (Settimana... 1995a.)

This version from the magazine exhibits a few added or slightly modified details, with respect to the relevant example from ALIBI. (Cf. in the previous papers on ALIBI.) Indeed, in the version from the magazine, the event took place on the night of February 18, and the defendant is an ex-con. The input of ALIBI is not so specific.

Moreover, in the magazine, the defendant got in through the door, instead of the broken window glass, as with ALIBI. Besides, the syntax is more complex: more than you’d hear from your typical ex-con, let alone ALIBI (that only handles rudimentary parataxis, not hypotaxis).

ALIBI does not handle complicitous agency. One thing it does, however, is having a variant of the above pretext claim that the defendant fell because he was pushed. That variant also appears in the magazine: as a cartoon. (Settimana... 1994a). That cartoon is as follows. The moon shines in the dark. A grim-faced policeman, holding a cudgel, stands in front of two stereotyped underworld characters. These stand sheepishly, one in front and one on the side of a jeweller’s broken window glass. The second offender fingers his companion, and says (as common sense has us ascribe to him the caption):

— È colpa tua: mi ha dato una spinta. ("It’s his fault: he pushed me.")

Being turned into a cartoon is the way another ALIBI story also went, or at least so it seems (Settimana... 1995b). The respective cartoon shows a lawyer, smiling sheepishly to the judge, and pleading the case of the ultimate ex-con cartoon character standing by, bald and unshaven. A halo on his head, his eyes stare upwards, cunningly. The caption goes:

— C’è stato un malinteso, signor giudice! Il mio cliente non voleva affatto rapinare la banca: voleva semplicemente un prestito e mostrava il mitra come pegno... ("It’s all a mistake, Your Honor! My client had no intention to rob the bank: he just wanted to get a loan, and showed his submachine gun to pawn it... ")

This is an adaptation of another one of the ALIBI-generated pretexts. Only, the defendant makes his statement in the first person (the setting of the claim is not mentioned), claims he got into the bank to get money at the counter (in the usual way: presumably, out of his account), and didn’t realize the employee got frightened because of the weapon he forgot he was carrying (for some unrelated goal, which, depending on the kind of weapon and circumstances, may be to having it ground—if a knife—somewhere else, or otherwise to go hunting, or while on military regular or reserve duty). In the cartoon caption, showing the submachine gun (robbers’ stuff) to pawn it is an adaptation intended to increase the vis comica at the expenses of the already meager verisimilitude.

This is ALIBI stuff. A pretext I had ALIBI (re)invent, and described in my lectures. You may reckon that the pretext is funny for the very reason it’s short on verisimilitude. Not that ALIBI is not trying hard: it produces several alternatives.

On exculpation and self-exculpation in a penal context, see e.g. Jackson (1996) in a semiotic perspective of anchored narratives. This is very different from persuasion in the context of, say, scientific argument and scholarly exposition (for which, cf. Nissan and Shimony (1996a)). Cf. however the ABDUL/ILANA program, that simulates the generation of adversary arguments on an international conflict (Flowers et al. 1982).

4 "LOOK, NO HANDS": AN APPRAISAL

In respect of humor research, the “career” some ALIBI stories had (unacknowledged) in the joke...
circuit does not detract from ALIBI being deeply problematic in respect of humor research. No explicit model of humor is embodied in this automated inventor of pretexts. "Look: no hands!" is no legitimate boast. The very fact some items are humorous, and some are not, testifies to the fact that ALIBI is not actually capturing a category in humor. It's an intersection that occurs. Is it chance? Let us get a closer look.

Some of the output pretexts may seem to be reasonable. A few are rather insipid. And some again, sound funny instead: funny, but not insipid. There is an Italian proverb predicated even by some non-Italian mouths. It's not as well known as the punning it Traduttore, traditore (which is why they don't translate it in the first place). The one I mean, is: Se non è vera, è ben trovata. "If untrue, it's well found". It's the very acerbity that attracts amused appreciation. It may be this is a subclass of the exaggeration class in the perspective of humor. Indeed, when the input is ALIBI face damning evidence, the simulated (very rudimentary) agent trying nevertheless to exculpate "himself" makes a display an exaggerated effort at concocting an explanation. Exaggerated, in face of the likelihood of achieving persuasion. (Cf. Shimony and Nissan (1993) on likelihood in legal evidence: it's not the same).

One ALIBI input has a person charged with breaking a jeweller's window glass, getting inside, wounding the jeweller, grabbing some jewels and running away. ALIBI, in the first person, claims there was an accident: "he" fell on the glass, this trigger the weapon "he" was carrying for some legitimate reason, and this wounded the jeweller. Or, then, "he" broke the glass accidentally, and got inside to leave a note. (A new accident is not so economic for the purposes of exculpation.) Or, then, "he" may prefer to play the hero: on hearing suspicious sounds from inside the shop, "he" got inside, and shot the jeweller by mistaking him for somebody ill-intentioned (towards the jeweller). Why did "he" run away, then? Well, as "he" has no medical qualifications, and there was somebody wounded, "he" ran away to get qualified help. Why did "he" run away with the jewels? Don't worry: the incapacitated owner could not guard his property, so the indicted man took it in order not to leave it unguarded. Another favor "he" was doing the jeweller. You're not convinced? Try something better than that yourself. It's an extremely damning charge. If you recognize it was you, and still are trying to exculpate yourself, you may be clever, but we are unlikely to be fooled. We may even recognize you are making the best in a bad fix, we cannot empathize but can be appreciative. We are not gazing: we are just amused.

5 VARIATIONS CON BRIO

ALIBI tries to generate a variety of alternative pretexts for the same accusation. This structure of the output, as a list of alternative exculpatory attempts based on the same input, is significantly different from, yet somewhat similar to a sequence alternating agent X's plans and agent Y's counterplans to thwart them in turn (cf. Carbonell 1981). Such a loop of planning and counterplanning is the structure of predatory/prey plots in Hanna & Barbera animated films. In these, the sheer repetitiveness of the motivational pattern in the inventively disparate quick action, is itself a humorous ingredient.9

Consider ch. 39 of Genesis: the story of Joseph in Potiphar's house (a theme proposed by the Association Internationale Bible et Informatique for a session at AIBI'94) is analyzed in HYPERJOSEPH. This project yielded a hypertext system in literary studies, with several retrieval capabilities: in respect of the lexicon, morphology, narrative motifs, etc. It is augmented with an knowledge-representation model of the narrative, as a blueprint for content-based processing (Nissan and Weiss 1994 sqq.).

The domain of HYPERJOSEPH is a narrative from the Bible, along with all depending commentaries or literary reworkings or references, across cultures and eras. The focus, however, is on the original Hebrew text of the core narrative, and on the Midrash, i.e., Biblical legendary exegesis as preserved in sources as early as the beginning of the common era, down to Medieval collections, drawn upon in the glosses of Medieval commentators of the Bible.

Genesis 39, 10 just says that Potiphar's wife tried again and again to seduce Joseph, but does not elaborate, and then the next few verses narrate the crisis in detail. Legendary exegesis, in both Jewish and Islamic traditions, makes very much of the narrative silence of v. 10. A list is given in the Midrash, of stubbornly renewed, inventive attempts at seduction or coercion, ascribed to Potiphar's wife.

The version of the legendary account in (a) the Babylonian Talmud, Tractate Yoma, folio 35, side b, features a longer list. Shorter versions
are found elsewhere in Midrashic sources.

In his history of the Hebrew folktales, Yassif (1994: 105) points out that (a) shares the format with (b) the account in the Testament of Joseph, in the Apocrypha, yet these are independent accounts, generated on the spur "of the narrative and erotic potential of Genesis 39, 10". Indeed, (a) shares the format with (b), but they don't share the specific pleas, wiles or threats Potiphar's wife reportedly resorts to.

Instead, Yassif states, this list is shared by (b) and an item in the roughly coeval Hellenistic folk literature: namely, (c), the romance of Phaedra and Hippolite. In this case, there must be dependence (Yassif, 105, and 584-5, n. 15; Perro 1975; cf. Hollander and de Jonge 1985).

6 GOODBYE: 'COLUMBUS'

Louis Ginzberg's (1873–1953) lasting legacy is essentially a compilation: a multivolume work, published in English between 1910 and 1954, in which Bible legendary exegesis from the Midrashic sources is presented syncretistically, as a coalesced comprehensive narration. The respective sources are identified in endnotes. The list of plans Potiphar's wife enacts, is coalesced, from various Midrashic sources. This editorial intervention is not innocuous, as the "ping pong effect" is made stronger, Hanna-Beer. Medieval collectiones of Midrashic stories also rearranged them. In the following, instead, we are interested in humorous creativeness that emulates the Midrash.

In this century, we find such emulation among (or interspersed within some of) the works of Hebrew novelist S.Y. Agnon. Agnon's language draws heavily on Rabbinic and Midrashic Hebrew. Weiss (1985) analyzes one of his works, in which a substantial part is outright Midrashic recreation of national history in Biblical times, through the eyes of a narrator: Agnon's own unborn soul, about to be born from this or that Biblical heroine, were the event not thwarted by mystic circumstances. The effect is elegantly humorous, but the main effect is one of fairy irreality.

In the COLUMBUS model, shortly introduced in this section, a goal/plan based analysis is developed. The object of this analysis is the incipit—the very first paragraphs—of a parody by Gershon Rozensweg, printed in Vilna, Lithuania, in 1894. The book was printed with the same layout as Talmudic treatises, namely, with the main text in the center, surrounded by commentaries. Also the structure of exposition is an emulatio. The very locale, in the publication data (thus, part of the paratext, in Genette's terminology) is very significant. Both the Babylonian Talmud and the Midrashim printed in Vilna are still the standard popular edition. (Publication data on a frontispiece, like, e.g., epigraphs, form what in literary theory is known as the paratext, in Gérard Genette's terminology (1991); cf. Nissan (1996).)

The text under discussion, the parodic Tractate 'America' (allegedly part of the Talmud Yanka'i, 'The Yankee Talmud'), pokes fun at the ongoing emigration to America. The opening page I analyzed is reproduced in the Encyclopaedia Judaica, Vol. 13, col. 127, s.v. Parody, Hebrew.

The language, style and narrative devices are like in the Midrash. America being a refuge for malefactors, and thus falling into disrepute, is made into the reason America is not named after Columbus, prophetic foresight (described in terms emulating a Midrashic locus classicus on Potiphar's wife) had him pray to the effect that such an association be spared him. The latter is narrated, phrased as to transparently emulate another promptly recognizable Midrashic passage, where Jacob prays for his name to be omitted, from the text of the Biblical verses giving the genealogy of two future miscreants: the Levite Korah, and the Simeonite Zimri. The plain sense of the relevant verse in the Genesis account of Jacob's deathbed blessings, is, instead, otherwise motivated anger directed personally at Simeon and Levi — his second- and third-born sons, and the eponyms of two tribes descended from them respectively.

Regardless of context, the default most prominent feature of concept 'America' is its being a place. In association with Columbus, however, the most prominent feature of America is inasmuch an object of discovery. In association with emigration to America, America stands out as being a "new" place: a new concept — in Old World collective geographic cognition (cf. Robinson 1992) — and, for emigration purposes, a new place having become available, conceptually, physically, and poetically (cf Gregory 1995). It's America in imaginative geography that is relevant.

The violation of a "norm" to the effect that the name of a discovery is bestowed on it, is reinforced by knowledge to the effect that
America was named, after all, after somebody (Amerigo Vespucci). For the purposes of humorous inventiveness, it’s not an economic explanation that is to be discovered in the first place; it’s rather a puzzling one. To construe it, in turn, the criterion of economy is possibly relevant. In certain parts of the analysis, the impact of the conventions of genre (the Midrash) is felt heavily. Intertextuality intervenes at this level.

Elsewhere, I am going to propose a graphic formalism of the analysis. An essential feature is that the goal/plan structure mainly applies to expression and to the genesis of the narrative, besides applying to the motives and actions of the characters.

Instead, a goal/plan based model of a category in humor, IRON-FINDER (*sic*), was developed as a Master’s project by John F. Reeves (1986) under the supervision of Michael Dyer and Margaret Flowers at the University of California at Los Angeles (cf. 1988). A sequel of Dyer’s own doctoral project, Boris (Dyer 1983)—which in turn had been done in Yale by adopting the conceptual dependency paradigm of natural-language processing—IRON-FINDER makes display of the ability to realize what the developers termed ‘situationally ironic’. Actually, however, the material they dealt with encompassed a category of jokes. The texts analyzed are understood according to plan-failure marrying characters’ planning. IRON-FINDER has much to recommend it as one of the earliest implemented NLP prototypes in computational humor, and is as relevant as ever to this emerging area of research.

Notes

1. Here are a few instances discussed in the framework of the AKDENIZ model:
   - Turkish AKDENIZ (‘White Sea’, for the Mediterranean), patterned after KARADENIZ (‘Black Sea’, for the Black Sea);
   - American English cheeseburger, back-formed by replacing the non-etymological hom- in hamburger (which actually is from HAMBURG);
   - Positivist thinker Max Nordau’s pen name, NORDAU, coined in opposition to his original family name, SUDFELD;
   - In pre-state Israel, Hebrew BOAZ for a landowner (after the Biblical Ruth’s rich second husband), vs. YAKHIN, the name of a trade-union owned agricultural company, patterned after the Yakhin Column opposite the Boaz Column in the Temple of Solomon);
   - DASH, the name Thomas A. Edison gave his son, born after his daughter Dot;
   - The punning proper name QARAMA, to be interpreted, in Aramaic, as ‘just some Mr. Coldfish’. According to an anecdote from the Babylonian Talmud, it was used at a certain academy by the faculty, after arguing with Hamnuna, a guest lecturer from a rival academy. The pun results from a mock interpretation of his name: ‘Prof. Hotfish’; as though;
   - South American Spanish ‘to go from Guatemala to Guatepeor’, for ‘to go from bad to worse’. The actual toponym Guatemala is following by one devised for the sake of the pun, by replacing mala for ‘bad’, with peor for ‘worse’.
   - In the terminology of imaging or printout orientation in computing, seascape, for the rotated image specular with respect to the more usual landscape.

2. In 1986, my students at the Ben-Gurion University of the Negev in Beer-Sheba, Anat Diani, Yaakov Iluz, and Eli Dahan, worked at unpretentious small programs, in Prolog, that somehow generated context-sensitive mutants of an input joke. Once one term is modified, the other concepts are also replaced. To suit the new context, they are replaced with concepts that respectively share the features of the corresponding original lexical concept, except the feature that discriminates the context. A contrast can be inverted, and so on. These programs did not analyze syntax, but just used to check the lexical semantics (and some pragmatics) in a semantic network. This kind of manipulation was restricted to elementary terminological or assertional knowledge about lexical concepts. In the 1986 stage of the project, processing was too rudimentary to amount to a model of punning or joking. Instead, Shahaf’s criterion of selection, based on conceptual distance, introduced an element of reasoning with more serious claims in terms of computational humor research.

3. In a paper on linguistic prototypes (Nissan 1995: 327–9, 352–3), a discussion is provided of examples of extended use of ‘Pharaoch’; in a present-day English account of the archeology of Nubia (Adams 1974), and in Medieval Latin as referring to contemporary Mamluk rulers. For our present purposes, consider the gaps between Northern and Southern Nubia. Once the Egyptian pharaonic sovereignty (ca. 1580 to 1000 B.C.E) was withdrawn from Nubia, the court and governing institutions of the indigenous Nubian empire of Kush were modelled on those of Egypt. From 750 to 650 B.C., the Nubian emperors reigned in Egypt too (the
XXV Dynasty). as the Pharaohs of the XXV ('Ethiopian') Dynasty. South or Roman Egypt, Nubia was still a single imperial state, with a Pharaonic court. However, the institutions of political control were far from uniform in the different provinces. Government in the North (i.e., Lower Nubia) was relatively secular, feudal, and with a more open economy, as opposed to the two-class society and religio-politically Pharaonic power of the South. Whereas reasoning on patterns of geographic difference and imbalance is legitimate, assuming a universal North vs. South role pattern is not. Such a pitfall is evident in (and not realized by) a didactic program, Naivpolitiķer, implemented by my students in 1989. It hypothesizes rules about certain attributes drawn out of a database of facts, and generalized, if possible, when new facts are added that contradict the old rules. Let α stand for 'has allegiance', and π stand for 'is a particular case of'. Take the following facts, that if given a semantics in real-world geopolitics, are simplistic:

- (North(Korea) α Communist)
- (South(Korea) α anti(Communist))
- (North(Vietnam) α Communist)
- (South(Vietnam) α anti(Communist))
- (Communist π leftist)
- (anti(Communist) π rightist).

The program proposes these rules:

- (North(X) α Communist) ∧
- (South(X) α anti(Communist)).

Now, suppose we add these "facts":

- (North(Britain) α Labour)
- (South(Britain) α Tory)
- (Labour π leftist)
- (Tory π rightist).

The program modifies rules to become:

- (North(X) α leftist) ∧
- (South(X) α rightist).

Once again, add a few "facts":

- (North(Italy) α progressive)
- (South(Italy) α conservative)
- (leftist π progressive)
- (rightist π conservative).

New rules are:

- (North(X) α progressive) ∧
- (South(X) α conservative).

With the first version of Naivpolitiķer, no furtherly generalized rule is found when new facts with contrasting evidence are entered. For example, on the old South Africa, where the southern city of Capetown is deemed to be relatively more progressive than the perceived norm in the former Boer republics. A second version of Naivpolitiķer is able to consult different attributes, in order to find why the new fact is different from those facts that before, had us find the latest rule. To keep it, though modified, it's necessary to circumscribe its applicability, according to some constraint on some attribute. We consider attributes as dimensions: once the distinction north/south on a national basis fails to capture political leanings, another dimension (the hemisphere) is probed; the counterexample happens to differ, as the program finds out, because of the hemisphere the country that the counterexample differs because of the hemisphere the country belongs to. A modified rule is displayed:

If (X) PART OF Northern Hemisphere THEN (North(X) Is leftist) ∧ (South(X) Is rightier).

Nothing more that the most elementary machine learning. Yet, it becomes interesting to computational humor once an explicit account of the semantics is provided, in respect of the gap between epistemic and mock-epistemic meta-information. This is conducive to accounts stereotypes belonging in a repertoire owned by an agent that in practice believes them (or claims to), yet would apply them in humor mode (whether this is legitimate or not). A mock-autoepistemically motivated goal could be: "Justify the assumption at all costs." This, however, was not part of Naivpolitiķer.

4. The ruminant Odocicilus hemionus hemionus is called Mule Deer in English. The compound reflects the Aristotelian notion of genus (here, 'deer') vs. differetia specifica (here, 'mule'). The mule analogy is apparently about shape. However, as the compound under consideration names an animal by reference to another animal ('mule') that is not its genus ('deer'), this hybrid name (which is how we may categorize the signifier) reminds of the mule main specific perceived feature: its being a hybrid. Yet, the mule-as-hybrid metaphor is not intended, in the compound. Not only: it's apt in one more unintended way. "The Odocicilus species are known to hybridize and sharing of mtDNA haplotypes was observed" (from the abstract of Murray et al. (1995)).

Prototypical 'deer' is the genus with respect to which, a 'mule deer' is so named. Taxonomies are the traditional realm of the cognitive and linguistic prototype approach to lexical concepts. Partonomies, instead, are about the organization of individuals as being made up of parts. Tversky (1990) argues that partonomies are akin to taxonomies, in that individuals of a category are [perceived to be] made up of prototypical parts, just as individuals within a category (in a taxonomy) are organized around one member that is taken to be the prototype. Tversky's discussion takes off from an example of a taxonomy about a 'pippen' being a kind of 'apple', that in turn is a kind of 'fruit' (334). The rest of this note considers associations, and then analogy or mock- analogy.
To a present-day inhabitant of Italy, Advertising makes Italians associate pippins (the mela renetta of dictionaries) with a certain region of Italy. For years, magazines have been running an ad about the mela renetta del Trentino: pippins grown in the mountains of Trentino, the region in the Alps near the Austrian border. It was annexed to Italy after World War I. The region's northern province is the German-speaking Sudtirol, whose Italian name, Alto Adige for the 'Upper Adige Valley', originated in hydrography and then entered dialectology in the 19th century. In Israel, a Druze student of mine, from the Golan Heights, once noticed, in my office, the bag of a conference in artificial intelligence I attended in Trento. He inquired. I explained to him that Trento is the chief town of Trentino, the mountainous northeastern region of Italy. I also told him that an ethnic barrier runs through that region, and that there is a nasty record concerning the political border, too; the terrain is kurst; the population is, relatively, affluent, and lives out of growing apples and of tourism. As I went on, an amazed grin broadened on the student's face, until he pointed his finger and finally blurted: "But it is just like . . ." And I: "The Golan, yeah. And they even grow Riesling vines." Riesling in Italy being typically produced in Trentino. There are also ethno-cultural, not just geological reasons for that particular vine to be grown in Trentino. As to the Golan — I learned from a chat with an export agent on a plane — there is some export of wine. "Riesling?" I had asked the expert, so he told me it was the case indeed: wine production in the Golan Heights resorts to locally grown Riesling vines (inter alia). A test-taler myself, I had inquired on Riesling in the first place, because of my own ambivalence between analogy and mock-analogy with Sudtirol. The very variety, Riesling, actually makes sense for the Golan, given the geological features of the Golan. Apart from the collective canon of knowledge (on the Syrian troublesome border, I got from primary school — and the news — in Israel back in the 1960s), an individual's subjective map of knowledge is relevant for how prominent a feature is, when coming up, say, with the Riesling association. In my case, it was a vivid memory of "Karstic" by way of a concise recollection of a visit, followed a mention of the Golan in the mouth of a high school teacher of mine in Milan. Now, a question: it doesn't occur to me which apples in particular are grown in the Golan. Pippins, perhaps? Of course, in asking that question, I am reasoning by stretching the (mock) analogy with Trentino. We could put it in terms of taxonomies and paronomasies: may the Golan and Trentino be, perhaps, instances of a fanciful abstract concept, 'Troublesome Karstic Northeast'? Or is it just the relative concept 'Northeastern region', with everything else being assertions: associated troubles, its mountainous character, its hydrographic importance, its growing apples, cattle, etc. Tversky himself discusses inference in taxonomies and in paronomasies (341–2) in such a way that is amenable to metaphor. The difference with respect to mock analogy, is that you know there is a fallacy, yet propose it. The fallacy of analogic reasoning in the example discussed, becomes all the more evident if we are to consider that the two regions, the Golan and Trentino, belong in different climates: true, the Golan (a plateau, plus mountains) boasts of the only ski resort in Israel, so Israeli adopters of that sport do currently have an alternative to traveling to the Alps. The ski resorts of Trentino are internationally well-known. But: local agriculture in the Golan makes the most of the fact the region is well inside the date-palm area. (Go down to Aden and the Horn of Africa, and the coconut area begins.) Consider this quotation from a Newsweek report on the Golan: "The climate mimics the ideal conditions of France's Burgundy and Champagne regions. Cabernet Sauvignon, Chardonnay, muscat and merlot from the Golan Heights Winery have all won international awards" (Bartholat 1994). Take notice: no mention of Riesling.

5. Does the image of a serpent stand for any exemplar of kind 'serpent'? No, if the serpent is there as a character, with any detectable intertextual link to the Genesis narrative of Eden. In the following example, both taxonomy and invoked contexts are relevant for 'apple' as well: taxonomy, because of replacement of an overspecific kind for a more general one; and invoked context, because of a jarring juxtaposition of kinds of discourse and cultural repertoire (agromarketing vs. sacred texts). In the Augusta syndicated cartoon (McGill and Poelsma 1994), successive panels in the strip show the child Augusta as she tells a snake: "I've been hearing nasty things about you!" And then, angrily: "Garden of Eden! Adam and Eve! GOLDEN DELICIOUS!" Then, as she snares angrily at the snake, her arms crossed, the snake is shown bowing its head and thinking: "It was all a terrible misunderstanding". In our present-day competence of apple varieties we may find in the market, 'Golden Delicious' is fairly central, prototypically, for concept 'apple'. To the cartoonist, it must have been a prototype of 'apple'; or, then, the signifier's prosody was appealing; see below. However, conspicuously so, 'Golden Delicious' does not belong at all in Biblical narrative. The lit-
tle girl is instantiating, with the respect to the somewhat anthropomorphized snake, a communicative situation of class 'reprimand', knowledge of which is indexed with two players, typically an adult and a child. Actually, this prototype of the event is suggested by the angle at which Augusta and the snake stare at each other, as their heads are at different levels. In the strip under discussion, we also have a play of confusing the kind ('snake'), an individual (the particular snake that Augusta is addressing), and representatives of the kind in two narratives: the strip at hand, and Genesis, which in the cultural repertoire is the prototypical narrative about 'snake'.

An account of the use of overspecific terminology in this strip (naming 'Golden Delicious' instead of 'apple'), has to pay attention to the girl's tactic of displaying knowledge of events by listing facts. She is being precise. She mentions the setting, then the two characters in the story she is referring to. When it comes to the apple, she is even too precise. She is conjuring up the image of the apple, and this image is of the Golden Delicious variety: it must be the prototype. Or, at least, it should be, because we know this is fiction, and that Augusta is no eyewitness. (An eyewitness is confronted with images at hand, however peculiar the individuals involved, as being instances of a kind. But as Augusta is no eyewitness and we know the strip is fiction, there must be a criterion by which the particular variety of 'apple' is selected.) The girl even admits she is relying on hearsay (which is an element in humor: there is a point in life when a child learns the well-known narrative from a canonic repertoire, and she treats this as hearsay about an individual in her environment).

It is likely that one of the reasons for Golden Delicious having been selected, is metric structure within the list in all three items, the stress is on the first and fourth syllables. (Only Adam and Eve have also a fifth, unstressed syllable. This should be no problem; e.g., in Italian poetry, syllables after the last stressed one are not counted among the $N$ syllables of versed classified as $N$-syllabic.)

This suggests that Golden Delicious is not necessarily the most prominent variety, or the prototype, of 'apple' is the cognition of the authors of the strip. It is more prudent to assume that prominence was a factor along with metrics, and that the authors may or may not have in mind other apple cultivars as prominent as Golden Delicious'. Therefore, this example is amenable to default by example, a notion introduced and exemplified in Nissan (1995):

"Let $F(S)$ indicate the set of features, $\{f_1(S), f_2(S), \ldots, f_n(S)\}$, of concept $S$ (this excludes features of some but not all subordinate concepts, $\{c_1, c_2, \ldots\}$, of $S$). Let $R$ be a set, $\{\rho^1, \rho^2, \ldots\}$, of semantic ranges, where at least some subordinate concepts of $S$ "pick" one out of possible features: let $\rho^1$ be the set of possible features $\{\rho_1^1, \rho_2^1, \ldots\};$ then if $\sigma_4$ has any feature in domain $\rho^1$, it may be, e.g., $\rho_1^1$, or perhaps any out of a subset of $\rho^1$. For example:

$$R: \quad \rho^1 \quad \rho^2 \quad \rho^3 \quad \rho^4 \quad \rho^5 \quad \ldots$$

$$S: \quad ? \quad \rho^2 = f_1(S) \quad \rho^3 = f_2(S) \quad ? \quad \ldots$$

$$\sigma_1: \quad \rho^1 \quad \rho^2 = f_1(S) \quad \rho^3 = f_3(S) \quad ? \quad \ldots$$

$$\sigma_2: \quad \rho^1 \quad \rho^2 = f_1(S) \quad \rho^3 = f_3(S) \quad \rho^4 \quad \ldots$$

Then, default by example allows you to replace, for a missing value (an interrogation point, in the above matrix) in domain $\rho^1$, among the features of $S$, some value from the same column, taken from the rows of $\sigma_i$.

Constraints likely to apply are coherence (which translates into "faithfulness" to picking values from the same row, if they are available there, or into some more sophisticated rule based on a few compatible rows), and representativeness, i.e., a function that selects the subordinate concept (the row) according to some degree of conspicuousness applying to the subordinate concepts, possibly concerning specific domains of semantic contexts" (330).

One example I discussed is the pictorial rendering, in Islamic art, of the rhinoceros (based on Viré 1978). Provided it had one horn on its snout, the shape its body was given may be any of canine, feline, bovine, urinate, etc. Every such instance of depiction is also an instance of the notion introduced above.

With respect to default by example, the following is, in a sense, the reverse process of exemplification: instead of imposing specificity (and unwarranted diversity with respect to the "real" thing), on a concept known only in its generality, now we are going to discuss the imposition, on a given, known kind, of generalization, and, by intentional exaggeration, of unwarranted diversification: that is, $X$ is a subordinate concept of $Y$, and the process replaces $X$ by $Y$, and then replaces $Y$ itself by the set or a subset of the subordinate concepts of $Y$.

This process is instantiated in an ad for an anti-lice lice medication, broadcasted by the Israeli radio in 1991. We have the pragmatic knowledge that lice carry a social stigma, and that still, it happens sometimes that some children carry them home; this knowledge helps us to identify, in the humorous setting the ad was given, the execution of a plan intended to
achieve a goal of detraumatizing the issue and the discussion thereof.

The ad starts with the loud, low-pitched sound of (therefore) seemingly large wild animals trumpeting; then a child says: “I used to have a zoo on my head.”

The idea for the ad is based on lice on one’s head being referred to (by jocular, contemptuous, or euphemistic generalization) as ‘animals’ on one’s head; on the other hand, ‘a multitude of animals’ is humorously described as ‘a zoo’, even though actually it is not a gamut of different kinds of animals that is involved, in the situation described by the ad: just a variety of instances of the same kind of animals is (though the ad, in the following, refers to eggs, to depict the variety of phases of the life of lice). It is as though, by exaggeration, diversification of kinds of animals was introduced.

The humor in the ad involves misleading in two stages: acoustics at the very beginning has us expect something large, and, more specifically, it has us expect a zoo, which is where we experientially know a comparable mix of sounds (unless it’s a film in an exotic setting: the dominant voice is the trumpeting of an elephant).

6.

Co-invasion of different conceptual repertoires is involved in the conventions of Stone Age cartoons. They often feature dinosaurs, even though the author is likely to be aware of (and of widespread popular knowledge to the effect that) in the vulgata of scientific discourse dinosaurs are not considered to have existed along with the caveman. Admittedly, such masalliances, so to speak, of repertoires are neither necessary, nor sufficient for humor to be involved.

Nineteenth-century painter Vedder was not being humorous, in depicting a Giza Sphinx-like crumbling statue in a riddling role like the Greek Sphinx. Consider Listening to the Sphinx, a painting by 19th-century American artist Elihu Vedder (Boston Museum of Fine Arts. Plate 55 in Wilmerding (1975)). It shows an anciortie among the ruins, kneeling, his ear directed to the huge mouth of a (silent?) statue of a Sphinx-like head, in the painting Listening to the Sphinx, by 19th-century American artist Elihu Vedder (Boston Museum of Fine Arts. Plate 55 in Wilmerding (1976)). Vedder is conveying a grand discourse, with no humor intended. Concept ‘anchorite’, as popularly perceived, fits in the role of the inquisitive agent, in a setting befitting the Sphinx. ‘Inscrutability’ is the theme, but the active character is being inquisitive: the latter rather befitting (through ‘riddle’) the Greek Sphinx, instead of the Giza Sphinx. Not only: an anchorite would be expected to abhor a pagan image. And indeed, when Anatole France, in That’s, his novel of 1890, has Pharonic-age images coming to life inside a small frescoed pyramid, this stands for the temptations to which anchorite Paphnuitus is finally succumbing. Both Vedder and Anatole France convey the accepted view, in popular culture, of anchorite ascetism. This one is related to authority within a monastic order (cf. Rousseau 1985). Vedder, instead, apparently uses the symbol to celebrate lonely, rigorous quest, disregarding the contradiction “anchorite ethos” vs. “pagan culture”.

This is a conceptual mixup, but contamination may involve narrative, instead. Contaminated narratives (coalesed versions of a story, or a merger of different stories) are a standard staple of folklore studies and literary studies. Humour has nothing to do with that. For example, given the role of the Bible in Vedder’s America, the hypothesis is warranted that there is an intertextual reference to prophet Elijah’s flight to Mt. Sinai, where the theophany he experiences, paradoxically climaxes into “a subtle voice of silence”. This paradox being humorous is prevented by active Context ‘Sacred Text’.

As to narrative contamination, may find it amusing that, say, an influential Hebrew tenth-century chronicle—the Book of Yosippon—merges the Jacob narratives of Genesis and the Roman myth of pristine Italy. Which comes with exhilaratingly garbled accounts of Aeneas (funny, if you insist on sticking to Virgil), and characters who mutate into each other: Esau’s grandchild, Sefo, defeated and imprisoned by Joseph but evaded to Italy, takes on the identity of Saturn and then Janus, and is enthroned after defeating the . . . Vandals (edn. Flusser 1978: 10–14). Present-day readings are stirred to make fun of this only because in our standard cultural expectations, a merger occurs of narratives we map far apart. In tenth-century Southern Italy, instead, it made sense, stemming as it did out of the religio-cultural urge to construct an integrated explanation of current realities of political and cultural dominance (the translatio imperii). The available, inviting building blocks were the doxa and narrative repertoires possessed, and their very disparities constituted the challenge. In a land at least nominally still under Byzantine rule, it makes sense that it’s the Vandals who retained the most vivid association with epically defeated invaders from North Africa (by Belisarius, in 534 C.E.). This way, the Carthaginian Hasdrubal meeting his end in the Punic Wars (in 207 B.C.E.) is described as leading the Vandals. His defeat, and the Vandals’, are coalesced in the Yosippon. An editorial footnote states (Flusser: p. 16) that this results in an identification of Hasdrubal with the Vandal leader Gelimer. Strictly speaking, as Gelimer is not
named, the footnote is practical, but the identification is not straightforward. Indeed, it assumes it’s legitimate to map the account on the version of history we possess. Then, to the extent we are amused, it’s by poking fun at “ignorance” (actually, different beliefs). Then, the class within humor (if applying) is funny inadequacy (cf. jokes on fools). The present remark, in turn, prevents its application by setting on a higher-priority condition: “more blâse than thou”, thus, if you are sneer at the ignorance of old, then it’s you who are inadequate, on the authority of the presently most prestigious standards of the trade. In the present terminology of the history of science, the derogatory Whigish and Whiggery denote a present-minded attitude towards the development of science, “a tendency to judge all past scientific activities by standards set by currently prevailing theories” (Stachel 1987: 60). At a meta-level, the very fact historians of science use ‘Whiggish’ in a derogatory sense, instantiates a Whiggish attitude to previous historiography.

7. In evaluating expected receptions (or making sense of observed receptions) of discourse drawing on cultural repositories, strictly speaking we need meta-knowledge on the “social” delimitation of the spread of knowledge of an item. This can be an extension of mutual belief models originally devised in AI for individual agents. See also my work on epistemic metaproperties (Nissan 1987a; 1985: Sec. 3). Consider first how this issue arises for, say, punning riddles (cf. Binsted 1995, Binsted and Ritchie 1994a,b). Suppose the setting is among locals in Milan, Italy, or its region, thus assuming that both agents in communication possess local general culture as reasonably can be expected of locals. These agents also possess mutual beliefs concerning expected culture. Moreover, let us suppose the riddle is put by a teacher of Italian literature to a teacher of Greek and Latin at a classics high school. Then, mutual beliefs become relevant, about the respective specialist knowledge and expected shared knowledge, as soon as an attempt is made to solve the riddle. Let the riddle be something like:

“What ought they to treat us to, if we were to take a flight between Milan and Budapest?”

Admittedly, there are several distracting leads. This does not detracts from the appreciation of the solution, as could be expected on the part of the principals. “Cream (Italian: panna, Milanese: pannera). Come on, between Pannonia and Pannepoli!”. The association is illusory, of the name for ‘cream’ to Pannonia, the name of Hungary prior to the Hungarian conquest. The mildly specialististic knowledge domain relevant for ‘Pannonia’ is ‘the classics’. Instead, Pannepoli is a fairly well-known, transparently formed nickname for Milan. Its originator, the well-known poet Ugo Foscolo, sarcastically mean by it ‘the city of cream-eaters’, in resentment over the reception of his tragedy Aias (‘Ajax’). It bombed because—but that is another story—of the incredibly unforeseen double sense of the exhortation: O salamis! (‘O people of Salamis’, the island near Athens). Out of context, the universally, promptly recognized sense is, instead, ‘You little fools!’ This involves devotional morphology (cf. Nissan 1987b, 1989), as a plural masculine diminutive of salame for both ‘salamis’ and ‘fool’. Sudden realization had even the cast burst in laughter (Menicanti and Spiller 1973: 132). The Salamis pitfall is related to the class of puns based on a false diminutive. Here is an example I am making up:

“Alaska? Once upon a time, it used to belong to Moscow. But in summer time, it belongs to the mosquitoes.”

8. The relation between exaggeration and humor is not straightforward. Exaggeration inspires especially awe in the following passage, translated from the ancient Egyptian by Hoch (1994a: 259):

“Whenever an amphora full of Kedy beer is opened, and people go out and get out the cups, there are 200 large dogs along with 300 jackals—a total of 500—every day standing ready at the door of the house every time I go out, because of their having smelled the sbr [‘dregs’, ‘less’?] when the jar was opened.”

Take notice of the large, unrealistic numbers forming a progression, in which two items are followed by a third number which is even larger, being admittedly their addition. It’s likely a reader was expected to detect and interpret accordingly the conventional device of exaggeration. This opens a gap in which humor is potential, instead of actual; e.g., concerning literal implications of having to face so many scavengers as a price for a drink. All of this is filtered through the screen of stylistic conventions.

For humorous exaggeration, consider the following, more recent example. It’s about trackless trolleys, a means of urban public transportation that can be viewed as a “hybrid” of a bus and a tramway. In Italian, they are called fitobus (‘wire-bus’). As the story goes (Sagredo 1961) When, around 1915, trackless trolleys were first adopted in Italy, it was in a town in Piedmont. It happened to be a town which used to be the butt of prejudices and jokes. Local patriots of neighboring towns or cities, in line with their prejudices about its inhabitants’ sup-
posed stupidity, turned its display of progress into ridicule. They put it this way: in that town, 'a tram that avoids hitting people' had come into service.

The joke is based on exaggeration: a particular quality of trackless trolleys — their silent wavering motion, whereas tramways, somewhat more noisy, proceed rigidly, without deviating from their tracks — was amplified hyperbolically. Trackless trolleys and the tramway share some features: among the attributes that set the difference, a conspicuous one is the mode of motion: rigid for tramways (on tracks, 'on iron'), versus on tires (su gomma, i.e., 'on gum', in present-day Italian transportation terminology) for buses and trackless trolleys.

The difference between the latter two is the trolley, shared by trackless trolleys and the tramway. The trolley is a constraint on deviation: trackless trolleys can only move, laterally, within a range, that is often shorter than the width of the road. Hence the wavering motion, on which attention is attracted by the movements of the trolley with respect to an axis: the wires.

In maligning their neighbor town, those spreading the joke interpreted the difference with respect to tramways (that were well known), by attributing a purpose (rather than interpreting as an effect), and the wrong purpose, to the wavering motion (absent in tramways) of trackless trolleys. They reckoned: wavering motion may be useful to avoid hitting something or somebody; silly people are expected to lack caution, e.g., when they move across the street. Thus, trackless trolleys in a town of such people have the advantage of moving waveringly: it is as though they wavered to avoid hitting people. More ignominiously, inanimate machines are ascribed cognition, volition, and skills, that passers-by are supposedly deprived of (if not of cognition and volition, then at least of the corresponding skill of avoiding the car).

9. Distinguish between the centrality of the pattern of predator-prey scheming or pursuit being repeated in Hanna-Barbera animated films, versus the frequent presence of one instance of the narrative element: e.g., at the end of several Donald/Scrooge Disney stories. Also, distinguish between a predator/prey pattern being applied again and again for the same two principals (who are ones of a kind), and the pattern being applied to individuals of a kind: repeated confrontations in Huxley's concept of the gladiatorial characteristics of nature, are neither humorous nor tiresome, but rather bleak.

Even when a predator/prey model reinstates pursuit again and again for the same actors, the effect is not necessarily humorous. It's not, say, in Grefenstette's illustration through prey/predator models, of techniques of genetic algorithms in multiagent learning (1992: Chs. 4 and 5). It's a matter of genre. Individual-based simulation of the dynamics of a model of a predator/prey system is a technical choice, which does not transform the agents into characters, even when — like in Wilson et al. (1986) — spatial scales and low vs. high mobility is involved. Readers keep in mind it's about the dimensionality of a habitat considered as an abstraction. It's not the setting of actions emplotted into a story. Or, more precisely, into a narrative that by genre conventions you would approach sub specie lepiditatis, with openness to mythical reactions.

REFERENCES


Binsted, K. (1995) "Using humour to make natural language interfaces more friendly". *AI, ALife and Entertainment Workshop, International Joint Conference on Artificial Intelligence (IJCAI'95).*


Fakher-Eldeen, F., Kufflik, Ts., Nissan, E., Pun, G., Salfati, R., Shaul, Y. and A. Spaniol (1993) "Interpretation of imputed behavior in ALMAHININ..."
(1 to 3) and SKILL". Informatica e Diritto Year XIX, 2nd Series, Vol. II, No. 1/2: 213-242.


Nissan, E. and S.E. Shimony (1996a) "TAMBALA- COQUE: For a formal account of the gist of a scholarly argument". Knowledge Organization 23(3), in press.


Nissan, E. and S.E. Shimony (1996c) "VRGEBD2G:
Formalism, vegetarian dogs, and partonomies in transition". Computers and Artificial Intelligence (in press).


Robinson, A.H. (1992) "It was the mapmakers who really discovered America". Cartographica 28(2): 31–36.


Settimana... (1972) §10653 in "Leggendo qua e là...". La Settimana Enigmistica, Year 42, No 2128 (Milan, 6 Jan. 1973): p. 4. [Data.]

Settimana... (1994a) Cartoon in La Settimana Enigmistica, Year 63, No 3266 (Milan, 12 Nov. 1994): p. 15. [Data.]


APPENDIX

Figure 1 shows a redrawn detail from Elihu Vedder's painting Listening to the Sphinx, referred to in note 6. Take notice of the contrast in degree: the Sphinx's size is large, but the volume of its supposed voice is all too low.

The remaining illustrations are related to Section "How Did ALIBI Leak Into Folklore". Figure 2 shows the cartoon (Settimana... 1994a). Figure 3 shows the cartoon in which the lawyer pleads that it is all a mistake. (Settimana... 1995b). These two cartoons may well have originated from the ALIBI examples I described in a lecture given in Italy. The cartoon in Figure 4, instead, correspond to no example from ALIBI, but it must have originated as a novel item in the same family of jokes. This one was published towards the end of October 1994 (Settimana... 1994b), two weeks before the cartoon of Figure 2.

Figure 2: "It is his fault: he pushed me."

Figure 3: "It's all a mistake, Your Honour. My client had no intention to rob the bank: he just wanted to get a loan, and showed his submachine gun to pawn it..."
--- Non riuscivo a leggerlo, il prezzo di quell'orologio, attraverso il vetro appannato!

Figure 4: I was unable to read the price of that watch through the glass; it was steamed up.
Humor Theory beyond Jokes: The Treatment of Humorous Texts at Large

Salvatore Attardo
Youngstown State University

The purpose of this paper is to begin to sketch a broad linguistic theory capable of handling any type of humorous text. Primarily, this means breaking the near-total hegemony of jokes in the linguistic analysis of humor. In Attardo (1994, ch. 8) I highlighted some useful developments in the treatment of non-joke humorous texts and reviewed the [scarce] literature.

I will assume in what follows a basic familiarity with the Semantic-Script Theory of Humor (SSTH; see Raskin 1985) and with its offspring the General Theory of Verbal Humor (GTWH; see Attardo and Raskin 1991). However, it seems fairly reasonable to assume that the broad theory to be developed will need to incorporate some technical apparatus which was unnecessary in the SSTH and GTWH.

1 Outline

The first section will very briefly outline these requirements. Because of the rather technical nature of the material I have deliberately left the detail of the implementation vague. Further elaborations will obviously have to flesh out this skeleton. Section two attempts to begin to chart the territory to be covered in a broad theory of humor by using some basic binary oppositions which include the narrative or dialogical nature of the text, the presence or absence of intertextual ties, the nature of the disjunctors (diffuse or discrete), and finally the degree to which the text is structurally similar to a joke (i.e., presents a textual construct functionally equivalent to a punch line). The next two sections develop the topic of "joke cycles" as an extreme case of intertextually based humor and irony as and example of diffuse disjunctors, in sections three and four, respectively. Section five uses the neologism "jab line" to describe a non-terminal, non-concluding punch line, discusses "hyperdetermined" (i.e., deriving from multiple sources at once) humor, and analyzes the role of repetition in humorous narrative. These three features are taken to be central in the description of longer humorous texts. The significance of the proposals outlined for a computational theory of humor is also assessed in each section.

A word of warning should be expressed about the tentative nature of most of the proposals I am about to put forward. As I have pointed out, very little study has been done on longer humorous texts and my goal is primarily that of stimulating further research in this area which is crucial to any real-world application of the theory of humor. The reader should be even more emphatically warned about the nearly non-existent computational treatment of the subject matter. I have provided a few cursory notes on how and why a possible future computational implementation may be interesting or problematic, but it is clear that first one has to have results (not necessarily coming from linguistics, obviously) and only then one can think about implementing them computationally.

2 Linguistic Apparatus

This section briefly and non-technically, in deference to the interdisciplinary nature of the symposium, outlines three areas of linguistic theory that will be used on and off in the broad theory: the "storage area" of all information accessed in the text, the notion of "script," and finally, the related notions of micro- and macro-narrative.

2.1 Storage Area

Whatever the shape of the final theory that will account for humorous texts at large there are a few features of this theory which appear to be clear. To assist the reader, I include a reference to the section of this paper in which the aspect of

---

1 For the reasons (some excellent) of this predomination see Attardo and Chabanne (1992).
the theory becomes most relevant. These aspects are:

1) the presence of a "storage area" for the information that is being assumed, shared, and developed by the text (be it a conversation between two parties or a monological text, such as a novel);

2) the fact that within the larger storage area there are privileged areas in which some or most of the normal (unmarked) features/requisites of the encoding of meaning in the text (both at the literal and inferential meaning levels) are suspended or deliberately violated2 (cf. Grice's (1975) Cooperative Principle);

3) the fact that the information stored does not travel in discrete units, but consists of "chunks" of information which in turn come surrounded by a web of associations and links to other chunks of information (cf. 2.2);

4) the fact that these chunks of information may fit like a Russian doll one inside the other (cf. 2.2);

5) the fact that the representation of the information in the storage area is not entirely linear, although there are portions of space that are linear and obey all the "Euclidean laws" of semantics (e.g., what happens before time T₀ cannot refer to time T₂₁), or, at least, can refer to representations of time which are not linear (cf. 6.4);

6) the fact that the representation of the information in the storage area admits of multiple strands of information being processed and accessed simultaneously (cf. 6.2).

I will not treat the nature of the storage area in any detail in this context. Let me only point out what I take to be the basic elements that will have to be included in a viable model of the storage area:

1. the propositional content of all the sentences which are uttered in the text
2. their presuppositions
3. all non-trivial inferences derivable from 1 and
4. all pragmatic presuppositions (cf. Caffi 1994)
5. all "accommodations" (Lewis 1979) and "bridgings" (Clark and Haviland 1977) presupposed by the text

While obviously I recognize the this is a large body of information, I would like to stress that important pragmatic principles (mostly, Relevance) prevent it from being infinite and/or indefinite. In other words, we are handling a finite, and potentially well defined set of semantic "objects." Let me finally point out that I have referred to a much more limited but similar concept as "presuppositional basis" (Attardo 1993) in the context of the communicative function of jokes.

2.2 TYPES OF SCRIPTS

As is well known the SSTH, as the name states is based on script-theory. "Script" is taken as a neutral term among the various proposals (e.g., frame, schema, daemon, etc.) and thus does not have the common meaning of the term AI. A script is defined as a complex of information associated with a lexical item. Thus the canonical example in Raskin (1985) is the script for DOCTOR, while in Schank and Abelson it is for GOING TO THE RESTAURANT. In this context, we will continue to use the term "script" as a neutral choice, equivalent to "frame."

Raskin (1985) introduces but does not exploit to their fullest potential the notions of complex and macroscript. As we will see, these could potentially be very helpful in our task.

The difference between scripts (frames), complex scripts, and macroscripts is primarily one of level: a script is the simplex form; a macroscript is a group of scripts organized chronologically (what some authors would call a script); a complex script is a script made of other scripts but without chronological organization. The RESTAURANT script in Schank and Abelson (1977) would be an example of macroscript while a complex script would be WAR. For example, if a text activates the script WAR the actants slots are likely to be filled with scripts such as ARMY or BATTLEPOW rather than individuals (such as Mary or Bob). Naturally, the presence of such "subscripts" makes it legitimate to activate such individualizing scripts as COMMANDER IN CHIEF or GENERAL.

Another interesting possibility in classifying scripts lies in the way in which they are activated: a lexical script is activated by having its lexicematic handle be instantiated as a token in an utterance (i.e., if a sentence using the word "cat" is uttered,
then we consider the script CAT to have been activated. An inferential script instead can be activated inferentially: suppose that a given text activated in rapid succession the scripts HUSBAND - LOVER - ADULTERY - PRIVATE EYE - WIFE - LAWYER - COURTROOM then a reasonable inference will activate the inferential script DIVORCE. Structurally scripts and inferential scripts are not different, and indeed the mere mention of "divorce" in the text would activate the script DIVORCE. To highlight the substantial identity between scripts and inferential scripts, I will refer to both as “scripts,” and distinguish between them only when necessary. I suggest that the use of the "inferential script" term is useful as a mnemonic device to remind us that inferential scripts are activated during the semantic/pragmatic processing of the text and can differ significantly from the surface manifestation of the text. It should also remind us that the interpretation of the text (be it that of the hearer/reader or of the analyst) is necessarily always a construct of the interpreter. Finally, it should also serve as a reminder of the fact that the larger the scripts activated in a text, the more other scripts may fill their slots.

2.3 MICRO AND MACRO-NARRATIVES

Narratology has long sought a definition of the “minimal story” i.e., the smallest possible narrative. In this context we do not need anything so sophisticated. We will assume that a micro-narrative is a story that consists exclusively of one “development” of events. A macro-narrative is any narrative which incorporates more than one development, i.e., a macronarrative consists of n micronarratives. Micronarratives may be nested within macronarratives. Thus “Mary bought a bottle of champagne to celebrate her Nobel prize” consists of one micronarrative: the purchase of the champagne with a given goal. “Mary bought a bottle of champagne to celebrate her Nobel prize. Once home she found many friends who wanted to help her celebrate and so she had to go buy more" contains instead at least two micro-narratives: the two purchases. It should be clear that I am leaving completely open the issue, which is far from trivial, of how to determine the status of a given event within a narrative. For example, the friends’ goal of helping Mary celebrate could be treated as a distinct micronarrative nested within a macronarrative. Simplifying we can summarize the distinction by saying that a micronarrative is a simple story, while a macronarrative is a complex story.

3 LONGER TEXTS: A TAXONOMY

This section sets the stage, so to speak, for the following ones by exploring and defining some concept that can be expected to play a significant part in the development of a broad theory of humorous texts. As anticipated, the discussion is organized around binary oppositions which are meant to serve a points for the triangulation of the field, as it were. Three of the topics brought up in this section (joke cycles, irony and texts structurally dissimilar from jokes) will then be explored in some further detail in sections 3-5.

3.1 NARRATIVE VS. CONVERSATION

A major difference between jokes, at least pre-theoretically, is between narrative (or canned) jokes and conversational jokes (see for fuller discussion Attardo 1994: 298-319). Canned jokes are typically told by a narrator who often prefaces the joke with an announcement of the humorous nature of the forthcoming turn and who holds the floor through the telling and releases for the reaction turn of the audience. Canned jokes are “rehearsed” i.e., they have been heard or created by the narrator before the telling. Canned jokes are generally detached from the context in which they are told. Conversational jokes are told as a regular turn in conversation, without preacing. They are created by the teller “on the fly" and are strongly context-dependent. The distinction between canned/narrative and conversational jokes has been recently questioned by the introduction of the concept of “recycling” which shows that canned jokes are adapted to the context in which they are told, often to a great extent (Zajdman 1991), thus making it virtually impossible to determine whether a joke is really conversational or if it is a clever recycling of a canned one.

Rarely do jokes occur in isolation. Canned jokes may be strung together with little or no concern for their connections, although stand-up comedians and other performers tend to introduce transitions and try to group jokes thematically. The situation is different in conversations.

3.1.1 Conversation

A humorous conversation is not the same as a sequence of jokes (cf. Norrick 1993b). Speakers

3A terminologically different but substantially equivalent way of seeing the above distinction would be to say that narrative jokes are monological, while conversational are dialogical.
tend to tell jokes that are related thematically (Norrick 1993b: 126) with the serious context and with one another. Here the GTVH seems to offer great promise. The GTVH specifically introduces a metric for "joke similarity," which essentially captures how similar or dissimilar two given jokes are based on the six KRs. The theory has been confirmed to a great extent by empirical studies (Ruch et al. 1993). Based on that metric it seems that it should be possible to determine a ranking of the likelihood of occurrence of a given joke before or after another one. Let me hasten to say that I foresee serious problems in the application of the GTVH to this problem: to begin with, despite its claims at generality, the GTVH was developed on the basis of canned jokes and its application to conversational humor may be less than straightforward. Moreover, it is not clear what the psychological reality of the various KRs is. The experimental studies mentioned above (Ruch et al. 1993) established that five out of six KRs produced ratings of similarity/dissimilarity as predicted by the GTVH. The KR that failed to perform as expected was the Logical Mechanism. It is not clear whether the abstract nature of the LM is responsible for the incorrect prediction of the GTVH. Similar factors may affect a rating of "thematic affinity."

Let me add an observation concerning the reason why speakers would choose to cluster thematically similar jokes together. After all, since they are engaging in a violation of the Cooperative Principle (Grice 1989) why aren't speakers violating the maxim of relevance as well, by choosing unrelated jokes? My suggestion is that speakers use thematic similarity to gain a certain degree of "justification" or "local logic" (Ziv 1984) for their jokes (i.e., of resolution of the incongruity, cf. Porobosco 1992: 55). By being topically relevant jokes "have a point" which topically irrelevant ones lack. The issue clearly deserves further discussion.

I will not pursue further issues relative to the narrative or dialogical nature of humor in this context. Let me however sum up by way of noting the importance of topical relevance for the modeling of humorous interactions in NLP. It is not enough to synthesize jokes. One has to come up with the right joke at the right time.

### 3.2 Textual vs. Intertextual

The next binary opposition deals with the intertextual nature of the text. The basic difference we are interested in is that between humorous texts that are "all there" (e.g., novel or play) and humorous texts that derive part of their meaning from the fact that they belong to a class of texts (e.g., joke cycles, serials such as newspaper cartoons). I will consider in some detail joke cycles in section (4) as well as briefly discuss the concept of intertextuality. These concepts are not too controversial, so we can let the issue rest until we return to it below. We instead turn to the next binary opposition in the field.

### 3.3 Discrete or Diffuse Disjunctors

Jokes have discrete, unique clearly identifiable disjunctors, while other forms of humor have diffuse disjunctors (i.e., markers). A diffuse disjunctor is any type of disjunct which does not occur alone in a humorous (micro)narrative. Register humor and irony are good examples of diffuse disjunctors since the incompatibility/inappropriateness between the context and some elements of the utterance is the sole necessary and sufficient marker of humorous or ironical intention. See below section (5) for discussion of irony and (6.3) for a smaller discussion of register humor.

The idea of a diffuse element in humor is central to Chlopicki's work on humorous short stories (1987). A diffuse trigger (or "dissipated" in Chlopicki's terminology) is "not any single word, but the formulation of the whole phrase or two, or even the whole text of the joke [which] is responsible for causing the script overlap" (Chlopicki 1987: 14). The introduction of the diffuse trigger per se does not introduce any significant theoretical difference between the SSTH and the expanded SSTH, since alliterative puns also present diffuse disjunctors (cf. Attardo 1994: 139).

The central issue at hand is the relationship between diffuse disjunctors and punch lines. Punch lines are equivalent to disjunctors (Attardo 1994: 87). Disjunctors2 are equivalent to script switch

---

4To a degree: LM is "behaving badly" in relation to the KR immediately before (SITUATION), but works fine in relation to all other KRs.

5Note that there are two levels of cooperation (or lack thereof): one within the jokes (and Relevance is violated without qualms) and the second in the larger conversation/narrative, where the thematic similarity principle holds.

6The technique is spelled out in Hockett 1973 and applied in Attardo et al. 1994.

7The disjunct is the item that cause the passage from the first to the second script.
triggers (cf. Attardo 1994: 82). From Chlopicki’s definition above we gather that a dissipated trigger is however not in fact a disjunct or trigger in the sense of the STTH, but rather a connector, i.e., the textual element that makes the copresence (overlap) of two scripts possible. In puns this is quite often simply an ambiguous element (cf. Attardo 1994: 134-135). Narratives do not necessarily have ambiguous elements, so it becomes harder to see how a dissipated trigger/connector should work.

In sum, diffuse disjunctors remain to be accounted for. I hope to provide a contribution to this problem by introducing the concept of “jab line” (3.4.2) and by considering diffuse disjunctors as a type of jab line. We now move on to the last (and perhaps most significant opposition, that between texts that are structurally akin to jokes and those who are not.

3.4 Structurally similar or dissimilar to jokes

The literature on humor has emphasized the importance of the punch line. However, not until recently (Ozing 1989, 1992; Attardo et al. 1994) have there been claims that the position and nature of the punch line structurally determined the type of humorous text. Roughly, Ozing’s point is that a joke must end on a punch line, which meets the requirements spelled out by Raskin (1985), i.e., the punch line must cause the actualization of a second script which overlaps with the other one in the text (script overlap) and is opposed to it (script opposition), in a technical sense discussed in Raskin (1985: 108). This theoretical claim has been confirmed by an empirical study (Attardo et al. 1994). Thus it seems a viable hypothesis that there will be humorous texts not commonly classified as jokes that are nevertheless structurally homologous to a joke (i.e., they end in a punch line). A case in point would be a micronarrative within a macronarrative. We will discuss this in further detail in (6.3).

The complementary claim is obviously that there will be a class of humorous texts that are structurally dissimilar from jokes, i.e., that do not end in a punch line.

3.4.1 Structural vs. local humor

A way to address the issue of the differences between jokes and other narratives would be to try to capture the difference between humor that belongs to the plot of the story and humor that is external to the plot. We can start by the interesting observation that they are analogs of the verbal/referential humor distinction: you can paraphrase or translate away the local humor, but you cannot paraphrase away the structural humor (Attardo 1994: 27-29). To work with the concept of plot is not very easy, given the difficulty in establishing what the “plot” of a story is. We will attempt to sketch a way to handle this issue in section (6.4). So it is perhaps better to try to attack the problem from a different angle, outlined in the next section, and discussed further below.

3.4.2 Jab line definition

“Jab lines” are humorous parts of the text which are essential to the narrative in which they appear or to the development of the text itself. Conversely, when we are dealing with a narrative not structurally equivalent to a joke, punch lines are not essential to the macronarrative (although they obviously conclude the micronarrative they are the disjunctors of). I suggest that by introducing this distinction we may be able to handle diffuse disjunctors in humorous texts not structurally similar to jokes; see (5.3).

4 Joke cycles

After having presented four oppositions we come to some concrete example and more detailed treatment of three of the central areas of the theory of longer humorous texts: joke cycles (below), irony (5) and, last but not least texts structurally different from jokes (6).

We will start out, appropriately enough since we are building up from them, with the consideration of jokes, but no longer seen in isolation or compared to other jokes individually (as the concept of joke similarity in the GTVH implies) but as part of large clusters of texts all mutually related.

The main purposes of this section are to present the notion of joke-cycle, and to sketch
the relationships among the texts involved in a joke-cycle.

4.1 Definitions

The notion of joke cycle originates in folklore studies. At a basic (and intuitive) level a joke cycle is a set of jokes that are related. The prevalent relationship seems to be that of thematic links between the jokes, often mirrored in the folk-taxonomies of jokes (lightbulb jokes, elephant jokes, sorority girl jokes, space shuttle jokes, Watergate jokes, Italian jokes, etc.). While the subject matter of the jokes is clearly important, the GTVH has argued that this is not the only link among the jokes, and moreover that it is not the most important link among the jokes.

4.2 A Little History

The “original” lightbulb joke runs as follows:

(1) How many Polacks does it take to screw in a light bulb? Five—one to hold the bulb and four to turn the ceiling (chair). (Dundes 1987: 143)

Clements (1973: 22) reports 28 versions of this joke in the Indiana University Folklore Archives, prior to 1969. By 1978-79 the light bulb joke cycle “had swept the country” (Dundes 1987: 144). The collections of lightbulb jokes currently available add up to more than a thousand of variants, targeting hundreds of groups and individuals (e.g., Guntheroth (1990) and Marcush (1996).

Thus, from the available historical evidence, it appears that lightbulb jokes originated as an ethnic slur, in the “canonical” form shown above (1), where the implied insult is stupidity. Soon a large number of jokes emerged where the charge of stupidity, essential in the original “lightbulb joke,” had been dropped, and instead the way in which given groups performed the action of “lightbulb screwing” was used to point out the peculiarities of the targeted group (Kerman 1980).

4.2.1 Two generations of jokes

These “second generation” jokes are based on an implicit intertextual (see Norrick 1989) reference to the original lightbulb joke, since otherwise the frame “joke” would not be established, and the texts would simply be absurd. It should be noted that the absurdity of the text when the intertextual reference is missed can be read as humorous, thus complicating the analyst’s task. It is however clear, at least theoretically, that the hearer would be laughing at a different joke if he/she does not understand the intertextual reference, see below.

A “third generation” of jokes emerged in which the teller fails to deliver a lightbulb joke, and in fact delivers a joke based on the fact that the hearer was expecting a joke and does not receive one. Or, to put it in Lefort’s words: “the incongruity is that there is no incongruity in this [...] type of joke” (1992: 154). This is known as a second degree joke, or meta-joke (cf. Attardo 1988: 359-361, Lefort 1992: 153-154).

4.2.2 A Definition of Intertextuality

I should begin this attempt at a definition of intertextuality by noting that the concept is fairly controversial, and there is disagreement on the boundaries of the phenomenon. In any case, a reasonable definition could be the following:

A text \( T_i \) will be said to have an intertextual relation to another text \( T_j \) when the processing of \( T_i \) would be incomplete without a reference to \( T_j \).

The nature of this “incompleteness” is essentially open, but it may involve reference to any of the elements that constitute a text (e.g., meaning, its formal organization, such as word choice, syntactic structure, the circumstances of its production, etc.). The most common forms of intertextuality are the quotation, in which \( T_i \) includes a fragment of \( T_j \) in its body, the paraphrase, in which \( T_i \) states the same contents of \( T_j \) (or a fragment thereof) in different words, and the parody, in which while reference is made to the formal organization of \( T_j \), \( T_i \) more or less subtly jokes fun at \( T_j \) by changing the contents of the text. It should be noted that originally, as is apparent in the etymology of the word paroody (para-odon) no element of ridiculing was present in the idea of parody, and in fact some literary devices can be seen as non-ridiculing parodies (for example, Joyce’s mapping of Ulysses on the Odyssey).

4.3 Reconstructing the Joke Frame

The basic problem of the intertextual relationships of joke cycles is to determine that the joke frame has been activated, or in other words, that
a given text is, say, a lightbulb joke (henceforth, LBJ) or a sorority joke.

(2) "How many Californians does it take to screw in a lightbulb?" "Ten. One to screw it in and nine others to share the experience."

Compare the text above (where the joke-frame is established) with the following one where it is not:

(3) **"How many Californians does it take to make toast?" "Ten. One to make it in and nine others to share the experience."**

(3) fails to activate any known script of jokes, and hence the hearer is faced with the problem of deciding whether (3) is a joke or not, whereas (2) successfully activates the intertextual script for LBJ and hence the hearer is certain that the text is a joke.

Naturally, given a sufficiently intelligent hearer, he/she will be able to process (3) and identify a script opposition between the triviality of the activity of making toast and the fact that one would want to invite nine people over to witness the event. If the hearer has available the humorous script that Californians are especially gregarious, then he/she will be able to decode the allusion to the script, and will draw the required inferences (Californians are acting according to stereotype) and draw some humorous pleasure from the facts.

In fact, even someone who has never heard a LBJ, and is faced with (2) may successfully decode the text, identifying the script opposition. However, he/she will be missing the intertextual reference to why the teller is specifically choosing lightbulb changing and not, say, making toast or washing one's teeth. Incidentally, let me point out that this is not an entirely hypothetical discussion. One of my students has claimed that she had never heard the "original" LBJ, and yet she was able to appreciate a LBJ similar to (2).

This last consideration opens the way to a question: what is the status of the primacy of the "original" LBJ? Clearly, no actual psychological primacy is likely to be the case, as the above discussion proves. It may be that the case recorded is not unique and perhaps many speakers who are familiar with many LBJs have never heard the "original" LBJ. Historical primacy seems to be confirmed by the available folkloric sources. But from a linguistic point of view, and from the point of view of intertextuality it is not obvious where the primacy lies.

Clearly, from the speaker's point of view, it is impossible to produce a LBJ without having ever been exposed to either the original LBJ or some examples of the second generation of LBJs. Therefore chances of randomly recreating the combination of KRs that make up a LBJ are negligible. Thus, we can safely assume that LBJs assume previous knowledge of the joke-frame on the part of the speaker. From the hearer's point of view, on the other hand, as we have seen, no previous knowledge is required, but that does not exclude knowledge of the LBJ frame, which is achieved inferentially. In other words, the hearer may well have been unaware of the existence of LBJs, but after hearing one he/she becomes aware of the existence of one such joke, which works in this case as an isolated joke, and not as a joke-instantiation participating in a joke cycle.

The previous discussion is based on the second generation of LBJs, roughly definable as parodies of the original LBJ. With the third generation of jokes the situation is different. Again, speakers must be aware of the existence of the LBJ frame, but in this case hearers must be aware of it too, otherwise the joke is impossible to understand.

4.4 Consequences for the NLP of humor

Essentially the above discussion has the purpose of pointing out that templatic generation of jokes, however sophisticated, will not match the naturally occurring data in two respects: intertextual and meta-jokes will fail to be accounted for because they are based on the subversion of the template. Naturally, some templates are broader than others, and the choice of the LBJs will tend to downplay the breadth obtainable by templatic generation, since the template of the LBJ is fairly restricted, if compared to the "What do you get when you cross an X with a Y?" where the only constant is the NS KR. Thus intertextual and meta-jokes (such as most LBJs) mark the upward limit of what can be achieved by templatic generation at its best.

5 Irony

As anticipated, we turn to the discussion of irony as a prime example of humor with diffuse disjunctors. The present treatment of irony depends
largely on an unpublished section of Raskin and Attardo (1994). A fuller discussion, including a review of the field will be found in Attardo (forthcoming). The background of the discussion is Grice's theory of conversational cooperation (Grice 1975; 1989), summed up in his famous Cooperative Principle (CP).

5.1 A CONTEXTUAL INAPPROPRIATENESS THEORY OF IRONY

Ironic is defined by Webster as "the use of words to express something other than and especially the opposite of the literal meaning." Let us take this informal definition as operational.

Suppose that the speaker (S) utters ironically an utterance of meaning M. The hearer’s (H) reconstruction of the intended meaning will be based on a set of shared presuppositions: H knows that S cannot mean M, and S knows that H knows that, and therefore S can count on the fact that H will not stop at S’s literal meaning M, but rather will look for a more suitable meaning among the infinite set of other meanings (i.e., non-M).

To be noted is also the fact that Grice’s CP, momentarily "suspended" when an ironic utterance is first encountered, becomes fully operational again once the first step of rejecting the literal meaning has been taken: for example in:

(4) S: "What nice weather." Context: it is raining.

H will assume that the utterance is relevant to the condition of the weather, and not, say, to the location of the cat.

Grice remarked that irony is problematic in a straightforward implicational framework because "irony is intimately connected with the expression of a feeling, attitude, or evaluation" (1989: 53). On the basis of the above observation that the CP is brought back to bear on the interpretation of the ironic statement after its violation by a contextually inappropriate utterance, it is easy to see how the expression of a speaker attitude towards the ironical target would fit the descriptive framework, since the ironical utterance would be interpreted as referring, cooperatively, to some element of the context.

We have seen that irony is non-cooperative at first reading. In what respect does irony violate CP at that first moment? This is an interesting issue, since every ironical utterance seems to be literally false and/or not appropriate to its context. Let us consider a few examples: If one says

(5) I love children so.

while, in fact, one dislikes them, clearly, one is technically lying, but one’s tone of voice or other signals may be making it clear that one is deliberately and conspicuously violating the maxim of quality and signalling it to the hearer(s). Then one is not "really" lying, since one wants to be "outguessed," but rather is being ironical.

Katz and Fodor's famous example (6)—the only well-known example of a seriously under-explored category of appropriateness—is neither true or false, when pronounced in the daylight, but it is inappropriate, i.e., it violates the rules that determine the deictic anchoring of discourse in reality.

(6) This is the happiest night of my life.

In the appropriate context, (6) could also be ironical (if for instance pronounced in the early morning by a speaker well known for his late-rising habit). Or consider the following situation:

(7) Two farmers in a drought-stricken area are talking and farmer one says: "Don't you just love a nice spring rain?"

while probably literally true (farmer A and B may like spring rains) the utterance is contextually inappropriate because it is not raining.

The earlier example, "What nice weather," uttered while it is raining, clearly belongs to the inappropriateness category of irony as well, but unlike (6), it also involves a literal non-truth. In other words, more than one maxim can be violated in an ironic utterance, just as the violation of more than one maxim at a time in a joke is a common phenomenon (see Attardo 1993, and references therein).

In fact it is possible to extrapolate these observations and define as ironical an utterance that, while maintaining relevance, explicitly or implicitly violates the conditions for contextual appropriateness, either deictically or more broadly in

10While I gratefully acknowledge Raskin’s contribution to the ideas discussed in this section, he cannot be held responsible for this treatment.

11Mere literal falsehood is insufficient to describe irony, as pointed out by Wilson and Sperber (1981). Literal falsehood is a case of contextual inappropriateness if appropriateness includes adherence to the CP.
terms of the knowledge by the participants of the opinions and belief systems of the speakers (see Searle 1979: 113 for a brief mention of an account of irony in terms of inappropriateness).

Let us note also that I am introducing here an interesting exception to CP, since we are drawing an inference on the basis of a rule not included in it: "be contextually appropriate" (which is not the same as be relevant). The most obvious consequence is that we are hereby extending Grice's CP.

5.2 CONSEQUENCES FOR HUMOR THEORY

The relationship between humor and irony is complex. Not all irony is humorous and not all humor is, obviously enough, ironical. As far as humorous irony goes, we can say that it is a prime case of humor in a text without a clear punch line (disjunct). While there may be clues or markers of humorous intention these are by no means necessary features of irony. Deadpan delivery of irony is a common, if dangerous, practice. If an ironical intonation (or other ironical marker) is present, one could argue that the suprasegmental marker is in fact the punch line (although this would be strained, since we would have a punch line lasting for an entire utterance). If no marker of irony is present, detection and processing of the irony can only be triggered by contextual inappropriateness, which means that either parts of the text will be incompatible with other parts of it, or that the text will be incompatible with the context in which it occurs.

In either case, the apparatus of the GTVH would, in principle, be able to handle the humorous nature of the text, but would fall short on the textual processing thereof, which requires the capacity to handle a disjunct which is not unique and cannot be pinpointed in one place. The broad theory proposed here would be able to handle the contextual appropriateness issues necessary to process irony, because of the concept of diffuse disjunct.

5.3 CONSEQUENCES FOR THE NLP OF IRONY

A contextual appropriateness theory of irony is significantly simpler to handle computationally than a mention or, worse, an interpretive theory of irony. It relies essentially on tools that would already be present in a NLP system capable of addressing contextual appropriateness and cooperative behavior, in the Gricean sense.

6 TEXTS STRUCTURALLY DISSIMILAR FROM JOKES

Before directly addressing some types of texts that differ structurally from jokes, we will consider two theoretical approaches at the problem. In Attardo (1994: 221-222) two approaches to the treatment of non-joke humorous texts from the perspective of the SSTH are presented: the expansionist approach and the revisionist approach.

6.0.1 The expansionist approach

The expansionist approach essentially reduces longer humorous forms to complex cases of jokes, by postulating that the underlying mechanisms of script overlap and opposition are the same across genres.

Chlopicki's expansion of the SSTH postulates a substantial identity between jokes and humorous short stories since they are both analyzed as reduced to oppositions between pairs of (shadow) scripts. Some discussion of Chlopicki's work in relation to the SSTH can be found in Attardo (1994: 209-211), but it will be useful to summarize briefly his methodology. First, all the possible script oppositions in the text are identified. This is an important step, since ordinary jokes usually have only one opposition. With short stories, the analyst is faced with many more script oppositions (66, in Chlopicki's first example). Analysis of the short stories reveals some
scripts that extend through several sentences and even through the entire text (the "main scripts"). The "shadow oppositions," which were already present in Raskin's formulation of the SSTH, are the deeper script oppositions, whose scope encompasses the entire text and which are responsible for the overall perception of humor, rather than for the individual surface oppositions (Chlipicki 1987: 19). These scripts are found to overlap with other scripts with which they bear relations of opposition.

6.0.2 The revisionist approach

The revisionist approach, on the other hand, consists of acknowledging the irreducible nature of generic variation and in incorporating a generic specification in the make up of the humorous text. Concretely, this was accomplished by the inclusion of the NARRATIVE STRATEGY (NS) knowledge resource in the GTVH. The advantages and drawbacks of both strategies are reviewed in Attardo (1994).

We now move to the consideration of four phenomena which play a significant part in non-joke humorous texts:

1. Repetition
2. Hyperdetermined humor
3. Jab lines
4. Plot

6.1 Repetition

Repetition is not a big player in jokes, although there are some examples of uses of repetition within canned jokes, reviewed by Norrick (1993a):

- the 1, 2, 3 formula (Norrick 1993a: 386-387 "repetition with variation")
- repetition routines used by children to embarrass the speaker (Norrick 1993a: 385-386)
- knock-knock jokes (Norrick 1993a: 388)
- intertextuality (Norrick 1993a: 389; 1989)
- alliteration (Attardo 1994: 139)

Repetition in spontaneous conversational joking is also documented where it primarily takes the form of the mention of a previous speaker's word's (e.g., for ironical or punning purposes; cf. Norrick 1993a for discussion and examples). However, the presence of repetition inside the jokes should not lead us to forget that it has been repeatedly noted that the repetition of jokes diminishes their humorous effect. Cf. Attardo (1994: 289-290) on the connection between implicit and surprise in jokes.

Conversely, repetition is a very big player in longer texts, for example, repetition is a well known feature of comedy (and a big headache for theories based on surprise, naturally). Cf. the following quote from Charney (1978: 82) "Repetition may be the single most important mechanism in comedy." Consider for instance the catch phrases of many sitcoms, or the repetitions of narrative motifs in Feydeau's comedy, or the repetition of phallic images in Allais (Attardo 1996). Repetition can be accompanied by slight variation, thus introducing an element of novelty as well as the pleasure of virtuoso variation. Also, introducing repetition will require progressively more elaborate justification (in the technical sense).

It seems that pure repetition of a given element can establish a "strand." "Strand" will be further defined below as a bundle of punch and jab lines. "Element" should here be construed broadly as ranging from semantic features to broad motifs and even to large cultural scripts. This implies that the storage area must keep a running tally of the frequency of tokens in the utterances that make up the discourse. That is yet another load for a computational analysis.

A major question, which remains open, is how to differentiate between the "normal" repetition of semantic features found across the board in language (e.g., agreement, anaphora, subcategorization, cohesion, etc.) and repetition for humorous purposes. We will not address this issue in this context and instead move on to the next phenomenon central in non-joke humorous texts.

6.2 Hyperdetermined Humor

We can define hyperdetermined humor as the presence of more than one active source of humor at the same time. For example in Han Rybeck by Allais (Attardo 1996), we find satirical references to writers of the time and the text is structured as a parody of the themes of exoticism which they were fond of (strand 1), but we also find a theme of sexual exuberance (reinforced by a number of phallic references) (strand 2). These strands, which by and large constitute the plot
(which interestingly starts on the first theme and ends on the second), are interrupted repeatedly by onomastic and topographical puns (strand 3) and authorial asides (strand 4). In one case, a punning "gag" hijacks the narrative for a large part of the text (about one third) and the self-contained narrative excursion is in fact built to justify a pun. It is interesting to note that this excursion has largely the structure of a joke (punchline at the end).

Let us note that the SSTH and the GTVH cannot elegantly handle this type of hyperdetermined humor, since they are programmed to identify a unique humorous trigger. The GTVH's NS knowledge resource could, up to a point, handle some of the complexities of these texts (for example by having as one of its options "short story," to parallel the question and answer, riddle, narrative, etc. formats which are some of the fillers in jokes), but clearly what is needed is a much broader reconsideration of the issue.

Let me point out that in a sense the basic intuition behind GTVH and SSTH remains valid: all these themes and their various humorous triggers and/or kernels can be handled individually by the idea of script opposition and overlap and are ultimately examples of incongruity and resolution. What is lacking in humor theory is a sense of how to handle more than one humorous line at the same time.

I would like to reiterate that these observations should be taken as clearly preliminary; they need to be expanded by further research and by broader sampling. Temporarily, I have proposed to expand the GTVH with a storage area for the build up of humor strands. At this point it is too soon to attempt to describe in any detail the structure of the storage area. Let us begin, however, by noting that strands are not structurally analogous to jokes, or, in other words, they do not have a punch line, although they can exploit a punch line to establish themselves. This fact will have to be reflected in the storage area.

6.3 PUNCH AND JAB LINES

We get now to what is perhaps the central issue concerning non-joke humorous texts. Oring (1989, 1992) has claimed that the presence of a punch line in final position is the defining element of the narrative joke: "the punchline [sic] is not a necessary element of humor but a literary device that characterizes the particular form of humor we label 'joke.'" (Oring 1992: 83).

The most interesting issue are longer texts which are structurally dissimilar from the joke (i.e., no punch line at the end). These contain two types of humor triggers:

- punch lines, which as the name indicates means that they are self-contained micronarratives (i.e., jokes) embedded within the larger narrative. A punch line effectively ends the narrative, unlike a jab line. See for example a punch line closing a scene in Peacock's Headlong Hall analyzed in Attardo (1994: 262-265)

- "jab lines", which are humorous turns/events which are essential to the macronarrative in which they appear (i.e., they are indispensable to the development of the "plot" or of the text). For example, the malfunction of Persky's machine which prevents Kugelmass from sending Emma Bovary back into her novel is humorous in and of itself but it is also the ultimate cause of Kugelmass' undoing (Allen 1981).

Naturally, not all jab lines are narrative elements. In the case of register humor, the presence of several markers of, say, a highly formal register in the context of a trivial situation, or of a situation which is usually associated with informal registers, will work as jab lines, since obviously they do not interrupt the narrative flow.

Oring (1992: 85) calls narrative endings that do not create an incongruity or resolution but resemble punch lines (in that they occur finally and are stylistically similar to them) "pseudo-punchlines." Some jab lines would qualify as pseudo-punch lines, but not all jab lines occur in final position.

The most significant issue concerning jab and punch lines is how the theory and the analyst are going to differentiate between them. The answer is that they can be differentiated only on the basis of their functions within a macronarrative. In itself a punch line and a jab line do not differ. Their textual function is different, however. The following are a small set of heuristics which should help in the process of formulating a complete theory.

- Does the line have the same actants as the macronarrative? If yes, no evidence, if no → jab line.
- is the line part of the (macro)scripts being developed? If no → jab line, if yes, no evidence.
- does the line cause the instantiation of a
(macro) script which is instrumental in developing the final (macro) script? If yes → jab line, if no, no evidence.

We can now return briefly to the definition of strand. As anticipated, I defined a strand as a bundle of lines (either punch or jab lines). It is clear that with the term "bundle" we are actually referring to the repetition of jab or punch lines. The punch line of a microromantic may naturally function as one of the lines in a broader strand. Summing up: lines are the smallest humorous "events" in a text. There are two types of lines, jab lines which do not disrupt the narrative and punch lines which end a narrative. The repetition of lines thematically or otherwise related creates strands, which appear throughout a text, often concurrently.

6.4 Plot

Last but not least, we finally come to the complex relationship between plot and humor. In Attardo (1994: 265) I suggested, in the context of the discussion of register humor, that one may want to distinguish between a non-humorous "narrative core" of a text, whose function is to make the story advance, and the humorous parts of the text. The idea was presented as speculative and extremely tentative. At the time of writing, I was unaware of the work of Palmer (1987) on humor in film and television. Palmer had presented essentially the same ideas I had advanced, but in a more elaborate form.

Palmer distinguishes two main cases:

1. "the narrative [...] consist[s] of nothing more than the articulation of jokes together in a joke sequence" (141)

2. "jokes will be linked by something which is not in itself comic, in other words some form of non-comic narrative" (Ibid.) In this latter case, the relationship between jokes and narrative may be as follows:

(a) "the non-comic narrative is no more than a series of links between jokes" (142)

(b) "the narrative serves some further purpose" such as character development.

Palmer discusses the work of another film scholar, Terry Lovell, who argues that all comic plots are in fact non-comic plots "turned into the comic mode through the inclusion of comic material" (Palmer 1987: 144). While I am not sure that Lovell and Palmer are right in their claim that there are no specific humorous plots, it is certain that some humorous plots are produced by the technique outlined by Lovell. Moreover, they account for the fact that the degree of disruption of the "naturalistic/realistc" narrative may vary, thus giving us "realistic" comedies (e.g., Austen's Emma) with low degree of disruption of the narrative frame and "crazy" comedy (e.g., Blazing Saddles, Space Balls, The Icicle Thief) with an extremely high level of disruption of the narrative (such as the superb self-referential scene in which the characters in Space Balls rent a video tape of Space Balls and fast-forward through it up to the point in the plot in which the characters in Space Balls rent a video tape of Space Balls and then are astonished when looking at the screen they see themselves looking at the screen).

Let us note some weaknesses in the picture outlined above: as I hinted, there seem to be humorous plots, which I have referred to as narratives that are structurally analogous to a joke (i.e., end in a punch line). An example is a little known short story by E. A. Poe analysed in Attardo (1994). The idea of disruption of the realistic narrative as the constituent element of humorous narrative is put under serious doubt by the observation that self-referential devices (which totally disrupt realistic narrative) have been used in tragic theater (e.g., Pirandello's Six Characters in Search of an Author) without comic effects. With a word of warning about the fact that Palmer himself is not uncritical of Lovell's position (Palmer 1987: 145-147), we may consider what these ideas have to offer.

Some, if not all, of the weaknesses pointed out above can probably be taken care of by a careful formulation of the theory, for example allowing only certain types of disruptions to count as humorous. In any case, the central point from our current perspective is that the development of the plot/fabula14 is stored in the storage area and is then accessible as a topic of humorous manipulation. For example, coincidences or other highly improbable events, would be avoided in naturalistic narrative, but are normal fare for humorous narrative. Palmer (1987: 115-140) presents an analysis of Faulty Towers focusing on the highly improbable "bad luck" of Basil Fawlty, the neurotic owner of the hotel who seems always to be

14 The fabula are the events narrated in the text in their chronological order, the plot are the events in the order they are presented in the text. Flashbacks, for example, present events that happened before a time T₀ after T₀.
having the worst day of his life. Consider the following example, which strikes me as paradigmatic:

(8) [Basil] turns the bathroom light switch, which is just outside the door [...] [Raylene, an attractive guest] moves to the wall by the bathroom door [...] Without looking, [Basil] reaches out of the bathroom for the switch. His hand engages Raylene’s left boob. He tries to switch it on, sense something is wrong, and feels it. Raylene looks down in disbelief just as Sybil [Basil’s wife] enters the room. (Cleese and Booth 1983: 202-203)

It is already highly improbable that the attractive guest would choose to position herself exactly in the only position of the wall where Basil would be looking for the switch, but that Basil’s wife would walk in exactly at the moment he is feeling her breast is totally improbable. Presumably the information available in the storage area and the known encyclopedic information about the likelihood of events (not to mention their social consequences) interact to mark the above violation of naturalistic narrative conventions as humorous.

This approach to the interpretation of humorous texts which has access to the narrative development of the text as it is processed has the advantage of accounting for metahumor, which can be easily explained as a play on the expectations built by the inclusion in the storage area of the opening sequence of a known humorous sequence/narrative, which are then deliberately thwarted.

Let me add a final note, again concerning the nature of the representation of the data processed in the storage area: in the case of narrative we have two tools which are complementary, namely scripts, which as is well known play a significant part in the interpretation of stories and acts’ (characters) plans, which are inferred by the interplay of their utterances, the knowledge about the context as it is represented in the story, and the CP. Plan/goal based interpretations of narrative are well known, e.g., Wilensky (1982); script-based approaches are also well known (cf. a review in Raskin 1985), but could profitably be enhanced by the use of the various levels of embedding of scripts and macroscripts. From this perspective, interpreting a story would essentially boil down to instantiating a series of lexical scripts and inferentially instantiating progressively larger scripts that would eventually yield the actual script of the individual story. Naturally this type of story interpretation would be ideal for our purposes, since it would provide a seamless juncture with humor theory.

Regardless of the fact that the two approaches are presented as antithetical, I believe that they are in fact complementary and could and should be fruitfully brought to bear on a text at the same time.

7 Conclusion

I would like to conclude this article by repeating the warning I gave at the outset: most of the avenues of research I mention in this overview are new and tentative, some are outright speculative, and some yet I came up with while writing this article. What little research has been implemented has been reported, but my goal in writing this piece was not to report on a field as it is but rather on a field as I think it should be. Several of the points that I make in one or perhaps two paragraphs should receive a much more thorough treatment in book-length monographs. So I apologize to the reader if I sound at times more prophetic than scientific, but that is the price to pay to try to change the state of the art. I will return to some of these points in the future, and perhaps to all of them, but for the time being I can only hope that my thoughts may bring others to work on these problems, if only to prove me wrong.

8 References


Guntheroth, Kurt. 1990. Canonical Collection of Light Bulb Jokes. Posting on rec.humor 6 Jul 90. (e-mail kurt@tc.fluke.COM).


A NEURAL INVARIANT OF HUMOUR

B.F. Katz
School of Cognitive and Computing Sciences
U. of Sussex
brucek@cogs.susx.ac.uk

ABSTRACT
It is argued that although it may be difficult
to discern an invariant of humour at the level of
the stimulus, one emerges as a consequence of
a neurally-inspired model. This invariant is the
activation boost formed when two relatively
incongruous interpretations are simultaneously
active. A number of predictions of the model
are discussed in the context of verbal humour.
It is first shown that the size of the boost is a
monotonic function of the degree of support for
the competing interpretations. Next it is shown
that the size of the boost is inversely propor
tional to the time to resolve the punch
line. Finally, the interaction between the
degree of incongruity and the support for the
expected interpretation is discussed. The paper
concludes with a brief treatment of the
phenomenological implications of the model.

INTRODUCTION
W.C. Fields once said that trying to
understand humour is like trying to catch an
eel in a bathtub. Fields based his statement on
his comic intuitions, but hundreds of years of
philosophical speculation, and approximately
40 years of scientific research in this field
seemingly confirm his view. Certainly, it is
unlikely that humour can be confined to a finite
set of topics. Anything can be the subject of a
joke. This includes items that are intrinsically
peculiar such as kippers, chopped liver,
politicians, and Formica. But so-called serious
topics are not immune to humorous treatment.
Death, taxes, Haley’s comet, and trees have all
been made hilarious at some time.

It is natural, then, to look to the abstract
structure of the joke in the search for an
invariant aspect of humour. But here, too, we
seem to be at an impasse. Humour can be
found in the pithy saying, as in

(1) One more drink and I’ll be under the
host - Dorothy Parker.

by question and answer, as in the riddle

(2) Why did the chicken cross the Mobius
strip? To get to the same side.

by dialogue

(3) Mugger: Your money or your life
Jack Benny: <pause>
Mugger: Well?
Benny: <longer pause>
Mugger: Well?
Benny: I'm thinking it over.

and perhaps the most common structure, the
story

(4) President of synagogue at a meeting
gets up and says "I'm nothing." The others
protest, you're rich, you have a wonderful
family, your department store is thriving,
etc. "Yes," he says, but compared to God,
I am but a speck of dust." Vice-president
stands and they go through a similar
routine. Ditto the Secretary. Finally the
caretaker, who has been listening in the
corner, speaks up and makes the same
claim. The Vice-President turns to the
President and says, "Look who thinks he's
nothing."

This list is by no means exhaustive. In
addition to a variety of surface mechanisms, it
is likely that the appreciation of the full corpus
of jokes calls upon every known cognitive
mechanism - schemas, analogy, learning,
syntax, semantics, pragmatics, and so forth -
and possibly some yet to be discovered.

All is not lost, however. One intuition that
appears in various guises throughout the
history of this field is that humour somehow
involves the juxtaposition of incongruities
(Beattie, 1776; Kant, 1790; Koestler, 1964). In
other words, although humour theorists may
be reduced to glorified taxonomists when
talking about content and form, they can
justifiably return to being scientists proper
when discussing Attardo and Raskin's (1991) final level of humour processing, that of oppositional dynamics. One of the clearest expressions of this view can be found in the incongruity-resolution theory (Suls, 1972), which can be summarised by the following pseudo-code:

Form <predicted ending>
If <actual ending> = <predicted ending>
    then no surprise
Else
    If actual ending not resolvable
        then puzzlement
    Else humour.

In this conception, there are two necessary conditions for humour, an ending which does not match expectation, but nonetheless makes sense. The incongruity-resolution theory is an advance over its less precisely specified predecessors, but still suffers from the following problems:

a) It cannot explain humour without an explicit resolution (e.g., nonsense, and much non-verbal humour).
b) It says nothing about the bias towards sexual and aggressive content in humour.
c) It is intrinsically discrete and non-dynamic.
d) It does not explain why humour is pleasurable.

I have discussed problems a) and b) in Katz (1993). Problem c) arises because the two tests of the model are binary conditionals; i.e., are either met or not. In addition, there is a very weak sense of time in the model, but we know that timing is essential to humour. The primary purpose of this paper is to suggest that an alternative, neurally-inspired model potentially addresses problem c). In addition, the model will have implications for the resolution of problem d), which will be discussed in the final section.

THE NEURALLY-INSPIRED MODEL

The proposed model assumes that humour works by unifying two disjoint interpretations (the condition that they are disjoint will be relaxed in the next section). The unification occurs because during the transfer from the expected interpretation to the unexpected, but resolvable interpretation, the neural units representing the two interpretations are jointly active.

This process is illustrated in Figure 1. Initially, at t1, the context primes the expected interpretation, represented by unit A. For example, in (1), the expected completion is "under the table." This interpretation is incompatible with the unexpected ending, "under the host", represented by unit B. Thus, these two units are part of a winner-take-all network, characterised by excitation from the units to themselves, and a bi-directional inhibitory connection between the units.

This sub-network ensures that normally only one interpretation will be active at any given time. However, at time t2, the ending primes

![Figure 1](image-url)
the alternative interpretation, represented by unit B. In addition, this interpretation receives support from the context, i.e., the body of the joke before the ending, as well as any previous knowledge the listener can bring to bear on the situation. In (1), the context is Parker’s louche, inebriated state. Were it not to receive this inferential support, insufficient activation would be received by unit B to ensure the transition between the two interpretations. The context also continues to support the initial interpretation, represented by unit A, because of the strong expectation that this will be the actual ending. This joint support for units A and B overrides the lateral inhibition, and permits these units to be simultaneously active for a brief time, before unit B wins the competition at t3.

In summary, the model predicts that two mutually conflicting interpretations can be jointly entertained if the context strongly primes the expected interpretation while the incongruous interpretation is triggered by the ending, but also makes sense in the overall context of the joke and the situation. The induced transition is equivalent to the resolution step in the incongruity-resolution model. The transition results in an activation boost, illustrated in Figure 2, which details the activation traces and the sum of the traces for the two units under typical parameter conditions. By measuring the size of this boost, one can determine the extent to which the two interpretations have been temporarily unified.

In the following section, the boost size will be measured as a function of some of the parameters in the model.

**PREDICTIONS OF THE MODEL**

Because the model is precisely specified, it is in principle a simple matter to vary the key parameters to determine the degree of unity as revealed by the height and size of the activation boost, and presumably, therefore, the degree of humour. The parameters that may be varied are as follows:

a) the amount of activation sent to unit A,
b) the amount of activation sent to unit B,
c) the size of the lateral inhibition between units A and B,
d) the size of the recurrent excitatory connections for units A and B,
e) the thresholds of units A and B, and
f) the delay between activating units A and B.

g) the degree of overlap between the two interpretations (to model this, A and B are replaced by multiple units).

The next section treats a) and b), and what implications these manipulations have for verbal humour. The section which follows treats the cases where two parameters are jointly varied, a) and f), and a) and g). The final section will briefly discuss c) and e).

![Figure 2. The activation boost. The bottom graph illustrates typical activation traces for units A and B. The top graph shows the total activity in the network formed by summing these two graphs.](image)
SINGLE PARAMETER MANIPULATIONS

In this section, the effect of varying the supporting activation to units A and B will be studied. The activity delivered to unit A may be construed as the strength of support for the expected or default interpretation. In general, the greater this support, the greater the perceived humour will be.

This has been demonstrated in experiments in which subjects lift a hidden weight that differs markedly in heaviness from the previous weight. Facial expressions of humour and subjects' ratings of humour have been shown to be proportional to the number of times the box with the first weight is lifted (Deckers & Kizer, 1975). The more such encounters, the more they are fooled, and the greater their laughter. Presumably, this result is an effect of the build-up of expectation with repeated liftings.

The role of expectation can also be seen in verbal humour with the following sequence of manipulations of (2):

(2') Why did the chicken traverse the Mobius strip? To get to the same side.
(2'') Why did the turkey traverse the Mobius strip? To get to the same side.
(2'') Why did the giraffe gallop across the Mobius strip? To get to the same side.

The greater the difference between the actual question and the related question "Why did the chicken cross the road", the less the priming of the expected answer, "to get to the other side." Thus, (2) is funnier than (2'), which is funnier than (2''), which is funnier than (2'').

The left graph in Figure 3, which shows the size of the activation boost as a function of the support for unit A, shows that these observations are consistent with the predictions of the model. The boost rises in a monotonic and near-linear fashion as this parameter is increased.

It is also the case that the greater the strength of the resolution, i.e., the more sense the incongruous ending makes, the greater the funniness. Verbal humour without an explicit resolution, such as nonsense humour is consistently found to be less funny than humour that in which the punch line is resolvable (Staley & Dekks, 1995). Wicker et al. (1981) also found a significant correlation between subjects' rating of joke funniness and their ratings of the their ability to resolve the joke.

Persona, a key attribute of the successful comic, is an important factor in creating a believable resolution. Compare (1) with the following role change:

![Figure 3](image-url)

Figure 3. The activation boost as a function of the support for units A and B. The contour graph on the left shows the activation boost as a function for unit A, and that on the right as a function of the support for unit B.
(1') One more drink and I'll be under the hostess - Steve Martin.

Martin's persona is that of the everyday goof ball, if this is not a contradiction in terms. He rarely reverts to sexual humour, not because of prudery, but because this would be out of character. These facts combine to make (1') less likely than (1), therefore less resolvable, and therefore less funny.

The graph on the right of Figure 3 is consistent with these observations. The graph shows that the height of the boost rises in a monotonic fashion as the support for the incongruous interpretation is raised. That is, the more plausible the punch line, the greater the humour. Continuous predictions such as those shown in Figure 3 are a fundamental requirement of any theory of humour, because the human response is also not discrete. Further advantages of the proposed model are discussed in the next section.

**DUAL PARAMETER MANIPULATIONS**

The incongruity-resolution model and similar cognitive theories of humour, although not capable of continuous predictions, are at least consistent with the current model when expectation violation and resolution are both strong. But these models say little about factors such as timing, or the similarity between the two interpretations. In addition, these models are powerless to make predictions when these factors are altered in conjunction with other factors.

In this section, two dual parameter manipulations will be studied. The first involves varying the delay between the triggering of the first interpretation as represented by unit A, and the second interpretation as represented by unit B. The boost area is measured as a function of this delay, and the support for unit A; the support for unit B is held constant.

These results are shown in the graph to the left of Figure 4. The general trend is consistent with Goldstein's (1970) finding that humour is inversely proportional to the time to resolve the joke. But the graph reveals another interesting result. The rate of decline slows as the support for the expectation grows. The reason for this is that high support slows the decay in the activation of unit A.

![Graph](image)

Figure 4. Dual parameter manipulations. The left graph shows the activation boost area as a function of a delay between triggering units A and B, and the support for unit A. The right graph shows the activation boost height as a function of the similarity between A and B, and the support for unit A.
This result is consistent with the fact that humour that involves weak support for the expectation typically involves a rapid transition between the two interpretations. For example, slapstick almost always occurs at breakneck speed. This effect can often be enhanced by showing the relevant section at high speed. In these cases, the rapid alternation compensates for the fact that the initial interpretation perseverates for a short time only. These results await empirical verification in the case of verbal humour.

The graph on the right of Figure 4 shows the height of the activation boost as a function of the degree of incongruity between the interpretations and the support for unit A. The degree of incongruity was varied by changing the number of features that the two interpretation shared. For example, 25% incongruity means that there is a 3/4 overlap between the feature representation of the two concepts. The concepts were subjected to competitive learning (Rumelhart & Zipser, 1986), an unsupervised learning schema that forms classifications on the basis of similarity. They were trained in the context of other random concepts for verisimilitude. Five clusters of classification units were used, each cluster consisting of a winner-take-all network similar to that shown in Figure 1.

The effect on the transition from one interpretation to the next is best illustrated by reference to the inverted U-shaped curve formed in the low expectation case. In this instance, two factors determine the size of the boost. The first is the magnitude of the support the classification units, triggered by the first concept, receive from the second concept. This will be proportional to the number of shared features, because these features will remain active when the second concept is presented to the network, and will also be connected to the first concept. The second factor, which runs in the other direction, is the probability of a transition between two classification units. The greater the similarity between the concepts, the lower this probability. These two factors combine to produce the inverted-U, in which an optimum occurs when there is a possibility of a transition, but also some support for the first classification by the second concept.

Because completely disjoint concepts do not share any features, they do not produce a boost in the low expectation case. However, the situation changes when an expectation is added. In this case, the external support for the first interpretation compensates for the lack of feature support. If the external support is sufficiently strong, then the first interpretation will be almost completely active when the second interpretation becomes active.

This is illustrated in the graph in the high expectation case. In this situation, there is a monotonic rise in boost height as incongruity rises. These graphs show that it is possible to unify two relatively disjoint concepts, but only if there is strong support for the first when the second is triggered. This, of course, is the normal verbal humour mechanism, in which the joke context serves to create a strong misdirection towards the first interpretation, before the final (plausible) incongruity is presented in the punch line. Other implications of this graph are discussed in the next section.

COMIC CONSCIOUSNESS

Other predictions follow from the model. For example, topic salience can be shown to lower the effective thresholds of the corresponding units via an exposure effect (cf. Katz, 1995), thereby increasing the size of the boost. A similar effect can be achieved if the lateral inhibition, separating the two competing units, is lowered. In this case, however, this parameter corresponds to a trait rather than state variable, possibly bearing an inverse relation to Eysenck's P (Psychotic) factor, which is correlated with breadth of association (Eysenck, 1995).

Rather than pursuing these speculations in depth, I would like to conclude this paper by saying something about the model's implications for the phenomenology of humour. The study of humour must be first and foremost the study of a state of consciousness. No comedian would perform in front of an audience of zombies (although many have felt themselves to be in that situation). No one would listen to jokes if they did not generate pleasure. Therefore, any theory with pretensions to generality must address the issue of qualitative states.

The leverage for doing so in the current context comes from the identification of relatively high neural activity with pleasure. Elsewhere (Katz, 1994) I have defended this as
an appropriate means of gauging the hedonic response of a neural system. If this identity holds, then the current model reveals why humor produces such a heightened experience of pleasure. The unity of two relatively disjoint interpretations will potentially create twice as much activity in the system as is possible when only one interpretation is active.

As shown in the graph on the right of Figure 4, this contrasts with situations where items of moderate similarity are simply juxtaposed (the low expectation curve). Without the joke work that sets up the misdirection, the juxtaposition of highly incongruous concepts produces no boost at all, and no pleasure. But this does not mean that juxtaposition is not a useful mechanism. If the successive concepts share a moderate number of features, then a small boost can be achieved.

This, for example, is one of the main mechanisms of music, which consists of themes followed by variations on those themes (Katz, 1994). Thus, one would expect music to consist of less pleasurable episodes than humor. But because the boosts are relatively easy to achieve, there will be more of them. Humor, in contrast, will consist of a series of sharp but short pleasure jolts punctuated by relatively neutral affective periods where the next punch line is being set up. I believe this provides an accurate phenomenological account of the fundamental hedonic effects of listening to music and to humor. Brain imaging techniques may provide more direct confirmation of this view.

REFERENCES
WHY IS A RIDDLE NOT LIKE A METAPHOR?

E. Judith Weiner
Temple University
Computer and Information Sciences
Computer Building 306
1805 N. Broad Street
Philadelphia, PA 19122-6094
USA
e-mail: weiner@cis.temple.edu

ABSTRACT

Metaphors and riddles both are characterized by incongruous elements. Yet, riddles are funny and metaphors (often) are not. In order to highlight some specific types of incongruity which are conducive to humor, this paper offers an exploration of those features which these two forms have in common and those which they don't. The differential use of prototypicality in metaphors and riddles is also discussed.

INTRODUCTION¹

Many researchers have observed a connection between incongruity and humor (e.g., Navon 1988; Schultz 1976; Suls 1972, Raskin 1985, 1986, 1987). Their work has focused primarily on the role played by incongruity in general in humorous situations, for example, on the fact that humor can occur with incongruity which is later resolved. Other researchers, e.g., Paulos 1980, have dealt more specifically with the precise nature of some of the kinds of incongruity necessary for humor.

Our primary interest in past years has also been in discovering those specific kinds of rule violations which lead to humor (De Palma and Weiner 1990, 1992, Weiner and De Palma 1993, Weiner, in press). In this paper however, I am turning my attention more toward an understanding of forms featuring incongruity in which the result is not generally funny. Metaphors are forms which, while it is possible to find some that are humorous, generally they are not. Yet, like many types of humor, they are also based on incongruity and its resolution. Here, I will explore some of the pragmatic factors that are featured in metaphorical language and how these operate in riddles, where riddles are assumed to be funny² and metaphors, as a rule, not funny.

In selecting metaphors, I will often use examples like Ortony's familiar

1) Billboards are like warts.³

which fit the form

¹ I'd like to thank David Weiner for his useful comments on an earlier version of this manuscript.

² I am concerned also with riddles that are meant to be humorous, as opposed to, for example, those of the Riddle of the Sphinx variety.

³ For the purposes of this discussion, I have chosen not to distinguish metaphors and similes and will ignore the fairly subtle differences between them. I am also assuming the context, both linguistic and extralinguistic, to be as neutral as possible.
2) A's are (like) B's.

Riddles have been chosen as the domain of humor because it is possible to find examples of riddles which are formally simple (and thus potentially syntactically manageable) and because it is illustrative to try to relate them to formally simple metaphors such as 1). For example, I could make up a riddle which corresponds to 1):

3) Why are billboards like warts?

which is of the form:

4) Why is an A like a B?

**DOMAIN DISTANCE INCONGRUITY**

Metaphors are used in an attempt to improve communication, to make something that is difficult to express easier to express (Weiner 1984, 1985b). Poems abound with metaphorical descriptions of love and other elusive emotional concepts. In metaphors of form 2), if there is difficulty finding a literal way to describe A, one can produce a metaphor by selecting an appropriate B from a different domain.⁴

In their communicating function, metaphors are consistent with other aspects of language, which is what one might expect since metaphors are, without question, part of language. Yet there is an inevitable element of incongruity in metaphors since the A term (the *vehicle*) and the B term (the *topic*) of metaphors come from incongruous domains (Ortony 1979a, Weiner 1984, 1985a, 1987a, 1987b).

Riddles, on the other hand, are an attempt to prevent communication. The Gricean Cooperative Principle (Grice 1975) is violated deliberately, following instead a kind of “uncooperative principle.” Yet, in both riddles of form 4) and metaphors of form 2) there is a degree of incongruity between the A and the B terms. But whereas metaphors function by helping the hearer comprehend the inexpressible, the function of riddles is to confuse and prevent the hearer from understanding.⁵

The domain distance incongruity between A and B that occurs in metaphors is obvious in the billboard/wart metaphor. That it also occurs in riddles is illustrated in riddles like

5) Why is a harp like an elderly parent?

Both are unforgiving and hard to get in and out of cars.

in which there is no doubt that harps and elderly parents are in distant categorical domains from one another.

Generally speaking, categorical domain distance is difficult to measure although a few clear-cut cases do exist. Take, for example, the similarity statement

6) Dogs are like wolves

In an intuitive categorization scheme such as the following:

---

⁴ "Ortony's inexpressibility thesis (1975) deals with transferring from vehicle to topic "characteristics which are unnameable."

⁵ It appears that with riddles, one strives for as gross an incompatibility as possible between A and B. Interestingly, others have posited limits on this distance, notably Glora (1991). Obviously, more work needs to be done to reconcile these points of view. This should probably focus on the specific *kinds* of domain distance incongruity since it doesn't appear to be a simple matter of distance *per se.*
dogs and wolves are at a level of abstraction close
to Rosch's (1973) basic level categories and share
the same superordinate. They lack incongruity
because the relationship between them is so close,
*i.e.*, that of siblings. Such examples are potentially
Nor are similarity statements with domains of A
and B in a formal parent-child relationship
incongruous, *e.g.*, 7) John is a person.
8) Ice is a solid.
(Weiner 1984, 1985a, 1987b). These, of course,
are also literal.

A complete exploration of the domain
distance issue will not be presented here; it is
currently under investigation by myself and a co-
researcher. In any case, it seems clear from the
above discussion that domain distance incongruity
of the sort I have presented here is not sufficient
for humor.

**SALIENCE**

I use the term *salience* to refer to the
"prominence of a particular attribute with respect
to a concept to which it does or could apply" (Ortony 1979a: 162) with Ortony's later
(1979b:191) substitution of "predicates" for
"attributes." A predicate is anything which can be
said about a concept, including commonly-held
beliefs (*e.g.*, snakes are generally disliked) as well
those things traditionally referred to as attributes
(*e.g.*, chairs have legs). 6

In Weiner 1984, I investigated Ortony's
(1979a) claim that in order to have a metaphor, it
is necessary to have predicates of high salience in
the B term predicated of the A term in which they
are of low salience. In a metaphor like 1),

1) Billboards are like warts.

the A term (the topic) is billboards and the B term
(the vehicle) is warts. The ground is what they
have in common: 1) appears to be prediciating of
billboards the ugliness and prominence of warts.

In Weiner 1984, I also demonstrated that
metaphors don't always require low salient
predicates in A. In a metaphor like

9) Mary's cheeks are like apples

---

6 A more detailed discussion of salience is given in
Weiner 1985b.
meaning that Mary's cheeks are red and round, if Mary's cheeks are in fact like that, they are probably one of her salient features.

In speculating about a possible response to a riddle like

10) Why are Mary's cheeks like apples?

responding with highly salient predicates of apples is somehow unfunny. Rather, to dream up a funny response, one way would be to ignore high salient predicates of the B term as in the following example:

11) How would you fit four elephants in a VW bug?
Two in the front seat and two in the back.

For this riddle to work, it is necessary to ignore one of the most salient features of elephants, i.e., their size. This example is unusual in its simplicity, a salience violation. Other riddles are more complex in that many mechanisms are present at once to produce the humorous effect. But salience violations can be a factor in the production of humor (Weiner, in press).

PROTOTYPICALITY

In metaphors of form 2), prototypicality functions in a number of ways:

1. The B term is more prototypical of the relevant predicates than is the A term, e.g., in 1), warts are more prototypical of the predicate "ugliness" than are billboards (Tversky 1977, Weiner 1984).9

2. The B term is the epitome of predicates that are true of A, i.e., B is the prototypical representative of these predicates. A better example cannot readily be found (Tversky 1977, Weiner 1984).

3. The A and B terms refer to the prototypical representative of its category. (Weiner 1985a). For example, in

1) Billboards are like warts.

the sentence refers to prototypically ugly billboards as opposed to ones which might might not be ugly. It also refers to prototypical warts.

It is useful to compare and contrast the ways in which these uses of prototypicality function in riddles:

1. Unlike in metaphors, in riddles, the B term need not be more prototypical of the relevant predicates than the A term. In

5) Why is a harp like an elderly parent?
Both are unforgiving and hard to get in and out of cars.

both harp and elderly parent are equally prototypical of the relevant predicates.

2. As in metaphors, in riddles, B will still represent the epitome of the relevant predicates. The B terms function differently, though, as we've already observed in the discussion of salience. In riddles, this may also be true of the A term.

7 Navon (1984:213) writes that jokes must ignore "a most essential piece of information which, nonetheless, is not explicitly stated." The not explicitly stated requirement would rule out ignoring elephants and giving a (very unfunny) response like "They weren't elephants. They were mice." (Weiner, in press)

8 My use of the terms prototype and prototypical is somewhat different from the ways they have been used by others, e.g., Rosch 1973, 1978, Rosch and Mervis 1975.

9 Tversky (1977) observed that the B term is the more prototypical of the two in literal similarity statements. Ortony (1979a) has made similar claims for certain types of metaphorical statements.
3. In riddles as in metaphors, the A and B terms refer to prototypical representatives of their respective categories. In 5) then, the references are to prototypical harps and to prototypical parents, not to, let's say, miniature harps.

**PREDICATE INEQUALITY**

In metaphors, the ground consists of those predicates common to both the A and the B terms. Generally, however, there is an inexact qualitative or quantitative match between them as in

12) Jane's eyes are like stars.

While in the case of metaphorical statements such as 12), the inexactness of the match enhances the metaphorical effect (Ortony 1979a, Weiner 1984), it is also possible to have metaphorical statements in which the predicates match exactly. In

13) The initial spectrum of fluctuations thereby gave rise to gigantic, irregular clouds of gas that resembled flattened pancakes.

(Scientific American, p. 72)

there is the possibility of an exact match of predicates in A and B. Yet the statement appears to be at least mildly metaphorical.¹⁰ The reason for this is due simply to the incongruity produced by the domain distance between the A and the B terms.

That a statement can be more or less metaphorical, that there exists a continuum of metaphoricity, in the sense that 13) is less metaphorical than 12) should not be surprising. The notion of graded structure has significant generality within categories of all sorts (Barsalou, 1983) and the category metaphor appears to be no exception; 13) is a less prototypical metaphor than 12). In both cases, though, there is considerable domain distance incongruity; both can be considered metaphorical to a greater or lesser extent.

In riddles of form 4), predicate inequality is one way of introducing incongruity but the incongruity of a simple mismatch of predicates (along with domain distance) is not sufficient to produce humor. As demonstrated above, 12) certainly is not funny. But other forms of incongruity in predicates can produce humor.

The first part of the response to 5)

5a) Both are unforgiving

is an extreme example of predicate inequality -- the predicates aren't even the same, since the use of unforgiving is polysemous in this case. This is an example of a pun, a common riddling device but one not usually found in metaphors.

**BEYOND METAPHORS: INCONGRUITY AND HUMOR**

While I have stressed in this paper that not all forms of resolved incongruity are funny, e.g., domain distance incongruity and ordinary (as opposed to punny) predicate inequality, I have also pointed out that the kind of incongruity produced by violations of salience can be funny. There are certainly other kinds of resolved incongruity that also produce humor. I will discuss two of those here: accessibility violations and parallelism violations.

**Accessibility violations**

There are categories of humor in which lexical ambiguity is a major factor. Take, for example, the old riddle:

14) What has a mouth but does not eat.

A river.

---

¹⁰ Elsewhere, e.g., Weiner 1987a, I referred to these types of statements as literal statements with a "metaphorical flavor" (p. 522).
What makes this riddle "work" is that the more accessible meaning of *mou\ th* is the animate one. This is the one which the hearer presumably thinks of first. But the riddle response is possible only if this highly accessible meaning is "ignored" and a less accessible one utilized.\(^{11}\)

In earlier work (De Palma and Weiner 1990, 1992, Weiner and De Palma 1993, Weiner, in press), we describe how this accessibility continuum might function in normal speech in order to be misused to humorous effect in riddles such as 14). As mentioned above, it has been demonstrated that seemingly unrelated types of categories share similar features, e.g., graded structure, the characteristic that causes some members of a category to be judged as more typical of that category than others. Graded structure is not limited to traditional taxonomic categories (e.g., Rips et al., 1973; Rosch, 1973) but may actually be a universal property of categories (Barsalou 1987), having been observed in such disparate categories as formal categories (e.g., odd numbers, squares) (Armstrong et al. 1983), linguistic categories (e.g., phones, phonemes, syntactic categories (Lakoff, 1986), even ad hoc categories formed to achieve a novel goal (Barsalou 1983, Murphy and Medin, 1985).

In the same spirit, we have posited categories (sometimes *ad hoc*) of homophonous/polylemous words/phrases (e.g., *mou\ th*) on which an accessibility continuum based on prototypicality of meaning is imposed. (De Palma and Weiner 1990, 1992, Weiner, in press).\(^{12}\)

For the phonemic category representing the word *mou\ th* then, there are members with different accessibility values. Out of context, the most accessible meaning is accessed first (causing the riddlee hearing examples like 14) to miss the "correct" answer, the one which violates the accessibility continuum.\(^{13}\)

**Parallelism violations**

A number of researchers have recognized parallelism as a force operating in language (Kuno 1974, Prince 1981, Weiner and Labov 1983). We have defined parallelism as the "tendency to remain in the same syntactic, semantic, pragmatic or cognitive track unless there is pressure to change to an alternative one" (Weiner and De Palma 1993, p. 190). The opposition to this force can be used as a mechanism in the production of humor (De Palma and Weiner 1990, Weiner and De Palma 1993, Weiner, in press, Giora 1991).\(^{14}\)

As an example, consider another old riddle:

15) What is black and white and red (read) all over?
A newspaper.

There is pressure from parallelism to parse the three conjuncts into the same syntactic category.

---

\(^{11}\) I am ignoring the issue of whether or not there is historically a metaphorical connection between these meanings since there are numerous examples of riddles which function similarly and in which there is no such connection (De Palma and Weiner 1992).

\(^{12}\) This has taken subtly different forms in other papers but the overall thrust is the same. Also, the terminology which I employ here is somewhat different from that of earlier work. For example,

\(^{13}\) Giora (1991) has written similarly about the role of accessibility in humor. In her discussion of the notions of accessibility and prototypicality she states that a prototypical member of a category is the most accessible while those that are "cognitively distant from the prototypical members" are "hardly accessible." (p. 473)

\(^{14}\) Giora refers to this tendency as the use of our "one-track-mind." (p. 481)
This is made even stronger by the fact that they can also be interpreted as belonging to the same conceptual category (colors). The riddle answer is obtained by resisting the force.

The following is a more sophisticated riddle involving multiple parallelism violations:

16) Why is the snowman smiling?
   Because he heard that the snowblower was coming.

Why is this funny? Although the personification of the snowman sets up an expectation that he will experience some human pleasure, the word snowblower returns (briefly) to a snow context before a return to the human pleasure context. The snow context meaning is irreconcilable with the query so the other is necessary to reconcile the conflict. The sudden shifts in contexts, multiple violations of parallelism, make it funny.

FUTURE WORK

Domain incongruity in metaphors is not usually funny. The example presented above,

9) Mary’s cheeks are like apples.

isn’t at all funny, in spite of the fact that there is considerable domain distance between Mary’s cheeks and apples.

But sometimes a metaphor can be at least mildly funny:

17) John’s nose is like a carrot. 15

Accounting for this difference lies clearly outside the domain distance incongruity issue since the distance between Mary’s cheeks and apples cannot be much different than that between John’s nose and a carrot. Neither does predicate inequality explain the difference since this also is pretty much the same in both examples.

An explanation might lie in the fact that people with extremely long (and maybe even orange) noses are usually considered "funny-looking" whereas those with extremely round, red cheeks are usually considered attractive. Perhaps it is an issue of socially attractiveness/acceptibility. 16 The kind of incongruity then, seems to be between the assertion of a physical condition and some social convention. 17

CONCLUSIONS

Humor can be produced by “misusing” our pragmatic knowledge in various ways often by creating certain types of incongruity. Here, I have demonstrated the ways in which salience, accessibility and parallelism can be violated to funny ends in riddles but also the fact that domain distance incongruity and the kind of incongruity found in ordinary predicate inequality are not sufficient to produce humor. For this reason, metaphors are not as a rule funny.

We need to look more closely at the nature of incongruity and its many manifestations. Incongruity may not tell the whole story though. We also need to investigate other pragmatic factors which may have little to do with incongruity, as I’ve tried to do with prototypicality in this paper, to derive some insights into the mechanisms operating in the production of humor.

15 In an informal survey which I conducted, people’s responses to this example ranged from “That’s hilarious” to “No, that’s not at all funny.” The people for whom it was funny tended to visualize the carrot superimposed on John’s face; the others didn’t.

16 I am grateful to Nancy Weissman (personal communication) for this insight.

17 This explanation is being put forth tentatively. More work needs to be done to determine with more confidence the factors operating here.
REFERENCES


Rosch, E. and C.B. Mervis 1975 "Family Resemblances: Studies in the Internal Structure of
Categories”, *Cognitive Psychology* 7: 573-605.


No Laughing Matter:
The Cognitive Structure of Humour, Metaphor and Creativity

Tony Veale,
School of Computer Applications, Dublin City University, Ireland.
Email: Tonyv@compapp.dcu.ie

1. Introduction and Foundations
The tale is told that Ludwig Wittgenstein, dismantler of logical atomism and a beloved benefactor of modern cognitive science, remarked late in life that it was a long-time wish of his to write a philosophical work composed entirely of jokes. When pressed as to the reasons why he had never pursued this goal, Wittgenstein supposedly replied that he simply didn’t have the sense of humour for the task. This apocryphal anecdote, whether true or invented, does suggest that the nature of humour has some philosophical content worthy of investigation. Accepting this challenge, this paper looks to an analysis of humour and metaphor to yield some valuable insights into the nature of creativity and the human conceptual architecture.

This paper attempts to accommodate humour within a broader computational framework of metaphor interpretation. Its main contention is the belief that humour and metaphor are both folds of a larger creative tapestry, wherein humour represents a perversion of an underlying metaphoric mapping. This introduction shall sketch some basic strategies by which humourists achieve this perversion, while providing a ideological/philosophical map for the unified model of humour and metaphor that will be sketched in the sections to follow:

Apposition: The juxtaposition of convention and unorthodoxy, or straight-man and funny-man, which is central to the operation of humour, is similar to that apposition achieved by metaphor and analogy, in which two disparate domains become co-active in a structurally coherent manner. To further appreciate the relation between metaphor and humour, consider that humour extends in a continuum from the humble pun, a simple word replacement based on sound similarity, to intellectual humour, as practiced by the likes of Umberto Eco, which is the clever extension and elaboration of a productive metaphor. At the level of intellectual humour, sheer creativity blurs the distinction between the clever and the funny.

But how central is the role of metaphor in human cognition? Romantic philosophers such as Rousseau and Nietzsche have granted metaphor pride of place amongst the human creative faculties, whilst more modern authors, such as the cognitive linguists Lakoff & Johnson (1980), argue convincingly that metaphor is the major architect of our conceptual systems. Metaphor provides the cognitive apparatus that allows us to view one concept through the conceptual lens of another, thereby throwing light on both. In effect, metaphor reshapes the conceptual landscape that structures, and is structured by, our world experience, often warping this conceptual space to bring two concepts closer together. For instance, to describe man as a wolf is to highlight not only the stereotypical lupine characteristics of mankind (e.g., ruthless, lonely predators), but also to accentuate the anthropomorphism inherent in wolves (e.g., they are family animals which kill prey to make a living, etc.). It should be untendentious then to suppose that the central cognitive role of metaphor -- the ability to relate and reconcile two diverse situations -- is also widely exploited in the creation of humour.

Analogy: Metaphor and analogy are cognitive mechanisms (or perhaps, different facets of the same mechanism) which establish cross-domain mappings between a tenor concept (i.e., the
metaphorized subject) and a vehicle concept (i.e., the metaphorizing object). For example, when comprehending the metaphor "Set theory is the religion of mathematicians", cross-domain mappings are established not only between set theory and religion, but also between mathematicians and priests, and possibly between articles and faith, lectures and sermons, theories and prayers, text books and bibles, and so on. This demonstrates clearly the essence of good metaphor, for this contrived latticework of mappings allows a rather dry domain (mathematics) to wear the conceptual clothes of another (religion); the abstract world of maths is thus imbued with all the ecclesiastical flavour and pomp of religion, and visually-bland entities such as mathematicians are given colour by the metaphor (one almost imagines them in cassocks performing a holy ritual with a compass in one hand and a slide-rule in the other). But this mapping lattice provides more than mental imagery - each mapping may also provide the basis for further inference. For example, the AXIOM: ARTICLE-OF-FAITH mapping opens the door to considering God as blind faith as being established elements of the mathematical endeavour; this inference contradicts the prototypical view of the domain, contributing to the tension of the metaphor. This tension prompts the reader to think in novel ways, as he/she attempts to reconcile the domains of the metaphor. In this sense, good metaphor demands that the reader, as well as the speaker, exhibit creativity.

For a less obvious use of metaphor in humour, consider Groucho Marx’s apocryphal quip to a woman, who, when pressed as to why she had so many children, explained "I love my husband very much". Groucho is said to have replied "Lady, I love my cigar too, but I take it out of my mouth once in a while". The core of this joke is a classic Freudian metaphor which we leave to the reader to analyse further.

Truth (or "It’s funny because it’s true!"): Brevity has long been a perceived attribute of wit, as advocated by a diversity of sources, ranging from Shakespeare’s Polonius to the more academic Sigmund Freud (1938). However, conciseness might more accurately be considered a signature of the cleverness of a witicism, inasmuch as a good epigrammatic metaphor can be highly productive, relative to its lexical cost; this point is perhaps best stated by a past master of both:

"Wit has truth in it; wisecracking is simply calisthenics with words"

[Dorothy Parker, in the Paris Review]

A striking metaphor is appreciated mainly for the insights -- that inspired AHA! feeling -- it brings to the hearer. Novelty in itself is not enough, the metaphor must be seen to be apt, a true description of its subject, if it is not to be dismissed as a failure. Likewise, humour too must contain truth, even if this truth is merely an acknowledgment that a situation is open to an alternative, though less conventional, interpretation. The issue of truth is closely related to that of cleverness, for what is it to be clever other than to see a veiled aspect of the world as it really is, or should be?

King Kong is [...] amusing nonsense punctuated by such reflections as why, if the natives wanted to keep the monster on the other side of the wall, they built a door big enough to let him through.

[James Agee, film critic, 1933]

Humour often approaches truth from an oblique angle, wherein the hearer might reach an accepted truth (or belief) through a bizarre or incorrect chain of reasoning. For instance, the Steve Wright (SW) gag "There was a power outage at a department store yesterday and twenty people were trapped on the escalator" is a product of reasoning from a bad analogy (escalators = elevator), but it leads the hearer to a stereotypical (if unfair) Eurocentric conception of the average middle American, that of an overweight and lazy mail-dweller.

The American poet laureate and Russian exile Joseph Brodsky was once asked whether he felt any bitterness toward the Soviet Union for subjecting him to two years in a labour camp for dissent writing; Brodsky replied that, on the contrary, by Soviet standards the experience was “positively homeopathic”. Looking past the immediate ironic humour of this throwaway remark, it is possible to discern a deeper mapping between the concepts of intermittent and homeopathy. Consider that upon Brodsky’s release he was offered voluntary exile in America, which he readily accepted. Compare this with homeopathy, the dubious practice of
subjecting a patient to a diluted form of an ailment in the hopes of inducing a cure and expelling unhealthy agents from the body; a brief spell of imprisonment clearly induced a similar desire in Brodsky to leave the Russian body politic. While the remark is witty at a surface level, it reverberates all the more at a conceptual level because of an essential truth.

**Pain:** It is an undeniable irony regarding the nature of humour is that while jokes are uttered to bring laughter to oneself and to one's friends, humour invariably involves pain, for it is ruthless in its demands for a victim. In this sense, humour is directional -- it establishes an ameliorative/pejorative perspective shift between *funny-man* and *straight-man*. Humour without a well-defined straight-man often fails to work; for instance, the humour of Monty Python might easily seem directionless and unfunny. Sated more intuitively, humour is most often effective when it exhibits cleverness to achieve pejorative goals. For example, the insult "*Your family tree must have Dutch-elm disease*" is a clever extension of the Ancestry: Tree metaphor which exploits the pejorative connotations of the concept Disease to impart a disparaging view of the victim's genetic heritage. Dispositional accounts of humour which rely on the debasement of a straight-man date as far back as Aristotle's *Poetics* (see Hutton 1982) and underlie the disparagement model of humour analysis (see Zillmann 1983).

Of course, in the case of self-deprecating humour, the speaker is both straight-man and funny-man, and the ameliorative/pejorative shift is established between the comic's view of himself and the hearer's view of the comic (or more mind-bendingly, the hearer's view of the speaker's view of himself). For example, "*My gene-pool needs regular doses of chlorine*" extends the gene-pool metaphor into the realm of swimming pools, whereupon a derisive view of the comic's family (and by implication, the comic himself) is conveyed. Humour is most disarmingly potent when it exploits self-deprecation, or simple modesty, in this fashion. For example, when after winning the Nobel prize for literature Joseph Brodsky proclaimed "*It is a giant step for me but a small step for mankind*"; a deep sense of modesty here serves to lift an unconventional word ordering from the realm of simplistic spoonerism to the status of lasting humour.

**Prototypicality:** A standard strategy in the comic's repertoire for enforcing a literal context (i.e., establishing a straight-man) is the exploitation of polysemy in language. As Lakoff (1987) amongst others has noted, concepts and their associated lexical forms are probably best considered to form radial categories, conceptual structures in which the core meaning or prototypical usage of a word-concept resides at the hub, whilst other, less representative meanings populate the area surrounding this hub, at distances concomitant with their prototypicality. Within this framework, humour can be seen as the evocation of prototypical meaning with the subsequent dashing of hearer expectations, by the evocation of a secondary, altogether less-representational interpretation. In this sense, humour is often an excursion to the outer-limits of a radial category, running against the hearer's inertia toward the prototypical hub.

Lakoff (1987) cites a classic case of prototypicality in our conceptualization of the bird category. Robins and sparrows are examples *par excellence* of the bird class, but penguins and ostriches, while still valid examples of the class, are much less representative. A story related about a bird, then, in which the particular species is not explicitly mentioned, is likely to evoke an image of the prototypical bird (a robin) rather than a category outlier like an ostrich, provided the context is open to this interpretation. So, a comedian might exploit this tendency toward the categorial centre in a joke like the following: "*A month ago I found an injured bird lying in my garden. Being the animal-lover I am, I took it inside, tended its wounds, cared for it night and day, and gave it all the right food. Then, when I felt it was time for this bird of mine to rejoin his fellows in the wild, I opened my window and joyously threw the newly invigorated fellow into the air outside. Unfortunately, he fell four stories and went splat on the pavement below. God, how I had come to love that penguin."

Prototypicality also underlies the interpretations we give to polysemous words, and comedians frequently exploit this form of lexical expectation to humourous effect.
Consider the following apocryphal Q&A extract from a court transcript: "Lawyer: Are you sexually active? Witness: No, I just lie there." The humour content here clearly relies on the polysemy of the word "active", which in the context of the word "sexually" should suggest the notion of non-celibacy. Likewise, the SW gag "I went into a general store but they wouldn't let me buy anything specifically" relies upon the hearer incorrectly employing the broad-scope interpretation of "general", but SW dashes this lexical expectation by following up with the non-specific meaning via clever antonymy. Finally, consider the SW gag "I saw a restaurant with a sign that said We serve breakfast at any time, so I went in and ordered French toast during the renaissance". This joke exploits the social ambiguity inherent in the word "time", which can refer to local or historical temporality. A model of humour must therefore allow word meanings to be structured in such a way that certain readings are a priori more prototypical than others, but when given enough semantic priming (such as from "lie", "anything specifically" or "renaissance"), secondary readings can come to the fore.

Prototypicality also defines convention, and is thus central to our conception of expected and unexpected patterns of behaviour, another essential element of humour appreciation. The boss who says to his secretary "Will you ever forget the wonderful weekend we spent in Paris" surely does not expect to hear "I will for £2000", and for this reason (in part) it is funny. A misunderstanding of the communicative goals of the speaker, combined with a willful violation of the contextual expectations of the hearer, often lies at the heart of creative and humourous language use.

Transfer: If the mapping lattice established by a metaphor supports further inference, in what situations are such inferences actually mandated by the metaphor itself? Consider, for example, the classic Time-Life metaphor that views the Kennedy administration as Camelot; this metaphor obviously intends the hearer to infer that the early sixties was a magical time of promise and child-like wonder. However, the originators of the metaphor would have hardly endorsed the inference that RFK had a torrid affair with Jackie Kennedy (though Gore Vidal has since claimed as much), despite the fact that this a valid transfer given their rather more innocent mappings JFK: Arthur, RFK: Lancelot, and Jackie: Guinevere.

This inference is an example of a bad transfer or bad adaptation. Given the mapping nexus of the metaphor -- the core set of valid mappings such Camelot: Whitehouse and JFK: Arthur -- the perverse hearer simply over-extends the mapping to produce evermore tenuous inferences regarding the first lady's infidelity. But what criteria does one use to tell when a metaphor is being stretched beyond its operable safety limits? We get our first inkling of farce when the inferences and mappings produced in this red zone begin to work against the communicative goals of the metaphor as a whole. For example, the goals of the Time-Life metaphor are to imbue the Kennedy era with a sense of magic, purpose and innocent wonder (a classic media goal), but any inferences involving extra-marital affairs in the oval-office only tarnish this aura with the grubbiness of betrayal and infidelity. In essence, then, a metaphor interpretation must possess not only structural coherence (i.e., contain systematic mappings) and aesthetic coherence (i.e., the look and feel of the metaphor - in this example, a magical tone must be maintained), but also a pragmatic coherence, whereby the communicative goals of the speaker and the contextual expectations of the hearer serve to limit and restrain the scope of the interpretation.

To compose a general principle, then, an analogy can form the basis of humour when any reasoning process performed using the analogy produces results which are not considered well-formed, that is, (i) contrary to good sense (ii) contrary to intuitions about physical laws; (iii) contrary to established convention; or (iv) contrary to the expected goals of the metaphor speech-act (such as the Camelot scenario above). The comic Steve Wright is a master of the broken analogy form, as witnessed by the following instances of the above criteria: (i) "There's a pizza place near where I live that sells only slices. In the back you can see a guy tossing a triangle in the air." The core analogy here, between whole circular pizzas and pizza wedges, produces a reasoning chain which defies common sense; (ii) "I went to the hardware store and bought some used paint. It was in the shape of a house." While used
wood/metal etc. often retains the shape of its former usage, it defies physical sense to assume that, even if paint could be recycled, it too could retain its former shape; and (iii) "I have an answering machine in my car. It says, I'm home right now, but leave a message and I'll call when I'm out". Interestingly, the core analogy (car phones = house phones) and the resulting reasoning chain are both valid and sensible, but the outcome contradicts convention so squarely as to be jarringly odd.

**Re-Representation:** Analogical and metaphoric transfer is in essence a deep change of representation, one which causes the valid reasoning processes of the original domain to yield unconventional, bizarre or downright anomalous results in the new domain. For example, the gag "Y'know, my mother-in-law has a three-figure IQ -- that is, if you count in Roman numerals!", is funny precisely because it reinforces the conventional decimal numbering scheme in the hearer's mind (through the use of "three-figure" and "IQ"). But the sudden re-representation effected in the punch-line forces the hearer to re-evaluate the benefits of a three-figure IQ in this new scheme, and generate the opposite inference to that offered by the gag build-up.

**Creativity:** The most unifying trait of humour and metaphor is that both draw so heavily upon the wellsprings of human creativity. Metaphor is striking when it offers a novel perspective on a familiar object or experience; likewise, humour is perceived as funny when familiar scenarios are perverted to generate unexpected outcomes, and perhaps yield unorthodox observations upon the mundane. Both phenomena rely heavily upon their ability to surprise the hearer; metaphor does this by showing that surface dissimilarities often conceal deep conceptual similarities, while humour heightens the element of surprise by deliberately misleading the hearer along paths of convention which unexpectedly go astray. Metaphor and humour depend then on a creative speaker who not only possesses a unique view of a particular situation, but the communicative skills to convey this view to the hearer in such a way that a tension is created between the hearer's expectations and the speaker's goals.

*Idea rose in clouds. I felt them collide until pairs interlocked, so to speak, making stable combinations.*

[Henri Poincare, 1880]

Poincaré's description of the creative process, which most notably guided him toward a proof regarding the solution of general algebraic equations, refers to such cognitive events as ideas emerging from the contextual background, concepts colliding with each other, merging together, and interlocking to form stable and coherent patterns. This gels with our own discussion thusfar which has emphasized the role of concept juxtaposition and apposition, leading to analogical mapping between domains, and extended reasoning based upon these mappings. The creative process is viewed as acting essentially bottom-up -- individual cross-domain mappings arise due to local criteria of similarity and suitability, and combine to form larger mapping lattices which support the transfer of mental imagery and the generation of further inferences.

It is not our intention to bandy such colourful terms as **analogy, juxtaposition and mapping lattices** in lieu of a genuine model of creativity, metaphor and humour; such terms certainly have their role in motivating the discussion, and providing the hooks whereby theories can appeal to intuition, but will inevitably ring hollow if they lack a core computational model to describe. To support our discussion, then, this paper reports on a computer model of figularity, named Sapper, which builds upon Poincaré's insights to demonstrate how metaphors can be interpreted in mechanical/formal terms without doing violence to the ideals of human creativity, and in turn, creative humour.

2. **Sapper - A Network Model of Metaphor and Memory**

The Sapper framework, as described in Veale & Keane (1993), is a hybrid symbolic/connectionist model, embodying a basic philosophy which views the interpretation of novel metaphor as a process of connectionist **bridge-building**. From the Sapper perspective, metaphor comprehension involves the construction (or more accurately, the *awakening*) of new cross-domain linkages, which serve as bridges to bind the **analog-pairs** established by the metaphor. The novelty of a
metaphor may be measured by the extent to which it adds to the structure of the network, as it is accommodated by the system. This philosophy thus views metaphor as a dynamic, constructive, conceptual phenomenon, which evokes a response in a reactive fashion from an adaptive and accommodating knowledge-base.

A Sapper network is essentially a localist graph in which nodes represent concepts, and where arcs represent semantic relations between these concept nodes. Two significant architectural features arise out of a localist rather than distributed architecture: (i) knowledge isomorphism - the structure of a localist network directly mirrors the semantic composition of the knowledge represented therein; and (ii) knowledge parallelism - in a localist network any number of different concepts (i.e., noces) may be co-active simultaneously. While this first feature is simply a matter of design convenience, the second is arguably central to the operation of metaphor, the heart of which is surely the apt juxtaposition of different concepts (see Harnad 1982). Sapper is thus a computational model of conceptual juxtaposition, one which generates a mapping theory to resolve, or explain away, the initial incongruity of a concept pairing. From a humour perspective, Sapper's resolution of such conceptual juxtapositions corresponds well with the logical framework of Paulos (1977), while the juxtapositions themselves may be seen as the metaphoric equivalents of the identity synergies of Apter & Smith (1977), who work within the framework of the psychology of reversals.

A network model such as Sapper must support the functionality demanded by the analysis of humour and metaphor outlined in section one. More specifically, such a network must: (i) support the organization of different concepts via radial categories, that is, provide a weighting mechanism whereby prototypicality of meaning can be represented; (ii) support the determination of ameliorative/pejorative shift between different concepts (e.g., is X more insulting than Y?), and allow different this determination to occur in different belief environments (e.g., should I be insulted by this?); and (iii) allow creative reconciliation of different domains to occur, by inferring cross-domain linkages, or bridges, where previously none existed (e.g., AHA! if generals are surgeons, then an operating theatre can be seen as a battlefield).

New cross-domain bridges are laid down along dormant inter-concept connections, which have been established a priori by the memory management component of the model. These dormant connections provide the possible routes along which creativity can arise. Such dormant network linkages represent merely plausible, rather than fully established, semantic relations, and are thus not operative carriers of activation. The connectionist component of Sapper is responsible for the controlled propagation of activation energy, or zorch (see Hendler 1989), throughout the network, as it flows from the matriarch concept nodes as evoked by the metaphor. The matriarch concepts of a metaphor are known as the tenor, or metaphorized subject, and the vehicle, or metaphorizing object. The role of the connectionist component, then, is the awakening of dormant linkages, under suitable conditions, into active conceptual bridges that link the domains of the tenor and vehicle concepts.

At this juncture, the workings of the Sapper model are best illustrated with an example; consider then the metaphor "Surgeons are butchers." Sapper produces the cross-domain mappings of Figure 1 in response to this concept juxtaposition:

\[
\begin{array}{ll}
[.86] & \text{If Butcher is like Surgeon} \\
[.25] & \text{Then Abattoir is like Operating-Theatre} \\
[.75] & \text{and Meat is like Human-Flesh} \\
[.94] & \text{and Cleaver is like Scalpel} \\
[.98] & \text{and Carcass is like Corpse} \\
[.95] & \text{and Slaughter is like Surgery} \\
\end{array}
\]

\text{Figure 1: Sapper Interpretation of the metaphor/insult: Surgeons are Butchers.}

The number specified in square brackets to the left of each mapping is a numeric measure, between -1 and +1, of the perceived similarity of the related concepts after the metaphor has been comprehended. This measure combines a metric for both the literality similarity of the related items (e.g., Cleavers and Scalps are sharp and metallic), and the higher-order similarity that is now perceived between the two (e.g., the mapping of Scalpel to Cleaver supports the second-order mapping of Slaughter to Surgery, which supports the mapping of Abattoir to Operating-Theatre).
The Sapper rule component employs two distinct constructor rules to yield dormant conceptual bridges: (i) The Triangulation rule is invoked whenever two concept nodes share a common association or superclass, as is the case with the concepts SURGEON: BUTCHER & BLOOD, MEAT: HUMAN-FLESH & FLESH, SURGERY: SLAUGHTER & DEATH, OPERATING-THEATRE: ABATTOIR & LOCATION, and SCALPEL: CLEAVER & SHARP, laying down dormant linkages between the schemata BUTCHER and SURGEON, HUMAN-FLESH and MEAT, CLEAVER and SCALPEL, ABATTOIR and OPERATING-THEATRE, and SLAUGHTER and SURGERY; while (ii) The Squaring rule, which is a second-order constructor inasmuch as it may act upon the linkages laid down by the triangulation rule, is used to reinforce the bridges SURGEON: BUTCHER & MEAT: HUMAN-FLESH, and SURGERY: SLAUGHTER & SCALPEL: CLEAVER. The Squaring rule essentially builds bridges upon bridges, each new linkage extending the inter-domain reach of the last; in this way Sapper accounts for the phenomenon of domain incongruence (see Tourangeau & Sternberg 1981).

Following the construction stage of comprehension, the connectionist component proceeds to propagate activation from the matriarch nodes of the metaphor, SURGEON (tenor) and BUTCHER (vehicle), causing the dormant linkages between SURGEON: BUTCHER, MEAT: HUMAN-FLESH, SURGERY: SLAUGHTER, SCALPEL: CLEAVER, and OPERATING-THEATRE: ABATTOIR to be awakened as full, activation-carrying bridges. A dormant linkage is awakened in Sapper whenever it serves as channel for competitive activation waves originating at different matriarch nodes, that is, whenever it is seen as constituting a cross-domain bridge between the domains of tenor and vehicle.

The opening of these bridges allows activation to flow freely between the tenor and vehicle domains, altering the activation dynamics of the network in such a way that the tenor actually interacts with the vehicle at a conceptual level. The activation patterns of BUTCHER, SLAUGHTER, MEAT and CLEAVER interact with those of SURGEON, SURGERY, HUMAN-FLESH and SCALPEL to produce a response to the metaphor. Overall, a Surgeon is seen, through the lens of the metaphor, to be an altogether less skillful and precise tradesman, performing surgery which is akin to the slaughter of innocents, amidst the pain and screams of fear wielding a blood-stained scalpel to slash and chop liberally into human meat. A graphic interpretation to be sure, but one that, in its overwrought sense of pantomime, disparagement and exaggeration, is open to pointedly humourous exploitation while retaining an essential basis in truth.

3. Creativity in Metaphor

... intuitions give the appearance of miraculous flashes, or short-circuits of reasoning. In fact they may be likened to an immersed chain, of which only the beginning and the end are visible above the surface of the consciousness. The diver vanishes at one end of the chain and comes up at the other end, guided by invisible links.


If one accepts that metaphor lies at the root of the human creative faculty, a computational treatment of the subject then faces a serious philosophical dilemma: how does one explain human creativity without explaining away human creativity? That is to say, surely it is the very nature of creativity that it cannot be wholly reduced to a collection of algorithms and rules, for such rigorous descriptions are the very antithesis of creative thought.

We might define metaphoric creativity as comprising both a cognitive act, whereby a conceptual domain is viewed from a novel and startling perspective, and a linguistic act, whereby this perspective is subsequently conveyed to another via a clever and engaging juxtaposition of words. In essence, creative metaphor requires both invention and communication, as illustrated by the following sly extract from Umberto Eco (1994):

"... Insufficient consideration has been given to the new underground religious war which is modifying the modern world. It's an old idea of

---

1 This might be achieved by establishing a more pointedly immediate straight-man than simply the general class of Surgeons. For instance "My wife's brother Mortie is supposed to be a qualified doctor, but o' vey, he should have been a butcher/vet ..."
mine, but I find that whenever I tell people about it they immediately agree with me.

The fact is that the world is divided between users of the Macintosh computer and users of MS-DOS compatible computers. I am firmly of the opinion that the Macintosh is Catholic and that DOS is Protestant. Indeed, the Macintosh is counter-reformist and has been influenced by the 'ratio studiorum' of the Jesuits. It is cheerful, friendly, conciliatory, it tells the faithful how they must proceed step by step to reach—if not the Kingdom of Heaven—the moment in which their document is printed. It is catechetical: the essence of revelation is dealt with via simple formulae and sumptuous icons. Everyone has a right to salvation.

DOS is Protestant, or even Calvinistic. It allows free interpretation of scripture, demands difficult personal decisions, imposes a subtle hermeneutics upon the user, and takes for granted the idea that not all can reach salvation. To make the system work you need to interpret the program yourself: a long way from the baroque community of revellers, the user is closed within the loneliness of his own inner torment.

You may object that, with the passage to Windows, the DOS universe has come to resemble more closely the counter-reformist tolerance of the Macintosh. It's true: Windows represents an Anglican-style schism, big ceremonies in the cathedral, but there is always the possibility of a return to DOS to change things in accordance with bizarre decisions; when it comes down to it, you can decide to allow women and gays to be ministers if you want to.

... And machine code which lies beneath both systems (or environments, if you prefer)? Ah, that is to do with the Old Testament, and is talmudic and cabalistic..."
response to the metaphor. Intuitively, it is as if both ends of the chain (tenor and vehicle) are pulled at once, lifting the chain out of the creative pool.

The Sapper interpretation system alluded to earlier in this paper embodies this intuition to provide, in Figure 2 below, a succinct description of the mappings both implicit and explicit in Eco’s article.

If metaphor creativity is a process of building upon known similarities to construct figurative edifices, then metaphor interpretation becomes the mirror process of deconstruction, whereby each edifice is dismantled to reach its foundations, those points where the metaphor is grounded in literal similarity. This, in a nutshell, is the guiding view of metaphor, and metaphor-driven humour, adopted in this work.

| [.86] If Operating-System is like Religion |
| [.86] Then Customer-Satisfaction is like |
| Sense-Of-Belonging |
| [.73] and Brand-Loyalty is like Religious-Faith |
| [.75] and Consumer is like Believer |
| [.71] and User-Base is like Congregation |
| [.42] and Windows-O/S is like Anglicanism |
| [.93] and Graphical-Window is like |
| Stained-Glass-Window |
| [.75] and Graphical-Icon is like Holy-Icon |
| [.76] and Wimp-Enviror is like Cathedral |
| [.79] and Mac-O/S is like Catholicism |
| [.75] and Command-Line-Interface is like |
| Protestant-Work-Ethic |
| [.63] and Ms-Dos is like Protestantism |
| [.87] and Graphical is like Baroque |
| [.04] and High-Level-Software is like |
| Religious-Text |
| [.65] and Software-Command is like |
| Moral-Directive |
| [.86] and Machine-Code-Instruction is like |
| Cabalic-Message |
| [.79] and Hexadecimal is like |
| Hermeneutic-Code |

**Figure 2: Sapper Interpretation of Eco’s Operating-System as Religion Metaphor.**

4. Tying It All Together: A Metaphoric Agenda for Humour Analysis

Thusfar we have attempted to characterize the cleverness that is a signature feature of humour in terms of a computational model of metaphor. By grounding humour in the cognitive mechanics of cross-domain mapping, we give it conceptual depth, and in doing so give it scope (we hope) for profundity, and thereby truth. This desire to allow humour to express non-trivial truth, a characteristic of genuine humour as argued in section 1, means we are loath to characterize humour in terms of simple logical contradiction. Certainly, contradiction is sometimes responsible for humour, as in the classic piece of poetic nonce “One bright day in the middle of the night, two dead boys got up to fight...”. However, one quickly tires of such formulaic absurdity. When contradiction is othertimes (rarely) witty, we believe the inherent humour derives from the cleverness of the contradiction, not the contradiction itself. The Irishman ² who wanted to fly a rocket to the sun, but intended to go at night to avoid being burnt to a crisp, demonstrated superficial idiocy to be sure, but also a remarkable caniness for playing with the facts of reality to suit his own ends.

For this reason, we are also reticent to endorse the view of humour proposed by Minsky (1980), who suggests that humour is a cognitive device for recognizing (and thus later censoring and avoiding) the lapses in logic that common-sense reasoning can create. With humour, we believe that cleverness is often its own cognitive reward, and any lessons of logic learnt in the pursuit and exercise of cleverness are simply an added bonus.

4.1. The Role of Localist Memory and Background Knowledge

The Sapper model employs spreading activation over a localist concept network; such a network is essentially a melting-pot for relationships ranging from the phonetic (as exploited by puns) to the conceptual (as exploited by striking metaphor), in which contextual priming can alter the relative salience of concepts (see Hofstadter & Mitchell 1988). An early experiment with humour in such an

² No, it was neither of the present authors. Feel free to insert your own ethnic stereotype here.
environment concerned the generation of MAD Magazine movie spoof titles, such as "RoseMia's BooBoo" (starring Mia Farrow), "A Crockwork Lemon", "Days of Blunder", "FlopLiners", "Stench of a Wench" and "Presumed Impotent". While we would argue that these represent not simple puns but sound assimilated metaphors, inasmuch as they express a certain aesthetic belief (Lemon, Crock, Flop and BooBoo being highly critical terms that demand metaphoric treatment), this earlier approach was predominately lexicosemantic, in which the same vein as Lessard and Levinson's 1995 treatment of 'Tom Swiftys' (pun-like expressions such as "I hate chemistry, Tom said acidly").

A more challenging use of a localist memory is to incorporate acontextual or low salience elements of background knowledge into a joke, thereby giving rise to the truth and unconventionality criteria of section 1. For instance, the Jay Leno witticism which goes "The U.S. Postal Office today announced that they are issuing a Marilyn Monroe stamp costing just 28¢: Hey, just one lick and you're an honorary Kennedy!" exploits a web of low salience background knowledge concerning both JFK and RFK's sexual dalliances with Ms. Monroe. While not metaphorically grounded, these jokes score highly when measured against our current agenda of truth, pain and unprototypicality, and are readily accommodated by a memory-situated model of metaphor such as Sapper.

4.2. Collecting the Pieces
We have seen above how the Sapper model of Veale & Keane (1995a) provide a memory-based metaphoric basis for humor analysis; In Veale & Keane (1992) we also present a two-tiered, metaphor-grounded model of meaning representation, called Conceptual Scaffolding, that accommodates immediate humour/metaphor interpretation at a broadly thematic level, in which basic perforations, ameliorations and ironies can be recognized (as with Brodsky's remark concerning "Soviet penal homeopathy"), whilst allowing for the future elaboration that accompanies deeper conceptual appreciation (using Sapper, for instance).

Veale & Keane (1994) also describe a belief calculus for incorporating speaker and hearer perspectives into the interpretation process, one that is geared toward use in a localist network like Sapper, and one which places special emphasis on the pejorative/ameliorative parallax introduced by a metaphor. This calculus was initially designed for evaluating the likely changes to a hearer's mental representation as caused by a metaphor, following the principle that learning is diminished if a hearer believes himself to be insulted, or believes the speaker to self-flattering. However, this calculus is thus very much suited to the determination of the pain perspective so valuable to humour's success. It is also particularly suited to the determination of goal cohesion, that is, in evaluating how a speech-act meets (or fails to meet) its expected descriptive goals. As with our Time-Life example of JFK: Arthur and the question of Jackie: Guinnever's adultery, such a mechanism allows us to detect when a metaphor crosses over from the realm of prosaic device to potential jape.

5. Conclusions
Mark Twain once delighted the members of a German Philological society with a speech (delivered in German) which poked fun at the oft-contorted rules of German syntax. While discussing the tendency of the German separable verb to distance itself from its subject to such an extent as to force the reader to use a telescope to locate it, Twain remarked anecdotally that one German historian had inserted an entire history of the Thirty Years War between the subject and verb of a single sentence. One is to be entirely grateful, Twain added, that the same author was not allowed to attempt a history of the Hundred Years War. The German audience roared their approval.

Humour, like metaphor, encourages us to view the familiar as strange, and the strange as familiar. This paper has proposed a computational model of humour that is built upon a metaphoric substrate, our intention being to ground humour directly in what we consider the main cognitive engine of human creativity. The elements of this computational metaphor substrate, such as the Conceptual Scaffolding model, Sapper, and its complementary analogue retrieval mechanism called Scout, have each been implemented and are available from the following metaphor-dedicated internet web site: http://www.compapp.dcu.ie/~tony/metaphor.html. In addition, empirical systematic comparisons
of these models with other computational theories of analogy and metaphor, such as the Structure Mapping Engine (SME) of Falkenhainer, Forbus and Gentner (1989) and the Analogical Constraint Mapping Engine (ACME) of Holyoak & Thagard (1989) show Sapper & Scout to rest on the firmest of epistemological foundations. Currently, our research centres upon the exploitation of these models for a robust implementation of the humour agenda outlined in section 4. In parallel, we are also exploring the application of these models to the analysis and automation of pastiche, or conceptual blending, particularly in a cinematic setting, e.g., Star Wars = King Arthur + Sci-fi and West-side Story = Romeo & Juliet + Jazz (see Veale & Keane 1995b).

Coulson (1995) argues also for the role of blended spaces much like these in the construction of humour. When these strands are taken together, we expect humour to show itself as a revelatory contributor in the human creative process.

References


Bad Vibes:
Catastrophes of Goal Activation in the Appreciation of Disparagement Humour
and General Poor Taste

Tory Veale¹, Mark Keane²

¹School of Computer Applications, Dublin City University, Ireland.
²Computer Science Dept., Trinity College, Dublin 2, Ireland.
tonyv@compapp.dcu.ie

1. Introduction
This paper attempts to accommodate humour within a broader computational framework of metaphor interpretation. Its main contention is the belief that humour and metaphor are both folds of a larger creative tapestry, wherein humour may represent either an aggressive and mischievous perversion of an underlying metaphoric mapping, or perhaps more opaquely, employ the mapping mechanics of metaphor for the more subtle purposes of coreference and narrative symmetry. This contention is quite similar to that of Koestler (1964), who bridges the gap between humour, wisdom and art with the notion of mental bisociation. While not making an explicit commitment to metaphor as the basis of creativity, it is clear from Koestler's writings -- and in the light of recent work on metaphor (from Black 1962 to Lakoff 1987) -- that bisociation and metaphor are different terms for the same cognitive process. In a sense then, it is a major goal of this paper to adapt a computational model of metaphor to provide an algorithmic level treatment of Koestler's theory, which like Black's Interaction Theory of metaphor, is comprehensively appealing yet nebulously underspecified. The flurry of highly energetic research in recent years into computational treatments of metaphor and analogy (Paul Thagard (1996) refers to this intensely competitive period as the Analogy Wars), means we are in a prime position to produce a more algorithmically formal theory of humour.

Additionally, we argue that when considered from the perspective of speech-act theory (as originated in Austin 1962), further significant parallels may be drawn between computational models of verbal metaphor and of verbal humour, for a theory of metaphor which does not account for the ability of a metaphor to satisfy the communication goals of the speaker and hearer is as sterile as a theory of humour which fails to explain why some jokes are funny and others are not. Both phenomena are equally sensitive to the contextual expectations of the hearer and to the proper (if mischievous) execution of the speaker's communication goals. But if the academic study of metaphor and humour can both claim to have originated with Aristotle, in his Poetics (see Hutton 1982), it is fair to say that metaphor has been the subject of considerably more theoretical thought in the two millennia that have followed.

In this paper then we exploit this parallel to construct (or borrow) some of the necessary elements of a computational theory of humour from the corresponding elements of an established model of metaphor comprehension, the Sapper model of Veale & Keane (1993; 1994; 1995). Sapper is a memory-situated model of structural mapping which employs spreading activation (see Collins & Loftus 1975; Charniak 1983; Hendler 1989) for the purposes of contextual priming, hypothesis gathering, and connotational attribution. Memory in Sapper is modelled using a localist network of concepts, which may be partitioned into different contextual viewpoints, or belief-spaces (see Wilks, Barnden & Wang 1991), each yielding perhaps different connotations for the same concepts.
Our discussion will focus on four different levels of humour/metaphor operation, the dispositional level (concerning speaker attitude), the contextual level (concerning hearer expectations), the representational level (concerning the speaker's use of domain knowledge) and the narrative level (concerning issues of coherence, timing, delivery and punchline).

Structure of this paper: with these goals established, the paper assumes the following structure: section 2 outlines Arthur Koestler's bisociation theory of humour and creativity, perhaps the most credibly comprehensive theory available to us; section 3 addresses the dispositional level in humour, and presents a network model of belief ascription, which employs polarized marker passing, or spreading activation, to detect disparagement or praise in a speaker's utterances (as attributed either to the speaker or to the hearer); section 4, concerning the context level, then builds upon this model of pejoration/amelioration to explain the context-sensitivity of utterances -- metaphoric or humorous -- in terms of speaker goals and hearer expectations; this provides a computational basis for analyzing the aptness, or downright bad taste, of a given utterance; section 5 then turns to the knowledge-level, and describes how polarized activation waves exploit the dependency structure of a domain during the comprehension of insults; section 6 outlines a model of the narrative level in humour analysis, grounding the notion of humourous surprise in the mathematics of catastrophe theory; the paper then concludes with a tying together of these ideas, with closing remarks, in section 7.

2. Koestler’s Theory of Bisociation
Arthur Koestler outlines a comprehensive theory in his "The Act of Creation" (1975) that both reconciles humour with other creative acts, such as those of art and of the intellect, but which also reconciles the cognitive role of humour with its physiological manifestation. Koestler does this by describing a mental process called bisociation, which he defines as the mental juxtaposition of two unrelated or incongruous frames of reference, domains or knowledge "matrices".

Koestler is careful to use the generic term 'matrix' to refer to the nature of the structures which may be bisociated. A bisociation of word sounds produces a simple pun, while a bisociation of complex ideas produces a more insightful joke. As demonstrations of bisociation, Koestler analyses two examples from Freud's earlier treatise on humour (see Freud 1938). The first of these is not strikingly funny, but nevertheless clearly demonstrates bisociation at work:

Chamfort tells a story of a Marquis at the court of Louis XIV who, on entering his wife's boudoir and finding her in the arms of a Bishop, walked calmly to the window and went through the motions of blessing people in the street.
"What are you doing?", cried the wife.
"He is performing my functions", replied the Marquis, "so I am performing his."

The main protagonist here employs pantomime to express the intended bisociation of himself with the unchaste bishop; in this instance therefore, bisociation is a goal of a character in the narrative. In the second example from Freud, bisociation is the last thing on the protagonist's mind:

The Prince, travelling through his domains, noticed a man in the cheering crowd who bore a striking resemblance to himself. He beckoned him over and asked: "Was your mother ever employed in my palace?"
"No sire", the man replied, "but my father was."

Here the bisociation is conveyed not through crude pantomime, but through unstated common-sense inference. The haughty prince, smug in the assumption that this lowly subject is a bastard son of the king, is dismayed to realize that he himself is the bastard, while this peasant is of royal blood. Bisociation drags him down to the status of commoner, whilst elevating the poor subject to a more noble status.

But why do we laugh at humour? Koestler's theory cleverly uses bisociation to bridge the divide between cognitive cause and physical effect. Quite simply, the endocrinal system upon which the more fundamental human emotions are based (such as aggression, tension, fear and surprise) has greater chemical
inertia, and thus a considerably slower state-switching time than the reasoning apparatus of the brain; one can switch to a radically new train of thought much faster than one can switch between different moods. Thus, when presented with a sudden humourous bisociation, the brain reacts faster than the endocrinial system, causing the latter to discharge the excess emotive tension through laughter, or some other physical outlet.

2.1. Bisociation as Coreference
Going one step further, we argue that many of the structures that are subject to bisociation in humour are referential in nature, and that for the most part, bisociation is humorously employed as a coreference mechanism. To see this point, consider the following example from Koestler:

One mother meets another, who happens to be a good friend, at the supermarket in Brooklyn. The second mother, noticing worry in the face of the first, inquires as to whether anything is wrong.

"Well", says the first, "my son came home from school yesterday with a note from the school psychologist. She says my son is suffering from an 'Oedipus complex' or whatever it's called."

"Oedipus Shmoedipus" replies the second, "what do these psychologists know anyway? Everything will work out fine so long as he loves his mamma".

This joke clearly hinges on the issue of misdirected reference: if the second mother, like the reader, only knew that Oedipus Complex is a pathological over-love of one's mother, she would not be so quick to offer her home-spun advice.

Most of what constitutes humourous farce also hinges on mistaken reference with the reader/hearer acting as omniscient observer, ranging from Shakespeare's Twelfth Night, to Mozart's Cosi Fan Tutte and The Marriage of Figaro, to Charles Dickens' The Wrong Box to Oscar Wilde's The Importance of Being Earnest, to the dopplegänger slapstick of Laurel & Hardy and the Hollywood farce of films such as Some Like it Hot, What's up Doc? and Noises Off. The dénouement\(^1\) in such farces, which corresponds to the punchline in a narrative joke, invariably involves the bisociation of multiple characters into one (e.g., "What? You're really my dead husband?" or "I see! Bill is the murderer, and that must mean Bob was the man on the roof!").

This notion of bisociative coreference shares much with the idea of domain mapping in modern computational theories of metaphor, analogy and conceptual blending (see Falkenhainer, Forbus & Gentner 1989 Holyoak & Thagard 1989; Fauconnier & Turner 1994; Coulson 1995; Veale et al. 1994,1995). Such theories provide an algorithmic basis for linking entities in one domain or space to entities in another, and thus, in their broadest application, may be seen as providing algorithmic accounts of bisociation itself. For the remainder of this paper then, we make the assumption that metaphor, bisociation and coreference are intimately related processes, sharing much of the same underlying cognitive mechanisms.

3. The Disposition Level: A Model of Disparagement and Praise
We argue that there exists a pragmatic imperative, in order to capture the richness of humourous and figurative utterances alike, to view jokes and metaphors as full speech acts (see Austin 1962). From this perspective an insult (say) is uttered by a speaker with a specific communicative intent, in the context of a speaker-specific world model, and subsequently interpreted by a listener relative to a local world model which is similarly listener-specific. To arrive at a full interpretation of an insult, a system must therefore characterize that component of meaning which is common to the world models of speaker and listener, before the information content, as opposed to the semantic content, of the utterance can be represented. So, a proper analysis requires the system to model the belief structures of the speaker and listener relative to each other in the context of the tenor (the target of scorn) and vehicle (the descriptive force). Recognizing the communicative intent of the speaker is also important if the speaker's beliefs are to be correctly modelled by the

---

\(^1\) Literally, and ironically, dénouement means 'an untying' of strands in French. Given the purpose of denouement in modern fiction, 'a tying together' would perhaps be more apt.
listener (e.g., is the speaker conveying a pejorative or ameliorative account of the tenor?).

Veale & Keane (1994) describe a calculus for the determination of pejorative and ameliorative connotation within different belief spaces. The existence of such multiple viewpoints within semantic memory allows for considerable meaning shift, or conceptual parallax, to occur in the comprehension of metaphor and humour; as an utterance may be interpreted differently within conflicting agent belief-spaces. This issue is of particular relevance in quantifying the overall ameliorative/ pejorative content of an insult, this being a dimension that must be to quantified in any disparagement theory of humour (see Zilliran 1983). Consider for instance the insult "Surgeons are Butchers", in which an ameliorative shift occurs when considering the metaphor/ bisociation from the viewpoint of the vehicle, and a complementary pejorative shift occurs when a tenor viewpoint is adopted.

In drawing two conceptual domains closer together, the metaphor attempts to reconcile the myriad associations of the tenor and vehicle into a cohesive whole. However, while a systematic 1-to-1 mapping between these domains is achieved, the metaphor nevertheless fails to marry the connotative meaning of the two domains completely, but this is as it should be. What makes a good metaphor interesting is the volatile nature of the juxtaposition - while there are good structural reasons why both domains should be united, as demonstrated by the mappings above, there also exist conflicting associations that imbue the metaphor with a certain unsettling quility. Thus, bubbling below this neat structural reconciliation is a tension that threatens to tear the domains apart, rather than bring them together, and it is this tension that lends metaphor its characteristic flavour of disturbing unorthodoxy.

For instance, in the Surgeon as Butcher insult/metaphor, Human-Flesh is seen to resemble Meat, which consequently diminishes its connotation of sacred inviolability. Furthermore, this analogue opens the door to anthropophagy, allowing human flesh to be viewed as edible, and perhaps more disturbingly, as even tasty (consider for instance the scope for black humour even in a character as repulsive as Hannibal Lecter). Likewise, an operating theatre is seen to resemble an abattoir, which serves to accentuate the associations of pain and death which are normally so understated in the stereotypical conceptualization of surgery. The creation of this mapping also diminishes the connotations of brightness, and indirectly, safety, associated with operating theatres, highlighting instead the connotations of darkness and danger commonly associated (rightly or wrongly) with abattoirs. This of course assumes a reading from the tenor perspective; in contrast, from the vehicle perspective, the concept of abattoir would seem to be lightened somewhat by the association with clinical and bright operating theatres, places where pain, darkness, screaming and death assume a less salient cast. Likewise, from the butcher’s perspective, the apposition of meat with human-flesh imbues the former with a sacred overtone that is not usually attributed to the victualler’s profession.

A means of gauging the ameliorative/pejorative shift that the insult creates between the tenor and vehicle viewpoints is therefore necessary if the system is to accurately determine the cognitive, or emotive force of a insult, that is, a context-independent measure of the persuasive power of the insult regardless of the speaker’s intent. Such a measure is a principal determiner in updating the belief-space of the hearer in response to the speaker’s insult, and indicates much of what should be considered the imparted ground of the insult/metaphor. A connectionist means of obtaining such a measure is simply implemented within the Sapper model of metaphor interpretation of Veale & Keane (1993; 1994; 1995; 1996), as illustrated in Figure I below:
As shown in Figure 1, this approach designates two concept nodes, AMELIORATION (denoted A) and PEJORATION (denoted P), to serve as special activation collectors, which accumulate but do not redistribute activation from adjoining nodes. Descriptive concepts throughout the network which possess any measure of ameliorative/ pejorative connotation are connected to these nodes, whereby the association strength of each linkage reflects the extent of the connotation. In this respect these nodes resemble the pragmatic node that tempers the ACME analogy model of Holyoak & Thagard (1989) with contextual activation. Note that different belief-spaces may employ different connections (or connection strengths), reflecting the competing views of different cognitive agents (e.g., speaker and hearer). A further elaboration of this collector notion is outlined next in section 4, where such pragmatic nodes are used to encode speaker/hearer goals and expectations.

4. The Context Level: Speaker Goals and Hearer Expectations
A metaphor cited by Indurkhya (1992) compares marriage to an icecream cone, a comparison which reportedly caused an unsurprising degree of consternation among the attendants of one wedding ceremony, when unrealistically used by the presiding clergyman. The basis of similarity between tenor and vehicle, as offered by the injudicious priest, is that marriage, like an icecream cone, can be consumed at all once, providing a frenzied but short-lived source of pleasure, or alternately, can be enjoyed slowly and made to last a long time. The structural content of the metaphor is clear, but the impression it imparts is woefully unsuited to the context of a marriage ceremony. By viewing metaphor as a speech-act, we see that this particular metaphor carries a well-founded meaning but simply violates too many of the contextual expectations of the hearer (a participant at a wedding), who expects to be lavished with grand metaphors extolling the majesty and sanctimony of marriage. It is not that icecream necessarily imparts a pejorative impression; in many contexts it is actually evocative of warm feelings of summertime and fun. It is simply the case that it is devoid of the connotations of grandeur and opulence which one tries to evoke at weddings. A much more apt metaphor is one that compares marriage to a banquet, in which the same moral message is conveyed, but in a more suitable guise.

It is important to consider then that an utterance that is metaphor, or humourous, or both, when delivered in a specific context, will most probably be employed by the speaker to
satisfy specific communicative goals, but will likewise be expected to satisfy certain social and contextual expectations held by the hearer. Before a proposal to account for this goal / expectation sensitivity is outlined, let’s consider what form such pragmatisms might take:

**Example Speaker Goals:**
- Communicate Conceptual Structure [Model]
- Convey Pejorative Impression [Insult]
- Convey Ameliorative Impression [Praise]
- Emphasize Particular Concept Facet [Focus]
- Establish Mood / Preserve Tone [Conform]

**Example Hearer Goals:**
- Ego-Preservation
  - Less Likely to believe personal insults
  - More Likely to believe personal praise
- Objective Interpretation
  - Less Likely to believe self-praise
  - More Likely to believe self-deprecation

**Example Hearer Expectations:**
- Context-Dependent, Cultural Conformity e.g.,
  - Solemn in church, irreverence in parties
  - Honesty with parents / authority figures

The first of the speaker goals -- communicate conceptual structure -- is of course the default goal implicitly assumed in all models of metaphor, while the pejoration and amelioration goals (pejoration being a mainstay of mist humour) require an additional aesthetic judgment faculty as outlined in section 3. The focus goal of facet emphasis is an element of metaphor comprehension that has received considerable attention in the literature at large; for example, the Sapper model of Veale & Keane (1995) discusses facet emphasis in terms of analog bindings -- those facets which are relevant are those which possess equivalents in the new domain. Additionally, the salience imbalance theory of Ortony (1979) is ideally placed to account for this aspect of figurative communication. Similarly, the hearer goal of ego-preservation is modelled in the belief framework of section 3, through formalisation and quantification of the role of self esteem in belief ascription, while the hearer goals of objective interpretation form the basis of the communicative force metric formalized as part of the same belief framework (see also Veale & Keane 1994). The speaker conformity goal, however -- setting of mood and tone -- requires an extension of our treatment of pejoration and amelioration to account for other speaker tones, for instance, the mood of sanctimony and grandeur which was unfortunately undermined by the malapropos ice cream metaphor cited by Indurkhya. The flip side of this goal is the hearer expectation of conformity, which also demands that the moods established by the speaker’s metaphors are suited to the context in which those metaphors are uttered. It is to the computational treatment of contextual demands that we now turn.

First, let’s re-examine the malapropos marriage metaphor, in order to propose a formal diagnosis for its failure. Consider then Figure 2 overleaf, in which four different contextual expectations, or social norms as established within the prototypical wedding scenario, are illustrated. Figure 2 illustrates i) the tendency to accentuate the sacred rather than the profane, ii) a preference for all things happy rather than sad, iii) an emphasis upon things grand rather than meagre, and iv), an expectation that the special aspects of the event, rather than its more mundane features, will be highlighted. Such goals and expectations are codified in the Sapper memory network as pragmatic nodes, depicted in Figure 2 as bi-directional vertical arrows, with contextual tendencies illustrated in black. These pragmatic nodes combine the spatial intuition of the orientational scaffolding constructors of Veale & Keane (1992a,b) with the notion of an activation sink, or collector node, as employed in the belief-space augmentations of section 3. So these pragmatic nodes not only collect (and not re-propagate) activation energy from adjacent concept nodes, they also dynamically reflect a contextual observation of the aptness of the metaphor under analysis.
Pragmatic nodes operate then by determining the conceptual parallax that exists between the tenor and vehicle schemata. This parallax effect is quantified on the basis of the activation differential that is measured at contextually-determined collector nodes, as different conceptual perspectives are successively imposed upon the network. The ameliorative/pejorative shift which forms the basis of the belief-space model in section 3 can therefore be seen as a particular case of this broader pragmatic mechanism at work.

Note also that this collector mechanism can be used in reverse to contextually constrain the generation of apt metaphors. Simply, those pragmatic nodes corresponding to the contextual goals of the speaker, and the contextual expectations of the hearers as perceived by the speaker, are clamped. Activation energy then issues forth to either excite aesthetically suitable concept nodes or inhibit their malapropos neighbours. Thus, when higher-level structural criteria (such as plot, timing, etc.) are employed to generate a joke or a metaphor, the memory network will already have been primed to distinguish the wheat from the chaff.

5. The Knowledge Level: Insults Exploit Dependency Structure

Insults operate by highlighting a perceived flaw in the personality of the victimized tenor. It follows then that this weakness should be reflected in the internal knowledge representation of the tenor in the mind of the speaker. Thus, we can conceive of insult generation as a process that seeks out flaws in the dependency structure underlying a tenor domain, where dependency is here used to denote those representational facets of a concept through which ameliorative or pejorative connotations may most fruitfully be channeled. This notion is best illustrated with an example.
The American ambassador to India recently insulted that nation’s defence by describing the India/Pakistan border as “porous”; in doing so she exploited a well-worn metaphor schema (see Lakoff & Johnson 1980; Johnson 1987) which views a Location as a Container; the metaphor is comprehended by solving the proportional equation Porous is to Container as Water-Tightness is to Country, in the fashion of Aristotle’s fourth variety of metaphor as discussed in the "Poetics". Figure 3 above shows the ambassador to be employing a specialization of the Location as Container metaphor, namely a Secure-Nation as Watertight-Container schema. Sapper uses the following strategy to resolve insults based upon such proportional metaphors:

(i) Positive spreading activation issues forth from the tenor (Pakistan) and qualifying vehicle (Porous) nodes.

(ii) A suitable domain crossover point between the tenor and vehicle activation waves should eventually be reached. If the crossover point is an active conceptual bridge (Sapper’s term for the in-memory representation of a conventional metaphoric mapping), then the insult is seen as an extension of a conventional metaphoric schema (such as Location as Container). However, if the crossover is a dormant conceptual bridge (in Sapper terms, a potentially novel metaphor) the insult is seen to be that much more creative; this in turn can prolong comprehension time, as there will likely be many more dormant than active crossovers between the tenor and vehicle domains to investigate (see Veale et al. 1995).

(iii) The pathways between the tenor and the crossover point, and between the vehicle and the crossover point, are checked for a valid dependency structure. That is, the vehicle must relate to the crossover point in a coherent semantic fashion, and this coherence must be reflected in the tenor domain as well.

In this sample insult then, positive activation energy (or zorch, as it is termed in Hendler 1989) is spread from the matriarch nodes Pakistan and Porous. The word “border”, as used in the linguistic expression of the insult/metaphor, provides a ground for the interpretation, such that only those conceptual paths which pass through the concept node Border will be considered as valid interpretations of the metaphor. The path taken
through conceptual space is also directed to maximize the magnitude of zorch (regardless of polarity, inhibitory or excitatory) reaching each node; Sapper therefore pursues a trail which employs the most prototypical usage of each concept. For instance, the pathway through the vehicle domain proceeds from Porous to Water-Tight-Container to Container to Container-Seal, side-stepping the node Colander, as this is a less representative example of a container; this prototypical bias is codified in the conductivity of the appropriate taxonomic links.

Likewise, the tenor-domain path proceeds from Pakistan to Nation through National-Security to Border; the domain crossover point at the heart of the metaphor is thus Border: Container-Seal, a conceptual bridge which has previously been established by Sapper upon noting the superordinates Extent and Boundary shared by both (this form of heuristic bridgebuilding is termed Triangulation in Sapper). Tracing back from this crossover point, Sapper then establishes other higher-order mappings, in a process known as Squaring, between the concepts National-Security: Water-Tightness, Nation: Water-Tight-Container, and Complacency: Porous. In addition, the dependency structure of the tenor-domain path reveals that National-Security is that aspect of Pakistan most emphasized by the metaphor. However, as the energy reaching the crossover point from Porous is negative, having passed through an inhibitory link between Porous and Water-Tight-Container, the metaphor conveys an essentially pejorative account of Pakistan's national security -- that is, it is recognized conceptually to be an insult. In response then, the system augments semantic memory with a new inter-concept relation, a high conductivity linkage between the nodes Pakistan and Complacency (the perceived analogue of Porous).

5.1. More on Dependency Structure
As stressed in the foregoing discussion, the main issue of coherence in the interpretation of insults built upon proportional metaphors is not so much structural isomorphism, or systematicity (as stressed in theories of analogy/metaphor such as those of Gentner, Falkenhainer & Skorstad 1989 and Holyoak & Thagard 1989), but dependency structure. That is to say, the activation pathways pursued through the domains of tenor and vehicle do not have to mirror each other in a strict 1-to-1 mapping; it is more important that these pathways meet at a suitable domain crossover point, and that each pathway, when viewed as the composition of different semantic relations, exhibits a coherent semantic content.

The semantic content expressed in the pathways of Figure 3 is that of conceptual dependency: Pakistan is a nation, which depends upon its national security, which in turn is dependent upon its national borders; a water-tight container is a container, which depends upon its water-tightness, which depends upon its container seal; porosity undermines the water-tightness of a container, and this negative message is propagated through the dependency structure of both domains until a pejorative account is taken of Pakistan's national security. The dependency structure of a domain comprises not only dependency links, but taxonomic, causal, substantive and meronomic relations also, allowing such a pejorative connotation to propagate freely along these particular link types.

This notion of dependency echoes the idea of affiliation developed in Wolff, Smith & Murray (1934): persons adopt attitudes toward affiliated objects, such as possessions, spouses, children, and so on, that reflect their attitudes toward themselves (see also Zillman 1983). People therefore tend to exhibit positive dependencies toward affiliated objects, so to undermine a target via an insult, one should thus undermine an affiliate of that target.

The importance of dependency structure in conceptual description is once again illustrated in the altogether more humorous insult/metaphor of Figure 4 on the page following (this example is due to the BBC television series Blackadder, written by Ben Elton and Richard Curtis): "Baldrick, your family tree has Dutch-elm disease!". This jibe is also commonly echoed on mail signatures across the internet which humbly proclaim "my gene-pool needs a dose of chlorine".

Once again a weakness is highlighted in the dependency structure of the tenor, Baldrick, by emphasizing the well-established pejorative meaning of the vehicle, Dutch-elm disease. What makes this example particularly interesting, however, is its reliance on a well-
established conventional metaphor, that of a family tree. In fact, this is a prime instance of what we term, in Sapper parlance, a *lexicized metaphoric bridge*, a cross-domain metaphor that has been lexicalized into the language. If one were at all inclined toward the pursuit of *cognitive-linguistic archaeology*, there are many such instances of manifest conceptual bridges in English, such as Code Key, Blood Bank, Gene Pool, and River Mouth, which provide an empirical basis for the Sapper notion that metaphor interpretation is a process of conceptual bridge-building.

![Diagram](image)

*Figure 4: Network for the insult “Baldrick’s family tree has Dutch-elm disease.” Notation: Links labelled “M” indicate culturally pre-established (i.e., lexicalized) metaphors.*

### 6. The Narrative Level: The Mathematics of Catastrophe

Thusfar, there is a level of analysis that a focus upon insults has allowed this paper to conveniently ignore in its entirety, and that is the *narrative* or *structural level*. Included under this umbrella term are complex issues of delivery, timing, build-up, misdirection and punch-line surprise. In particular, the structure of *suspense-and-surprise* jokes, where the build-up evokes a set of expectations and some humourous tension, which is subsequently deflated by the punch-line, needs to be addressed in a serious theory of humour. When one considers the sudden reversal of inference/belief created by the punch-line, the process of joke telling would seem to share much with the concept of *catastrophe* in mathematics.

### 6.1. An Overview of Catastrophe Theory

Catastrophe theory is a branch of mathematics concerning surface and form (see Thom 1974), focusing in particular upon the *discontinuities of sudden change* (i.e., *catastrophes*). It has been proposed that catastrophe theory has broad applications in cognition to visual perception (e.g., for noticing the *high-energy* discontinuities created by edge-boundaries) and to the semantics of natural language, both of which need to model dynamic state-change in the real world (see Wildgen 1982). Presently, we demonstrate the applicability of what is essentially a mathematical model of *surprise* to the analysis of humour, an enterprise in which the dashing of expectations (or *continuities*) is a basic element.

Catastrophe theory is the basis of the study of *morphogenesis* -- the creation of forms,
whereby any complex structure is definable as the catastrophic unfolding of a lower dimensional germ form. Figure 5 below illustrates a 3-D subset of the unfolding in 4-D space of the 3-D germ $y = x^4$ into $y = x^4 + zx^2 + ux$:

![Diagram of unfolding](image)

Figure 5: A Surface exhibiting a “kink” or fold catastrophe. When projected on a 2D plane, the central area of instability exhibits the characteristic cusp of catastrophic surfaces.

Note how the unfolding introduces a pinch or kink into the structure, creating a central area of instability in an otherwise stable structure. Delimiting this unstable region are two major discontinuities, which, depending upon your orientation, can cause a sudden upward surge or downward plunge while traversing the surface. Furthermore, within in the catastrophe region itself, there exists a structural bifurcation caused by the pinched unfolding; again, depending upon one’s orientation, a given $(x, z)$ pair will specify one of two different points on the surface. A bifurcation is generated when part of a surface is folded over onto itself, thus causing two different areas of the surface to become co-referenced in space. We build upon this notion in the next section to flesh out our model and describe how mental bisociation can be interpreted as a bifurcation in activation space, thus explaining why bisociation can have such surprising (catastrophic) effects.

6.2. Catastrophe as Bisociation

To summarize our argument thusfar, bisociation, as discussed in section 2.1, causes coreference of narrative entities; as argued in sections 3 and 4, the goals, expectations and general contexts of such entities are represented via activation levels in a belief-partitioned localist concept network. Thus, coreference of two entities essentially causes a sharing or convergence of activation levels in their respective belief-spaces, as the same pathways in semantic memory are now open to both; in mathematical terms, this corresponds to a bifurcated folding-over of the activation-space. And in the metaphor perspective of the Sapper model, this folding corresponds to the warping of conceptual space that occurs when a new metaphoric bridge is constructed between two domains. For our current purposes, it is useful to consider the sudden and widespread reversal in network activation values that can occur as a result of such folding as a catastrophe, and to consider such a catastrophic event as being a primary characteristic of the comprehension of punch-lines in humour. Consider for instance the reversals inherent in the following joke:

Four gents go out to play golf one sunny morning. One is detained in the clubhouse, and the other three discuss their children while walking to the first tee.

"My son," says one, "has made quite a name for himself in the homebuilding industry. He began as a carpenter, but now owns his own design and construction firm. He's so successful in fact, in the last year he was able to give a good friend a brand new home as a gift."

The second man, not to be outdone, allows how his son began his career as a car salesman, but now owns a multi-line dealership. "He's so successful, in fact, in the last six months he gave a friend two brand new cars as a gift."

The third man's son has worked his way up through a stock brokerage. And in the last few weeks has given a good friend a large stock portfolio as a gift.

As the fourth man arrives at the tee box, another tells him that they have been discussing their progeny and asks what line his son is in. "To tell the truth, I'm not very pleased with how my son has turned out, he replies. "For fifteen years, he's been a hairdresser, and I've just recently discovered he's homosexual. But, on the bright side, he must be good at what he does because his last three boyfriends have given him a brand new house, two cars, and a big pile of stock certificates."
As the first three gents enthuse about their offspring, their self-esteem growing steadily in relation to their fatherly pride, the collector nodes *Pride*, *Success* and *Self-Esteem* swell with positive activation in the belief-space ascribed by the system to each. However, with the comprehension of the fourth gent’s final statement comes a sudden realization that each golden son is most likely a homosexual, a fact which is of no humourous import in and of itself, but quite horrific (and thus, funny in a disparagement sense) when viewed from within the belief-space of the stereotypical middle-aged, middle-class conservative man. With the delivery of the punch-line, negative activation thus floods into the *Pride* and *Self-Esteem* collector nodes in each belief-space, creating a sudden discontinuity, or reversal of activation level -- in effect, a catastrophe!

### 6.3. Timing: Punchline as Catastrophe

We argue that it is the fact that three such catastrophes occur, one in each of the ascribed belief spaces, with each one deconstructing and reversing the inferences made by the hearer during the course of the build-up, that makes the previous joke so satisfyingly *coherent*. Too often we are subjected to bad jokes in which the punch-line only obliquely acknowledges the preceding build-up. It should prove a worthwhile project then to chart the activation levels of different collector nodes over time, during the comprehension of different jokes, to ascertain whether *activation signatures* can be derived for different classes of joke.

Timing, of course, is a fragile issue here, one which is perhaps the most characteristic feature of a good jester. For there exists a delicate dynamic between suspense and surprise in a joke narrative: should a joke be told too quickly, no time is allowed for expectations to build in the hearer’s mind; and should a joke be told too slowly, tension and suspense diffuse and are superseded with boredom. Most importantly, the timing should be such that the punchline is delivered in a single narrative *beat*, that is, in as short a time frame as is necessary to create a sudden dashing of expectations. This clarifies why a joke that must be explained is not funny -- quite simply, an explained joke does not have the correct *temporal signature*, with a *multi-beat* explanation taking the place of a single beat punchline, slowly puncturing a balloon of expectations that was inflated for the purposes of explosion.

More deserves to be said about this notion of a ‘beat’. As with a musical score, a beat is a point of emphasis in the delivery of a narrative, and may correspond to an information-bearing statement, a meaningful look or even a pregnant pause. A beat is a meaningful abstraction in the sense that it frees a model of humour analysis from considerations of absolute time, allowing a theory to base itself instead on the progression of beats in a narrative. With each successive beat, new information may be gleaned which increases the connectivity of the system’s network representation of the narrative, allowing new connotations to become active and perhaps alter (usually in a continuous fashion) the activation levels of the goal nodes in a particular belief-space. The golf joke previously can be divided sensibly into six beats: an introductory beat which sets the scene and establishes prototypes (gents in a golf club); three positive character-building beats, one for each of the three proud fathers; one negative character-building beat, for the disappointed fourth father and his embarrassing son; and finally, a punchline beat which establishes, by coreference, a humourous reversal.

The punchline should correspond to the final beat of the narrative, for as argued above, to extend a punchline across multiple beats is to rob a joke of its sting. This final beat should be a bisociative event (we denote it as the *point of bisociation*), causing a co-reference of entities (with a folding of their activation spaces) and thereby generating a sudden reversal of goals against hearer expectations.

#### 6.3.1. The Inertia of Context

It is conventional in connectionist systems that the activation level of a given node in memory is determined as the sum of all impinging activation at that node (see Quinlan 1991). However, when one considers that the average joke will sensibly contain between three and ten beats, the beat notion provides little resolution within which to continuously chart the changing activation levels of character goals. If we cannot assume continuity as normative, then each intra-beat activation change may well constitute a catastrophe, thus blunting our theory and confusing the system.
We avoid this issue in the current implementation by assuming that the sum of impinging activation at a node merely establishes an asymptote for the perceived activation level of that node; the node level continuously converges toward this asymptote with each successive beat. In this way the preceding activation context exerts a dampening effect on a node's activation, thus serving two important roles: the expectations aroused during a narrative build-up possess an inertia that is not easily swayed; and only major activation changes of significant magnitude and rapidity will register as catastrophes. For our current purposes, a level change of 40% or more in a single beat constitutes a goal catastrophe.

6.4. Progression in Narrative Humour

We are now in a position to specify a procedure for the analysis of narrative humour in algorithmic detail. Consider an outline of this procedure in Figure 6 below:

0: Clear Narrative Context.
   Clear Activation in all Belief Spaces

1: Start Next Beat

2: Analyse Propositional Content of current beat, call this content P

3: If a new character is established in P, then
   • Create a belief-space for this char.
   • Chart all relevant goal nodes for this character in successive beats

4: Augment semantic memory with the semantic linkages contained in P
   This should increase the connectivity of one or more belief spaces

5: For every belief-space B in the narrative context do
   If a belief-space is newly created, then
   • Spread activation in that space to a certain horizon H
   else
   • Try and advance activation wave in B further, staying within H

6: If beat P is the punchline and final beat then
   • Take a snapshot of the activation of each goal in each belief space
   • Try to bisociate P with every event of the narrative (e.g., use Sapper)
   • Advance activation waves in each belief space one more time (as in 5)
   • Determine which goal nodes undergo a catastrophe in activation
   • Determine, from +/- changes, the wise-guys and the fall-guys
   • Evaluate the symmetry of the narrative by counting the number of bisociations

   else
   • Go back to step 1 and start next beat

Figure 6: Algorithmic Procedure for the Analysis of Narrative Humour

Steps 1 through 5 concern themselves with the establishment of characters, scene and context, usually in a way that confirms our prototypical expectations. For instance, in the earlier golf joke, the initial four beats establish a scene containing four middle-class, middle-aged conservative white men, though this information is not presented explicitly, rather it is derived from prototypical associations of the word 'gent'. Expectations are often exploited in this way both to maintain the conciseness of a narrative, and to suggest a false avenue of plot progression. The first four beats of the golf joke establish, in step 1, connections between each gent and his son, and between each son and some icon of wealth, and source of fatherly pride, such as an expensive car, house or stock portfolio. Step 5 exploits these connections to (initially positive) spread activation successively further in each belief space, gradually saturating the Pride goal nodes in each space.

Step 6 finally causes a bisociation of the punchline ("his last three boyfriends have given him...") with each preceding act of extravagance, recursively causing each son to be bisociated with the concept Homosexual-Boyfriend via a new Sapper bridge. A subsequent advance of activation in each belief space now has a new linkage to traverse, allowing activation waves to now reach this previously unprimed concept, and from there
flood into the concept Shame, which is negatively connected to the goal node Pride. Once the new level of activation at this goal node is compared with pre-punchline (snapshot) levels, the system can determine that a reversal has occurred. Furthermore, summing the activation levels of all goal nodes in each belief space allows the system to recognize the first three gents as the *full guys* of the narrative — characters whose egos are swelled in the build-up, and suddenly deflated in the punchline; in contrast, the fourth gent, who has already distanced himself from a disappointing son, gains by comparison and is seen as a *wise guy.*

This process of inflation and deflation is best illustrated with a diagram. Figure 7 above presents a chart of the Pride goal-node in each gent belief-space for this joke, as derived from a Prolog implementation of the theory; note the catastrophic activation reversal at the *point of bisociation* in beat 6.

Symmetry and coherence, which when combined with goal reversal, is the basis of the bitterest of ironic humour, may be evaluated as outlined in step 6 by determining the number of bisociations generated in the punchline. In the golf joke, symmetry is perceived due to the simultaneous crash of activation in three different belief spaces at the point of bisociation. The reader may recall that the jokes employed by Koestler (and borrowed from Freud) also rely heavily upon symmetry of this kind. In the tale of the marquis and the bishop, this symmetry is actually conveyed in a pantomime fashion by the cuckolded protagonist, while in the tale of the prince and the peasant, the haughty protagonist is forced to apply his snobbish pre-punchline reasoning to himself and his mother, and conclude that he, and not the peasant, is in fact a bastard. A system activation chart for these jokes is presented in Figures 8 and 9 below:

At best, this is a flimsy joke regarding a subtle revenge; at worst, it is frankly unfunny in a modern setting. Open to alternate readings, our reading of the joke is as follows: the marquis, on finding himself cuckolded by a bishop, refuses to react violently (as expected), but
instead humiliates the bishop by reminding him of just how wretched a clergyman he is. In pantomiming a prototypical bishop, he borrows the bishop's aura of pomp and social standing (salient goals of any royal personage), while suggesting that by violating the priest's vow of chastity, nothing remains of the Bishop's role that could not otherwise be performed by a layman. In contrast, the marquis claims sexual intercourse as a right enjoyed by non-priests. In particular, given that nobility is a goal prized by all of royal blood, conjugal rites are highly salient in the belief space of the marquis; thus a bisociation of bishop and marquis is a damning reminder of the bishop's own goal failure.

While relying more on subtle common-sense inference rather than explicit pantomime, the bisociation in Freud's tale of the prince and the peasant is a clearer example of wise-guy: fall-guy reversal in humour. This reversal is given graphical form in the activation chart of Figure 9. In this joke, the narrative build-up both establishes the prince as a man of importance and nobility, and leads us to anticipate the prince's line of reasoning: the peasant resembles the prince because the prince's father seduced -- as was the fashion of the day -- the peasant's mother; the peasant's mother is thus a slut, and the peasant a lowly bastard. At this point, the prince is clearly the wise-guy of the narrative, and the peasant the fall-guy. But as with most of narrative humour, the initial wise-guy is merely a fall-guy in waiting.

With the realization that the it was the peasant's father, and not his mother, that worked in the King's service, we must revise our assumptions and conclude that in fact the prince is the illegitimate son of the peasant's father; if we maintain our pre-punchline disposition (as Koestler's notion of dispositional inertia suggests), we must now view the prince as a lowly bastard, and his mother as a slut. Since nobility of birth is an almost defining goal of a royal personage, this is a catastrophe of great magnitude for the prince; in contrast, the coreference of the peasant with the prince elevates the peasant, in terms of pride and importance, from his lowly social status. The effect is a complete reversal of roles, with the prince metaphorically (and perhaps literally also) falling from his high horse.

7. Summary and Conclusions
Humour, like metaphor, is a goal-driven speech-act that must not only satisfy the communicative needs of the speaker, but must also cohere with the contextual demands of the hearer. Note that we do not say 'satisfy' the demands of the hearer, as it is the very nature of humour to subvert such expectations. However, this subversion must occur in a structured and systematic, rather than random and care-free, manner, if insightful humour is not to degenerate into absurd farce. It has been the guiding premise of this paper then that many of the same mechanisms that guide metaphor creation and comprehension are to an extent shared by the cognitive faculty of humour. There is a prima facie case for such a belief in the recent re-emergence of metaphor as a particularly deep-seated mental phenomenon (see Johnson 1987; Lakoff 1987; Veale 1996), for the deeper the mechanics of metaphor, the more likely it is to influence and contribute to the workings of other cognitive faculties.

We feel however that an analysis of joke humour cannot be properly realized without a prior computational treatment of attitude (i.e., how to recognize pain, superiority, self-deprecation), context (i.e., modelling the situation) and knowledge (i.e., modelling truth and reality). A consideration of metaphor comprehension provides us with a first cut at cognitively modelling these notions, and a basis upon which to construct more ambitious, catastrophe-based, theories of humour.
This paper has attempted to address each of these issues in computational terms, believing that only an algorithmic account will suffice to transparently ground a theory of a phenomenon so pervasive and entrenched as humour. In this respect humour is again very much like metaphor -- widely accepted theories such as the Interaction View of Max Black (1962) have great philosophical appeal, but are ultimately vague and nebulous, open to competing and often contradictory interpretations due to their fundamental lack of algorithmic detail (see Indurkhya 1992 for a discussion of the ambiguities inherent in Black's writings). Just as computational models of metaphor have served to lock down the elements of Black's theory in a more formal manner, we view the algorithms presented in this paper as a computational complement to Koestler's seminal work on bioniscia.

In many ways the work reported in this paper represents not so much a research result but a research agenda: it now remains for us to employ the various described techniques, such as conceptual bridging, polarized activation passing, partitioned belief spaces, catastrophic narratives and beat progression to chart the myriad forms of humour in evidence around us, to perhaps organize and generalize them to a level where it will be possible to posit some basic humour schemas or signatures. Armed with such signatures, it should be possible to not only endow systems with a greater sense of social understanding but with greater powers of creativity.

8. References


SPECULATIONS ON STORY PUNS

Kim Binsted* and Graeme Ritchie
Department of Artificial Intelligence
University of Edinburgh
Scotland EH1 1HN

ABSTRACT

There is a class of joke which consists of an anecdote, which is sometimes quite long and often has no inherently humorous content, followed by a punchline which is a distorted form of some well-known phrase, proverb or quotation. Usually the punchline purports to summarise or draw a moral from the preceding story. This genre has some unusual aspects, from the viewpoint of conventional claims about the attributes of jokes. These jokes also have certain structural or formal regularities which suggest that they might be amenable to computer generation. We outline how this might be done, by decomposing the construction of such story puns into a sequence of stages; some of these are clearly manageable, others are less straightforward. We also make some observations about where such an endeavour would fit within the broader field of humour research.

Finally, we make some observations about what such a system would demonstrate about the nature of humour.

2  STORY PUNS

There is a particular genre of joke which is quite common within British society, and possibly elsewhere. It consists of a story which need not itself have any humorous content (and which may be quite long), concluding with a single punchline which is usually a summing up of the content or moral of the story, although it may also be a suitable last line of the narrative or an utterance attributed to a character in the story. What is peculiar to this genre is that the punchline is always a distorted form of some well-known saying, usually a proverb but sometimes a famous quotation. For example, there is a well-established proverb in the English language "People who live in glass houses shouldn't throw stones" (meaning that those who are vulnerable to a particular form of criticism should not themselves make that criticism of others). This has been used as the basis of the following joke:

Once upon a time, many years ago, there was a chieftain in a remote tropical village who owned an old and battered throne of which he was very fond. One day, a visiting dignitary gave him a brand new and ornate throne, which the chieftain had to adopt immediately out of politeness. However, he could not bear to part with the old throne which had served him so well, so he stowed it away in the roof area of his grass hut, in case it should be useful in the future. Unfortunately the interior structure of his hut was too flimsy to support the weight of the large object, and it crashed through the grass ceiling, falling on the chieftain and killing him.

*This research has been partially funded by a PhD studentship from the Natural Science and Engineering Council of Canada.
The moral is that people who live in grass houses shouldn't sow thrones.

Many examples of these story puns\(^1\) have been constructed ([MN91] contains around 150), although few survive into common circulation. It is possible to use the construction of such jokes as a game or pastime, in which one person stipulates the original (undistorted) maxim, and another person has to construct the story pun (this was what occurred in the radio show reported in [MN91]).

The structure of such a joke is interesting, as various attributes of the main part of the story and the punchline (and the relationship between them) are different from those which are often posited as central to more conventional jokes. In particular:

- Neither story nor punchline creates ambiguity.
- The punchline does not resolve any ambiguity.
- There is no incongruity within the story.
- The punchline is always a suitable (congruous) ending.
- The punchline does not (apparently) violate any expectations set up by the story.

These jokes have some other noteworthy features:

- The *meaning* of the original, undistorted proverb plays no role in the joke at all - it is necessary only that the *wording* of the proverb be sufficiently familiar to the listener that it is evoked by the distorted form. This is in contrast to jokes typified by the much-cited "One more drink and I'll be under the host" (Dorothy Parker), where one factor in the humorous effect is the meaning of the related phrase "under the table".

- If the hearer is completely unaware of the original version of the punchline, there is no humorous effect.

- If a punchline is used whose meaning is virtually identical to the distorted proverb, but which uses different wording, there is no humorous effect. This is little more than a statement that these are a subclass of puns, since puns inherently rely on their wording.

- It is essential for the humorous effect that the maxim be *distorted*.\(^2\) A story in which a character living in a glass house came to grief as a result of throwing stones, concluding with the ungarbled proverb about such situations, would not have the same humorous effect.

As with many other forms of pun, the listener's reaction to a story pun is often to groan rather than to laugh, which emphasises that the main story portion is essentially setting up a context in which a pun can be made. In this way, story puns are comparable to real-life wit where someone achieves a humorous effect by describing some (actual) event or situation in a garbled form of a well-known maxim.

### 3 SOME ASSUMPTIONS

In tackling a topic as broad and as deep as humour, or even just verbally expressed humour, there are a number of different methodological strategies that could be adopted.

One possibility (perhaps exemplified by the General Theory of Verbal Humour [Att94]) is to devise a very general theory which attempts to encompass all possible phenomena in the area under consideration. Particular studies of sub-species of humour are then carried out within this overall framework, guided by its form and its principles. This is logically a very sound approach, avoiding the temptation to posit *ad hoc* devices, and integrating all work under one theory. Its drawback, in the current state of our understanding of humour, is that such a theory may be over-general to the point of vagueness, or (if more specific) may contain arbitrary details included prior to the evidence having been gathered. This sort of approach could be loosely labelled as *top-down*.

Alternatively, one could adopt a more *bottom-up* approach, in which detailed studies are made of particular humorous phenomena, using whatever theoretical constructs appear appropriate for that domain. This has the advantage that research can be fairly concrete, to the extent of even leading to testable computer programs, and hence quite fine details can be examined. The drawback, of course, is that the analyses may exist in

---

\(^1\) One of the current authors prefers the classification "shaggy dog story" for this genre, while the other author believes that shaggy dog stories consist of long anti-climactic stories rather than complex puns.

\(^2\) Another genre might be possible in which the punchline is an actual *undistorted* maxim and humour is produced by the indirect or unusual way in which the proverb applies to this particular story. However, these would not be puns, and would be different from the type of joke we are considering here.
a vacuum, unconnected to other phenomena, and
the mechanisms devised may be over-specific.
In reality, some blend of these two extremes
is necessary, and is normally used. One cannot
device a general theory (top-down) without at
least keeping an eye on the data, and one cannot
work on particular data (bottom-up) without
assuming (perhaps implicitly) at least some
theoretical basis. The work presented here, al-
though sketchy and speculative, tends towards
the bottom-up end of the scale. We believe that,
in the field of computational humour, it is im-
portant that some research effort should be put into
finding phenomena which are tractable, or nearly
tractable, with the current state of the art, and
pursuing these in some depth. A more firmly
top-down approach, while valuable, might not reach
actual computational implementation for many
decades.

There is an analogy here with the way in which
linguistic theory has developed over past decades
and centuries. Much of the early work involved
the detailed development of grammars (or frag-
ments of grammar) for particular languages, in
the absence of any over-arching theory of lan-
guage. Although theory-development became a
prominent activity in its own right under the in-
fluence of Chomsky, even modern generative lin-
guistics started from concepts (e.g. phrase struc-
ture) which had been developed within much
more specific and pre-theoretical streams of work.

4 PROPOSED SYSTEM

4.1 Overview

We have devised a relatively crude computational
model of how puny stories could be produced.
(This is not a model of how such jokes could be in-
terpreted.) As we will explain in section 5 below,
some of the components are still underspecified in
this preliminary design, but we believe that the
role of each module is relatively well-defined.

Viewed as an algorithm, the model could be
summarised as follows:

Choose a maxim.
Distort this maxim.
Construct a semantic representation of
the distorted version.
Develop some constraints from this seman-
tic structure.
Devise a story to meet these constraints.

We will use a worked example to illustrate how
this might be carried out.

4.2 Constructing the punchline

4.2.1 The method

The original (undistorted) proverb or saying
could be an input parameter for the process, as
in the “game” form of joke-construction men-
tioned in section 2 above. Alternatively, we could
assume that our proposed joke-generator would
have access to a set of proverbs, quotations and
other well-known sayings, and that it would ini-
tially choose one at random.

Construction of possible target punchlines from
a given maxim is not too problematic. The
punchline should be “similar” to the original in
some sense. The notion of “similarity” is es-
sentially that which is used in punning riddles
(cf. [PG84]), where puns can be based on inexact
matches between words. Tactics that are suit-
able include metathesis (spoonerism) (e.g. “throw
stones” → “stow thrones”) and substitution of
a phonetically similar segment (e.g. “glass” →
“grass”). There might be a large number of
phrases or sentences which could be produced
from the same original saying, so heuristics might
have to be developed to choose the more produc-
tive or suitable ones. On the other hand, the next
step in the algorithm, analysing the punchline,
might succeed only on a few distorted maxims, so
that pre-selection becomes unnecessary; that
is, a set of possible punchlines could be passed
to the next phase, letting the analyser filter out
unviable variants.

4.2.2 Worked example

The maxim chosen for this example is “Look be-
fore you leap”. Substituting similar sounding
words, for example, could produce several differ-
ent distorted maxims, such as:

- Hook before you leap
- Rook before two sheep
- Leak before you loop
- Book before you reap

All of these could potentially work in a story
pun. We will use the last one, “Book before you
reap”, for this example.
4.3 Analysing the punchline

4.3.1 The method

The punchline, as constructed, is still just a sequence of words, as the garbling process operated at a very low level, manipulating phonetic information. For the punchline to influence the construction of a suitable story, there must be a representation of its content. To produce this, there must be a module capable of scanning a sentence and constructing a symbolic representation of its meaning in some suitable formalism (e.g. some type of logic).

4.3.2 Worked example

There are many different formalisms in use in computational language processing. For the sake of this example, we will use a logic-like language in which operators representing the speech-act or illocutionary force of the utterance take as their arguments propositions (similar to those of predicate logic). It is quite possible that in developing a complete system, criteria for choosing or devising a suitable formalism would become apparent.

The semantic content of “Book before you reap” could be represented as:

\[
\text{IMPER( book}_Y\text{(hearer, Y, Time1) \& reapy}_X\text{(hearer, X, Time2, Time1 < Time2).}}
\]

A possible simplification that could be considered would be to restrict the model to morals stated as commands, so that any variant of the punchline which could not be analysed as a simple imperative (e.g. “Rook before two sheep”) would be eliminated.

4.4 Constructing the story

4.4.1 Simplifications

Story generation is very hard, and there is no prospect of having a good quality completely computer-generated fictional story in the near future. Generating a story which is guaranteed to be summed up by a given punchline is therefore the hardest part of this preliminary design. In order to make some progress, we believe it is both valid and feasible to constrain the type of story in various ways, so that our generator will be operating in a much smaller space of possibilities. There are various genres of story which seem to have certain stereotypical attributes, involving the types of characters, the types of events, the types of goals that characters have, the situations that characters may be in, etc. In particular, fables (e.g. [Len67]), fairy stories (e.g. [CR80]) and Greek myths (cf. [SPT96]) all appear promising. These relatively stylised genres also have the advantage of comparatively straightforward narrative devices – they make little use of flashbacks, multiple viewpoints, etc.

In the program of research we are outlining here, we would select one such genre as our area of study. For the sake of the worked example, we will choose the moral fable genre, since story puns bear a certain resemblance to moral fables, in that they build up to and justify a one-sentence moral.

4.4.2 Representing the story

Once we have adopted a manageable genre of story, the next issue is that of story representation. Although there is a considerable literature on literary analysis of fables (cf. [Car86]), creative writing (e.g. [EF90]), and story generation (e.g. [Mee76] and [Kan90]), there is no consensus on how to represent, formally, the content of a story. We would have to postulate some abstract representation of the story, using some knowledge representation formalism (e.g. a traditional logic, or something of the “KL-ONE” family [BS85]), based on the ideas proposed in these existing disciplines.

This representation, for generality, would have to contain information not only about events, situations and characters in the world being described; it would also have to represent the narrative presentation (e.g. the order in which events are described), although our simplifications (section 4.4.1 above) might render this rather trivial for the examples to be studied.

4.4.3 Constraining the story

Most AI work on story generation has focussed on methods which would actually construct a story from some basic material, such as information about the characters and their goals. An alternative approach, which we would like to explore, is to attempt to define declarative constraints or preferences describing the formally represented structure of the story. That is, using ideas from
the various disciplines mentioned in Section 4.4.2 above, we would formulate a set of heuristic rules which, given a formally represented story (or candidate for being a story) would evaluate its merit as a story. We do not believe that there is really some abstract measure of "quality" that can be used to rate real stories. However, there are some very crude guidelines which can be used to separate those structures which barely count as stories from those which are at least well-formed stories. The basis for this belief is that literature on creative writing can be viewed as attempts to do just this. For example, when a textbook advises "Never switch viewpoints in the middle of a scene" (e.g. [DiB, Bic83]), it is formulating a heuristic that could be applied to a symbolic representation of a story to make some estimate of its competence or success. (Notice that it is crucial that we have some formal, abstract representation of the story, as it would be completely infeasible to apply such abstract heuristics to the bare textual form of a story.)

A constraint-based (or "evaluative") approach such as this has the advantage that the very heterogeneous requirements of the story pun can all be accommodated. A story pun has certain requirements which might not arise in more conventional story generation:

- The plot must (somehow) conform to (the meaning of) the pre-determined final punchline or "pseudo-moral".

- Enough suitable concepts and vocabulary items must be introduced in the course of the story to make the final punchline relatively natural.

- Nevertheless, for the story pun to be humorously effective, there should not be so much preparation within the story that the punchline can be easily predicted.\(^4\)

All these requirements are of a very varied nature, so some very general scheme is needed for their representation. (We do not suggest that formalising these requirements would be trivial.)

### 4.4.4 Producing the story

Let us assume that we have managed to achieve the various goals listed above - we have selected a narrow, stylised genre of story, we have devised a formal representation for the content (including presentation) of such stories, and we have developed a set of heuristics which will contribute to the rating of how well a story meets our needs. How then can we produce such a story? Although we may have ways of telling how good a story is once we have it fully represented, we need some way of creating the structures. Clearly, the space of possibilities is very large.

Here we introduce another major speculation. Let us assume that our formal representation is such that we can define a compressed encoding of it into a sequence (tuple) of values from some limited range. That is, suppose we can re-code a story-structure into the form

$$\langle v_1, v_2, \ldots, v_n \rangle$$

where each \(v_i\) is selected from some (fairly small) set of possible atomic values (and where the full structure can be reconstructed from this sequence of values if needed). This might well be possible, given the simplifications we have made already. In the narrow, stylised story genres, there are often relatively few options which go to establish the basic parameters of the story.

We are now in a position which is exactly appropriate for the application of a technique known as evolutionary algorithms or genetic algorithms [Mit96]. These methods (a highly sophisticated variant on what is sometimes called "generate-and-test") have proved to be very useful in problem areas where the search space for solutions is very large, but there are clear constraints on what constitutes a 'good' solution. Genetic algorithms are based on a loose analogy with the way that evolution can occur through genetic variation and natural (Darwinian) selection. They require the following:

1. a class of structures or items which we wish to produce (e.g. abstract representations of stories), regarded as analogous to organisms in an evolutionary setting;

2. some measure of "fitness" which rates how good an item is (e.g. our heuristics of story quality), and hence whether the item should be given preference in a "survival of the fittest" process.

3. a way of encoding an item as a tuple, so that the tuples can be treated as analogous to chromosomes and can be mutated or combined in ways which regard them simply as sequences of symbols.

\(^4\)It is easier to succeed with a story pun if the audience does not know in advance the approximate shape of the punchline; it is a sign of the skills of Muir and Norden [MN91] that their stories amuse even when the original (undistorted) proverb is known in advance.
4. a way of decoding a tuple as an item, so that fitness can be assessed after any mutation or combination takes place.

This technique works in the following way. Initially, a large number of such tuples (the pseudo-chromosomes) are generated at random (perhaps within some very minimal constraints on well-formedness). Then the following cycle is repeated many times.

1. Each item is evaluated for its "fitness", usually by decoding the tuple representation (chromosome) into the full structure and then applying various evaluation functions. This gives some rating of the relative "goodness" or "fitness" of the various items.

2. Some manipulation of the tuples occurs, based on their fitness, in a manner analogous to mutation or sexual reproduction in genes. For example, the two "fittest" tuples might be combined (reproduction); or a single tuple might have a small number of its elements changed randomly (mutation). This creates a new "population" of chromosomes.

3. The cycle then repeats.

This process can be run either for some predetermined number of cycles, or until some criterion is met (e.g. that the fitness of the best item has not improved on several successive cycles).

This computational method, although over twenty years old in its conception [Ho75], has become feasible only recently with increasing computer power. It has worked with extraordinary effectiveness in certain areas where the problem has a large search space, a clear notion of evaluation of "goodness", and the possibility of a quasi-chromosomal encoding. (See, for example, [BN95] for a range of applications, including scheduling, layout, stock control, financial planning, traffic management.) Whether it would serve the cause of story generation as well remains to be seen.

4.4.5 Worked example

Work in this area has not advanced far enough to give even a tentative representation of the story, let alone in a form appropriate for use with genetic algorithms. Instead, we will informally outline some of the constraints from the genre and from the parsed punchline, and give a story that satisfies those constraints.

Our chosen genre, the moral fable, has a typical setting, plot structure, cast of characters, and prop list. Although few of these are necessary features, they help to make the genre recognizable, and are not overly restrictive.

- **Characters**: All characters are either animals (e.g. fox, sheep etc) or human. If human, the character must have one of a limited number of occupations (e.g. farmer, doctor, fishwife etc). Both animals and humans have their own simple characteristics, beliefs, goals and relations (e.g. fox is cunning, sheep is stupid, fox wants to eat sheep, farmer wants to protect sheep etc). These characteristics do not change over the course of the story (i.e. there is no character development). Any character can converse with any other character. Characters can lie.

- **Objects**: Each character is associated with a small set of objects. Some objects may be associated with more than one character (e.g hay is both eaten by sheep and mown by farmers).

- **Actions**: Some acts can be achieved by any character (e.g speaking), others cannot (e.g. only farmers can mow, and only when they have a field of hay).

- **Setting**: There is a limited number of settings (e.g. the forest, a farm, the city, a house etc). Each character is typically found in one setting, and can venture into a few others. For example, a fox is typically found in the forest, can go onto a farm, but would not be found in the city. This constrains the characters that can interact.

- **Plot structure**: A character attempts to do something (consistent with its goals). It fails. The reason for its failure, and/or advice on how to avoid a similar failure, is summarized in the moral.

The punchline also constrains the story, in that the fable must justify the punchline as its moral. We assume that the system has access to lexical information. Our chosen punchline, "Book before you reap", might constrain the plot structure of the fable as follows:

A character attempts to reap something, and fails. The reason for its failure is that it failed to book some location before attempting to reap.
From the cast of characters, objects, settings etc, we select those that fit into this outline. Let us assume that, in 'fable world', only farmers can reap, and that the only object that can be reaped is hay. There are no constraints from the punch-line on the location, so we choose one consistent with farmers, hay and reaping: a field. This gives us the basic plot:

A farmer tries to reap hay, but fails. The reason for his failure is that he did not book a field before attempting to reap.

Although we suggest that the generation of the final version of the story could be done with some suitable search mechanism, such as genetic algorithms, here it is done by hand.

Farmer Joe needed some hay to feed his sheep. He picked up his scythe and went to the field to cut the hay. Unfortunately, when he got there, he found another farmer already working in the field. He had forgotten to make a reservation!

Moral: Book before you reap.

Note that the character has been given a name, a few objects and characters (the sheep, the scythe, the other farmer) not strictly necessary to the story have been introduced, and none of the key words in the punchline are explicitly mentioned in the story. Also, the style is relatively sophisticated, using punctuation, adverbs, etc. A good story generator, whether based on genetic algorithms or not, must be able to manipulate the basic story in ways consistent with 'good' creative writing heuristics, as discussed in Section 4.4.3 above.

5 HOW FEASIBLE IS THIS?

Let us summarise what would be needed to make this work.

Punchline construction. This initial phase of the process seems straightforward. Given some basic lexical information (including phonetics), distorting a sentence into a similar-sounding sentence is not difficult.

Punchline analysis. It is completely possible, in theory, to analyse a sentence of English into a representation of its literal meaning. In practice, it would (in the current state of the art) not be trivial to construct a working analyser which would provide correct analyses of arbitrary sentences with any level of syntactic complexity, formulated from a wide vocabulary. Although we are assuming that the story domain could be simplified, the punchlines come from a very unconstrained source, namely existing proverbs and sayings. Not only does this widen the vocabulary greatly, it could introduce some quite unusual grammatical forms.

Representing the story. To do this completely would require significant research, but (given the use of a narrow, stylised genre), it should be feasible to develop some (comparatively crude) form of representation. It has to be recognised, however, that taking short-cuts here could have a deleterious effect on the story-evaluation issue.

Evaluating a story This is the most difficult (and perhaps most interesting) part. It would involve significant research, with a synthesis of ideas from a number of subdisciplines.

Producing the story. If the previous steps, particularly the formulation of an evaluation measure, are achieved, then the use of genetic algorithms should be feasible. The only further prerequisite is the design of a quasi-chromosomal (tuple) encoding for story structures.

6 DISCUSSION

Let us imagine that we manage to construct a program along the lines outlined here. What would the import of this be for a (computational) theory of humour?

What is noticeable about the proposed design is that none of the stages purport to embody or rely upon a theory of humour in any way. Each is a non-humorous process. It is the overall effect which (we hope) will be humorous. This is not paradoxical, as it is quite acceptable for the whole to create an effect that is not attributable to any one of its parts. However, nothing in the design process was driven by any theory of what was or was not funny. Rather, it was based on observing regularities in a process which was deemed humorous, and attempting to model these regularities (cf. remarks in Section 3 above on the
"bottom-up" approach). If it were successful, we would have created a model of a class of jokes without explicit use of a theory of humour. Hence, it is not clear what such a working program would prove or disprove. It may be an experiment, loosely speaking, but what hypothesis does it test?

It is important to realise that a computer program (or even an algorithmic but unimplemented model) may embody theoretical claims or assumptions, even although these are not explicitly represented and even although the system designers were not consciously aware of them. Any design has implicit assumptions, and these may depend upon some unarticulated theory of how humour works. It is commonplace for novice research workers in artificial intelligence (typically, students) to build computer programs without a clear abstract or theoretical model, and to try (with difficulty) to determine afterwards what assumptions constitute their de facto theory. We are perhaps in that position. We have to confess ignorance of what theory of humour has led to the outline proposal in Section 4 above.

This may be a symptom of the very early stage of the development of formal (or computational) models of humour. At present, we have very little theoretical framework on which to base further steps into the unknown. In this situation, one possible (and perhaps essential) first step is to investigate and formalise what might be termed the major structural aspects of humorous texts. That is, a joke genre (such as story puns) will be based on certain (fairly gross) relationships between formal entities. These can be seen as necessary characteristics for a text to qualify as a member of the genre. However, some of the texts which meet these gross structural criteria will not be very funny at all, whereas others will be extremely funny. Study of why there are these variations in funniness should take us much closer to the essential aspects of humour, and so might seem to be more important than the modelling of the "major structural" aspects. However, the formal statement of the major structural aspects provides a foundation upon which more detailed and subtle studies can be based. Until the undergrowth has been cleared away, it is difficult to focus clearly on the fine detail.

A loose analogy can be made with the role of syntax in a theory of language. Even someone who believes that the most essential aspect of language is its meaning is likely to concede that it is unrealistic to attempt to develop a theory of natural language semantics in complete isolation from syntax. Indeed, the more clearly grammatical issues are understood, the less there will be unwanted distractions in the study of meaning.

Although the methodological points argued above have been stated with respect to a nonexistent program (i.e. the story pun generator whose content we have outlined), they could largely be illustrated with respect to a working program — theirie riddle-generator ([BR94, BR97]). That program generates punning riddles which (at their best) are of a standard comparable to those circulating amongst schoolchildren. It was designed entirely by observing the formal regularities within such riddles, and modeling these patterns in abstract rules. Although it could be argued that the design is influenced by, or illustrates, ideas about ambiguity, any relationship to a general theory of humour is very nebulous. Nevertheless, it clears the way for more fundamental questions to be asked, such as what types of ambiguity enhance humorous effect.

More generally, there is the point that was already made in Section 3 above — we have to try out lots of small scale studies just to stimulate ideas about what the ingredients might be of a final theory of humour.

7 WHAT NEXT?

We have argued that:

1. Story puns form a well-defined genre of joke with characteristics quite different from most told jokes.
2. There are formal regularities in story puns which suggest a generator could be decomposed into independent stages.
3. All of these proposed modules could be constructed, at least in some simple form, using currently known techniques.
4. This outline design, whether implemented or not, has no apparent links to any theory of humour in general, but might be a useful preliminary step in formalising structural aspects of one subclass of joke.

The next steps, for those with time, inclination and funding, would be to try implementing such a system and then to experiment with the various factors involved, to see what makes one story pun funnier than another.
REFERENCES


[Dib] Ansen Dibell. Plot. In [DCT95].


AN ATTEMPT AT NATURAL HUMOR FROM A NATURAL LANGUAGE ROBOT

Dan Loehr
Georgetown University
Hughes Information Technology Systems
loehrd@gusun.acc.georgetown.edu

ABSTRACT

This study builds on an earlier integration of a punning riddle generator with an NL "robot", who converses with users in a virtual environment on the Internet [Loehr 96a]. Although this integration allowed the robot to easily generate humor, the humor lacked both relevance to user's input and conversational flow. The current study attempts to correct these shortcomings. To address topical relevance, more conversational key words and their synonyms were identified, and a newer version of the pun generator was tested using a larger lexicon, with the idea of widening the scope of possible responses. To address conversational flow, the rather awkward question structures of the punning riddles were regenerated by the robot as declaratives. It was found that none of these approaches was successful. The methods used to improve relevance actually worsened the situation, by returning too many extraneous responses, while slowing the robot down past normal response times. The methods used to improve conversational flow still produced stilted discourse. The goal, therefore, of more natural-sounding humor based on arbitrary user input was not realized.

1 INTRODUCTION

Humor is traditionally regarded as a human capability. It has been viewed as something which separates humans from machines, as the latter have not typically been able to produce good humor. In the popular movie, "Star Trek: Generations", the humanoid computer "Data" is unable to keep up with his human counterparts in humorous dialogues (which, ironically, we humans find all the more humorous). In spite of his linguistic skill, Data's inability to joke reminds of his non-humanity.

It has been suggested, then, that the ability for natural language (NL) systems to generate humor would make them more human-like [Mauldin 94]. In an earlier study [Loehr 96a] integrated an existing pun generator, JAPE (Joke Analysis and Production Engine), with an existing NL system, Elmo, a robot in a virtual environment on the internet. That study's goal was not to test whether humor did indeed render NL systems more human-like. Rather, it sought to show that is was possible to implement an NL system which could spontaneously generate humor. Although the study succeeded, it was noted that the humor generated was not "natural", which we informally define for this paper as "being relevant to the topic, and conversationally fluid". This paper describes an attempt to generate natural humor from our natural language robot.

Sections 2 and 3 below describe Elmo, the robot, and JAPE, the humor generator, respectively. Section 4 recaps specific details of the first Elmo/JAPE integration. Section 5 reviews the earlier work's recommendations for improvements, and then discusses this study's implementation of those recommendations. Section 6 summarizes the findings, and ponders future directions.

2 THE NL SYSTEM: ELMO, THE NATURAL LANGUAGE ROBOT

A discussion of the NL robot "Elmo" requires a brief discussion of his environment, MondoMOO. (For a full story of how Elmo and MondoMOO came to be, the reader is referred to [Speers 95].) Elmo "lives" in MondoMOO, a type of text-based virtual world accessible from the internet. Users log in to the MOO, and interact with other users and the environment. They may move from place to place, examine objects or create new ones, or talk with other "players" logged in by typing input to them, and reading their responses. Virtual environments
such as MOOs have been described as a cross between a game of "Dungeons and Dragons" and a conference center. MOOs (which stand for MUD-Object-Oriented) developed as object-oriented versions of MUDs, which themselves stand for Multi-User Dungeons.

Elmo is a player in MondoMOO, albeit not a human one. He is a computer program, written in the C programming language. He was developed by Michael Mauldin of Carnegie-Mellon University to live in MUDs, and modified by d'Armond Speers of Georgetown University to live in MOOs. He is programmed to do nearly all the things a human player can do on the MOO: log in, move about, examine objects, and, in particular, converse with other players.

Elmo's original conversational ability was entirely based on pattern-matching. That is, if he detected an input pattern of the form "Thank you", "Thanks", or any variation thereof, he responded with "You're welcome." A team of Georgetown researchers led by Dr. Catherine Ball, and including Speers and Loehr, has augmented Elmo's pattern-matching ability with a formal natural language processing (NLP) system. This system, in Prolog, is capable of processing input strings to determine their meaning. Thus, input of "Thank you" would be processed by the NLP system, the meaning would be returned to Elmo, and he would produce appropriate output (again, "You're welcome"). Currently, Elmo's conversational ability is hybrid. His input is first handled by the NLP system; if that fails, he falls back to pattern-matching.

Elmo is surprisingly robust. Though he occasionally must handle "unknown" input with the fall-back reply of "I'm sorry, I don't understand", he is usually capable of providing an appropriate response. Even if he doesn't know an answer, his clever way of rephrasing the question lends him a certain amount of "intelligence", as the examples below show. (These are from the MOO, which echoes the questioner's words so all may "hear" them):

"Elmo, is there a schnoodle here?"  
You say, "Elmo, is there a schnoodle here?"  
Elmo says, "I don't know about any schnoodle here, Dan."

"What is the name of Fred's dog?"  
You say, "What is the name of Fred's dog?"

Elmo says, "I don't know what the name of fred's dog is, Dan."

As might be expected, Elmo is well suited for providing assistance, and is amply programmed with answers about all aspects of the MOO. In particular, he can help users with the operational tasks of navigating the MOO: how to move around, how to speak with others, how to handle objects, or, most frustratingly, how to log out:

"Elmo, how do I quit?"  
You say, "Elmo, how do I quit?"  
Elmo says, "Type @quit, Dan."

In fact, the original goal of Dr. Ball's team was to move Elmo on to a MOO such as SchMOOze University, which is designed for students learning English as a Second Language. Here, students from around the world may log in and practice their English skills by interacting with each other and their environment. It can be particularly frustrating for anyone, let alone those learning English, to guess that the way to log out is to put the unintuitive "@" sign in front of the word "quit". It would be much easier to place a natural language question to the human-like Elmo, in order to get the answer.

The explanation of Elmo's name, by the way, involves discussion of his heritage. As communication in MUDs/MOOs is by text, one cannot visually or auditorily verify whether a conversational partner is indeed human. The evolution of these virtual environments facilitated the creation of computer-controlled players, called 'bots' (for "robots"). Mauldin created one such "ChatterBot" named Julia [Mauldin 94], who lived in a MUD called TINYMUD. She was described as an "automatic conversation agent on the net, imitating an attractive female user. It was observed from the dialog log that some users who talked to Julia made serious effort to date her!" [Klano 94]. Mauldin then released Colin, a "generic" version of Julia, for use by others in creating MUD robots. It was Colin who was modified by Speers to live in a MOO, and subsequently outfitted with NLP abilities. Speers named his robot "Elmo", after the muppet on the children's television show Sesame Street. The muppet Elmo is meant to be childlike, with childlike language skills. As the robot Elmo's NLP skills were not very sophisticated, it was thought to be a fitting name.
3 THE HUMOR SYSTEM: JAPE

The humor system used in this study is JAPE-1 (Joke Analysis and Production Engine, version 1), developed by Kim Binsted and Graeme Ritchie at the University of Edinburgh. JAPE, a Prolog program, generates punning riddles. Though "Analysis" is in its title, JAPE does not analyze jokes in any way - this refers to the fact that JAPE implements an underlying linguistic model, discussed briefly below. (For a full description of JAPE, the reader is referred to [Binsted and Ritchie 94]. The discussion below follows this work closely.)

From the broad range of humor, Binsted and Ritchie focus on one particular type: the punning riddle. A classic example of this is:

What has four wheels and flies?
_A garbage truck._

The authors chose to investigate punning riddles for several reasons. The linguistics of riddles has been studied [Pepicello & Green 84], there is a large existing body of them [Webb 78], and they consist of regular structures, making them easy to model and to hopefully generate.

The authors describe Pepicello and Green’s view that humor is often related to ambiguity, which be of several types. Phonological ambiguity can result in

What bird is lowest in spirits?
_A bluebird._

[PG 84]

as the phonological entity [blu] can refer to either the color "blue" or the feeling "blue". Morphological ambiguity can produce

Why is coffee like soil?
_It is ground._

[PG 84]

because "ground" can be a past participle of "grind", as well as a synonym for "soil". Finally, syntactic ambiguity enables the riddle

Would you rather have an elephant kill you or a gorilla?
_I'd rather have the elephant kill the gorilla._

[PG 84]

since the question can have two syntactic structures. Of the three types discussed, Binsted and Ritchie chose to concentrate on phonological ambiguity, as they appeared to the simplest to generate computationally.

Punning riddles of can be further subdivided. For instance, ambiguity can occur in the question (as in the "elephant/gorilla" riddle) or in the punchline (as in the "bluebird" riddle). The punchline can be a noun phrase (as in "A bluebird") or a sentence (as in "It is ground"). Further, the ambiguity can be exploited in three main ways. There is syllable substitution, as in

What do short-sighted ghosts wear?
_Spooktacles._ [Webb 78]

or with word substitution, as in

How do you make gold soup?
_Put 14 carrots in it._ [Webb 78]

or with a reversal of sounds in words, called metathesis (or "spoonerisms"), as in

What is the difference between a donkey and a postage stamp?
_One you lick with a stick, the other you stick with a lick._ [PG 84]

Of these choices, the authors chose to focus on word-substitution, and then only on noun phrases in the punchline. In addition to simplifying the scope of the problem, this allows them to use readily available lists of homophones, or phonologically identical words, in their pun generation. In fact, it is the substitution of homophones into noun phrases which is the key to JAPE’s pun generation.

At the risk of doing injustice to Binsted and Ritchie’s excellent explanation of JAPE’s theory and implementation, here is an attempt to summarize its workings.

First, JAPE relies upon a lexicon of "lexemes", or abstract entities roughly corresponding to a meaning of a word or phrase. There is one entry per lexeme, so that if a word has two meanings, it will have two lexemes. Lexemes may have relationships between them, such as "subclass", "superclass", "homophone", "synonym", etc. It is important to note that JAPE’s lexicon is general-purpose, and was not designed with humor generation in mind. Any humor arising from it is teased out of it by JAPE, by using homonyms, synonyms, etc., to create ambiguity as described above.
JAPE further uses a set of "schemata", or structures of relationships between lexemes, to build a new, "fake", noun-phrase. This noun phrase will have the same form as an existing noun phrase in the lexicon, but with new meaning, derived from relationships based on a homonym. As one noun phrase will have two meanings (the original and the newly constructed one), ambiguity results. Then, JAPE employs "templates", which use those meanings to build a question with the "fake" noun phrase as its answer. Finally, the question and answer are converted to a surface form, and output as the riddle.

As an example, consider the riddle,

What's green and bounces?  
A spring cabbage.  

[Webb 78]

In the lexeme would be the following entries, along with their meanings, as well as any homophones, synonyms, descriptions, component parts, etc.

- spring_cabbage (springtime vegetable)
- cabbage (vegetable)
- spring (season of the year)
- bounce (to jump)
- green (the color).

JAPE would attempt to fit any relations between these lexemes into a schema, or structure, suitable for ambiguity. For instance, it first finds a noun phrase: "spring_cabbage". JAPE would know that "spring_cabbage" is made up of "spring" (the season) and "cabbage" (the vegetable). It would find the homonym "spring" (to jump) for the first of these. Taking this, it would find the synonym "bounce". JAPE would then find the adjective "green" for "cabbage". Using these, it now has a new, "fake" noun phrase. This noun phrase looks the same as what we started with ("spring_cabbage"), but now it has a different meaning: "green bouncy thing".

This new meaning is then molded into a template to produce a question. In surface form, the question becomes "What's green and bounces?" The answer is the ambiguous noun phrase: "A spring cabbage".

Though the overall level of JAPE's puns is not high, it is impressive that it can generate these out of a general-purpose lexicon. And the best of its output are actually quite good, yielding riddles such as:

What kind of murderer has fibre?  
A cereal killer.

What do you call a reverent numeral?  
An avowed number.

4 THE INTEGRATION OF ELMO AND JAPE

Before discussing the specific integration of Elmo and JAPE, let us look at general issues in adding humor to an NL interface, as highlighted in [Binsted 95].

First, humor can alleviate known problems with natural language interfaces. As Binsted puts it, NL systems have "limited NL coverage, limited knowledge, and 'bad manners'". It is difficult for any one NL system to parse all possible natural language input. Even if successful, the system may not be able to generate the proper reply due to inadequate knowledge of the input domain. In addition, the system's responses may appear irritating, frustrating, or repetitive to the user.

Humans often use humor (sometimes self-deprecating) to soften failure. An NL system could do likewise. For instance, our robot Elmo might say, "Sorry, I don't know. I'm only a robot, and a virtual one at that!" This might be more refreshing than an endless refrain of "I don't understand". One must be careful, however, as unwanted jokes only make the interface more irritating.

Another general issue has to do with exactly how the humor is to be produced. On the one hand, it could be generated from a linguistically motivated implementation. This is what Binsted and Ritchie have attempted to do, based on Pepicello and Green's theory of ambiguity and humor. On the other hand, the humor could be more "hard-coded", or deliberately crafted to be effective on the surface, but without underlying linguistic motivation.

Which approach is "better" can be moot, from the user's point of view. To him or her, it matters little what is behind the scenes, as long as the result is effective. In fact, Elmo himself is an example of this dichotomy. On the one hand, he first processes input using a linguistically motivated natural language system. Yet if this fails, he uses simple
pattern-matching. In fact, Elmo’s pattern-matching is currently more robust than his NLP system.

A more famous example of this issue is the Turing Test, in which Alan Turing proposed to replace the question, “Can machines think?”, with the question, “Can a machine fool a human into thinking that the machine is human?” [Turing 50]. Turing mentioned nothing about the underlying workings of such a system, or whether it emulated human thought processes. He proposed, rather, that if a machine could pass as intelligent, then it should be considered intelligent. It is perhaps no accident that Elmo’s predecessor, Julia, has been a successful contestant in modern judged Turing Tests.

The problem with hard-coded responses, however, is that they only go so far. A finite supply of responses can never match an infinite number of inputs. It is for this reason that we have tried, in this research, to use methods which will allow response to any arbitrary input.

Let us now recap the particular details and results of the first Elmo/JAPE integration.

Elmo, version 1.0, is written in C, and is running under UNIX and SunOS 4.1.3 on a Sun SparcStation. Its original “robot” model, as developed by Michael Mauldin, is Maas-Neotek, release 11. MondoMOO, Elmo’s home, was compiled from the LambdaMOO code, version 1.7.8p4, and is running on the same SparcStation.

JAPE-1 (version 1), is written in Prolog, and is running on the same SparcStation as Elmo. Though written under Siestus Prolog, for this implementation JAPE is running under Beta-Prolog 1.2(C). The lexicon and other data files were exactly as delivered from JAPE’s developers. Using these files, JAPE is capable of producing around 200 punning riddles. Depending on the input given to JAPE, it can produce puns based on a specific noun phrase, schema, or template. Alternatively, it can simply generate every pun possible from its data files.

The first integration hurdle to overcome was getting Elmo’s C process to communicate with JAPE’s Prolog process. This was accomplished through a third piece of software, expect 5.15 [Libes 95], running on top of TCL (Tool Command Language) 7.3. expect is designed to automate user interfaces to interactive programs. For instance, Prolog is an interactive program, typically receiving user input from the keyboard, and displaying results to the screen. expect can automate this interaction with a script, which feeds input to prolog as if a human were doing so, and likewise monitoring results from Prolog. As far as Prolog is concerned, expect is functionally equivalent to a human being.

By linking with expect libraries, C programs can also feed input and receive output to/from Prolog. For instance, to run JAPE, the C code can spawn Beta-Prolog, then issue it a command to load the JAPE files, then issue it Prolog queries to generate puns, and finally read the response from Prolog (the puns themselves). This query/response loop can continue as long as desired, after which the C code can shut down Beta-Prolog (and JAPE) by issuing the “halt” command.

As an aside, this is the same mechanism used by the C code to communicate with Elmo’s NLP system, also written in Prolog. expect allows the C code to interact with multiple processes simultaneously, so for modularity, a separate Prolog process is spawned for both JAPE and the NL system.

Once Elmo gained the ability to receive puns generated by JAPE, the next question became: Exactly when, and how, should Elmo use JAPE’s puns? This led us to four levels of integration.

The first level of integration is the simplest. Upon a request such as “tell me a riddle”, Elmo issues a query for a pun from JAPE, which he then repeats to the user (after issuing a disclaimer that the pun may not be very good). A randomizer is used in the C code, so that the same pun (and disclaimer) will not be produced every time. The results might look like the following:

"Elmo, can you tell me a riddle?
You say, "Elmo, can you tell me a riddle?"
Elmo says, "I only tell bad puns, Dan, but here goes."
Elmo says, "What do you call an initial fish?"
Elmo says, "a first bass."

"Can you tell me a pun?
You say, "Can you tell me a pun?"
Elmo says, "OK, but it won’t be a good one.
Dan."
Elmo says, "What kind of rain has presents?"
Elmo says, "a bridal shower."

"Do you tell jokes?"
You say, "Do you tell jokes?"

Elmo says, "Well, if you really want me to, Dan, I'll tell you one"

Elmo says, "What kind of jewelry can swim in a lake?"

Elmo says, "a cygnet ring."

"Can you tell me another pun, Elmo?
You say, "Can you tell me another pun, Elmo?"
Elmo says, "Well, if you really want me to, Dan, I'll tell you one"

Elmo says, "What kind of road has weapons?"

Elmo says, "a duel carriageway."

At this level of integration, Elmo is simply a front end for JAPE. Admittedly, it does make him a bit of a raconteur, able to produce humor on demand. Yet, by simply requesting a random pun from JAPE, Elmo has no way of making his humor at all relevant to what the user may have said. Were he to sput off a random pun on arbitrary user input, we might get the following hypothetical interchange:

"Elmo, is Randalf logged in?
You say, "Elmo, is Randalf logged in?"
Elmo says, "What kind of road has weapons?"

Elmo says, "a duel carriageway."

This brings us to our second level of integration - an attempt to make Elmo produce humor relevant to arbitrary user input. At this level, the C code searches JAPE's data files for certain keywords, based on user input. If such a keyword is found in the noun phrase punchline of any pun, then a request is issued to JAPE for the full pun. (Originally, the search was extended to all parts of the pun, not just the punchline. However, this caused an unacceptable delay of many seconds before the pun was ready). Elmo then issues the pun to the user.

An example of this first attempt at relevance is in the following interchange ("look" is a MondoMOO command):

look BabyDog

You see an intelligent-looking little grey schnoodle.
She looks friendly. Why don't you pat her?

"Elmo, will the dog bite me?
You say, "Elmo, will the dog bite me?"
Elmo says, "Hmmm..."

Elmo says, "What do you call affection that is used to store data?"

Elmo says, "a love byte."

In this example, the C code searched JAPE's data files for the keywords "dog" and "bite". It found a pun based on the noun phrase "love, bite", requested the full pun from JAPE, and repeated it to the user.

There are several problems with this attempt at humor. First of all, the humor is unexpected. The user expects a "yes/no" reply, and Elmo responds with a joke. To ease the sudden transition, Elmo was programmed to say "Hmmm..." first. Though this was not very successful, the idea was to give the user a subtle warning, through Elmo's stalling, that perhaps the answer would not be a simple yes or no.

The second problem, of course, is that the joke is barely relevant. Admittedly, it uses the homophones "bite/byte", but even if the user were expecting a joke, they would not expect one about computational affection when they had asked about a dog's temperament.

Another example of this attempt at relevance is shown below. Elmo is programmed to give assistance on a wide variety of topics, ranging from how to move about to how to obtain the source code for a robot. If one were to ask for further help, Elmo was programmed to reply, "I've run out of answers". In an attempt to add humor to this failure, a JAPE data search was done on the keywords "help", "aid", and the homophone "ade":

"Do you give help on building robots?
You say, "Do you give help on building robots?"
Elmo says, "The source code for Maas-Neotek robots is available on host NL.CS.CMU.EDU [128.2.222.56], directory /usr/mlm/ftp, file robot.tar.Z. set mode binary first, and cd to /usr/mlm/ftp in one step, Dan."

"Got any other help?
You say, "Got any other help?"
Elmo says, "I've run out of answers"
Elmo says, "What do you call a useless drink?"
Elmo says, "lemon ade."

Again, the humor is rather unexpected, and only slightly relevant. It is a bit of a stretch for the user to realize that Elmo is depreciating himself, by
calling himself useless. And the "drink" part of the
pun has no relevance whatsoever.

Recall that earlier we discussed the use of
linguistically motivated algorithms, vs. deliberately
crafted "tricks", to achieve a desired effect. These
first two levels of integration have belonged to the
former category. In the first level, Elmo simply
repeated puns generated purely on the theoretical
grounds underlying JAPE. In the second level,
Elmo did so as well, with the only addition of trying
to select appropriate puns based on keyword
searches. While interesting, neither of these
methods is entirely satisfactory from a user interface
point of view. This brings us to the next two levels
of integration, which rely on some manipulation to
improve the user interface.

The third level of integration again takes place
in the above "help" event. If, after issuing the
above help request, the user asks again for help, they
receive the following reply:

"No more help for me?
You say, "No more help for me?"
Elmo says, "I don't know what else to say,
Dan."
Elmo says, "What do you call a useless
assistant?"
Elmo says, "a lemon aide."

There are two differences from the previous
response. First, the line "I've run out of answers"
has been changed to "I don't know what else to say,
Dan". This has been done by the C code, purely for
variety.

More importantly, the pun used here is much
more relevant. Rather than joking about a useless
drink, Elmo now jokes about a useless assistant,
which is exactly what he is at the moment. How
was this pun selected over the previous one?

The answer is, "by human intervention". The
keyword search used in the former example simply
returned the first pun it found with the word "help",
"aid", or "ade". However, there were a number of
puns generated from the "lemon_aid" noun phrase,
with different levels of quality and relevance. In this
case, a somewhat "better" pun was hand-chosen
beforehand, from JAPE's offerings, to be presented
to the user.

Note that we have sacrificed the ability to
handle arbitrary input from the user. We had to
know ahead of time that in this portion of the code,
the user would be asking for help, and then supply a
pre-chosen pun for just this case. It would have
been simpler to have one general pun-reply module,
which delivered puns for any occasion. Yet, as the
two previous examples showed, it can be difficult to
guarantee quality and relevance.

Finally, at the top level of user interface, we
have the fourth level of integration. If the user were
to ask one last time for further help, they would see:

"That's all the help you have?
You say, "That's all the help you have?"
Elmo says, "Sorry I'm not more helpful"
Elmo says, "I guess all I can offer you is 'a
lemon aide.'"

Though still not wonderful, this approaches
what a human might do in joking about failure.
Again, the punchline ("lemon aide") was generated
from JAPE, as in the previous example. This time,
however, it was inserted into a carefully scripted
reply, to achieve a "smoother" response.

Though these are simple examples, they could
be easily reproduced throughout Elmo's code. On
the one hand, we could have a central, general-
purpose pun-reply module which produced puns on
arbitrary user input. Alternatively, we would need
carefully scripted, pre-selected responses, to handle
specific user utterances.

It is interesting to note that, of the four levels
of integration, the first and fourth appear to be most
successful. In the first, users have no expectation
that Elmo will try to "fit" a joke into another
response. They simply ask for a pun, and he
supplies one. In the fourth level, though the user is
not expecting humor, Elmo more skillfully weaves it
into his reply anyway. It is only in the middle
levels, where the user is not expecting humor, and
yet where it is poorly delivered, that problems occur.

5 ATTEMPTS AT MAKING THE
HUMOR NATURAL

As mentioned, although humor was
successfully generated in the first integration, it
lacked in relevance and flow. Several
improvements suggest themselves.

We hypothesize that greater relevance could be
obtained merely by "casting a broader net" over the
conversation, to find more concepts to joke about. Specifically, we identified more key-words from the discourse, found their synonyms, and matched all of these against a larger lexicon used by the humor generator. The hope was that a greater number of concepts would be selected; the more concepts to joke about, the greater the chance of generating humor relevant to the conversation.

We further hypothesized that a syntactic structure other than the question form of riddles would aid conversational flow. As a simple approach, we converted the questions to declaratives.

The details of the methods used, and their results, follow. Let us start with the problem of relevance.

In the first study, the keywords from user input were identified by hand. We chose nouns and verbs as key-words. For instance, on the sample input Will the dog bite me?, we hard-coded Elmo to send the words dog and bite to JAPE, in the hopes that the humor generator would be able to create a pun based on these key-words. While this worked in the specific case, it would not be possible to hard-code Elmo to search for key-words in the general case of infinite arbitrary input.

Therefore, we consulted a part-of-speech list to determine the syntactic category of each word input by the user. In the spirit of casting the broadest net possible, we chose to use the Moby II part-of-speech list. With 230,000 entries, it is described by its developer, Grady Ward, as the largest part-of-speech list available. Moby is also easy to use, being a simple ASCII list. Searching the list for the unique word bite returns

```
bite VtIn
```

meaning that bite is primarily a verb (both transitive and intransitive), and secondarily a noun. This qualified it as a keyword in our algorithm.

As mentioned, we also used synonyms of the keywords. These were originally chosen by hand as well. To extract them automatically, we used the WordNet 1.5 lexical database by George Miller et al [Miller 93]. WordNet has many uses. It can be thought of as a lexicon, which is how JAPE's newer version, JAPE-2, uses it (described below). Yet WordNet also provides synonyms of words. For instance, asking WordNet for synonyms of the verb bite returns

```
3 senses of bite

Sense 1
bite, seize with teeth
=> grip, hold fast, hold firmly
Also See-> bite off, bite at

Sense 2
bite, sting, burn
=> ache, smart, hurt

Sense 3
sting, bite, prick
=> pierce, make a hole into
```

We now needed to extract the keywords out of this response, else function words might be though of as key. Again, we used the Moby part-of-speech list to sift out the nouns and verbs.

We have so far discussed details of extracting more keywords, and their synonyms, from the discourse. We also hoped that the humor generator could use a larger lexicon, providing a larger variety of words to choose from in generating jokes. For this, we planned to use WordNet as a lexicon. Whereas JAPE's original lexicon had several hundred entries, WordNet contains over 95,000. Since the first version of JAPE did not interface with WordNet, we also needed to integrate Elmo with the new version, JAPE-2.

The results of our attempts to improve relevance were disappointing. First of all, using Moby proved too slow for Elmo to use "natural" conversational response times (for the purposes of this study, we defined "natural" response times as the expected time it would take a user to type the response: between ten and twenty seconds). Since the entire piece of the robot was to converse in a human-like fashion, it had to operate at human-like speeds. As Moby had nearly a quarter million entries, it took even the fast UNIX utility egrep ten seconds to search the list on a Sun Sparc4. At ten seconds per word, even the short five-word sentence Will this dog bite me? would take 50 seconds just to identify the keywords. Moby's vast coverage was paid for in speed. Section 6, however, discusses some ideas for solving this problem.

In terms of speed, WordNet was much quicker, returning synonyms almost immediately. (Let us leave aside for the moment the problem of using the slow Moby to glean the nouns and verbs from
WordNet's synonym definitions). Yet in coverage, WordNet was too thorough, returning senses which turned out to be irrelevant to the discourse. As an example, we used the synonyms of both the verb and noun senses of *bite*, since Moby informed us that this word could be both. The verb's synonyms were given above; the noun's are given here:

6 senses of bite

Sense 1
bite
   ⇒ wound

Sense 2
morsel, bit, bite
   ⇒ taste, mouthful

Sense 3
sting, bite, insect bite
   ⇒ injury, hurt, harm, trauma

Sense 4
bite, collation, snack, nosh
   ⇒ meal, repast

Sense 5
pungency, bite, sharpness
   ⇒ spicery, spiciness, spice

Sense 6
bite, chomp
   ⇒ eating, feeding

Note that Sense 3 includes the word *insect*, which is related to *bite*. We thus identify *insect* as a keyword to be passed to JAPE. JAPE then tries to produce a joke somehow related to *insect*. As an example, JAPE-1 comes up with

Q: What do you get when you cross a female relative with an insect nest?
A: an aunt hill.

leading to the following MondoMOO exchange:

"Elmo, will the dog bite me?
You say, "Elmo, will the dog bite me?"
Elmo says, "Hmmm..."
Elmo says, "What do you get when you cross a female relative with an insect nest?"
Elmo says, "an aunt hill."

The problem is that *relative* is somehow related to *aunt*, which is somehow related to *ant hill*, which is somehow related to *insect*, which is somehow related to *bite*; but *relative* is not related to *bite* (let alone to *dog*). Like the children's game of "telephone", the meaning drifts with every layer of processing. The broad net our algorithm cast for relevant concepts can also easily catch irrelevant ones.

As mentioned, WordNet is also used as a lexicon, by the newer JAPE-2. It was hoped to test out the WordNet lexicon, but attempts at getting WordNet to run on the same system as Elmo were unsuccessful for the following reason. To check synonyms, we used the C-language version of WordNet, which loaded fine. However, for use as a lexicon, JAPE-2 requires the Prolog version of WordNet. This would not load under the Prolog used by Elmo's NLP module, due to the large amounts of memory WordNet's lexicon required.

However, given the fact that using WordNet as a synonym source produced too many entries, it is predicted that WordNet as a lexical source would do likewise. Binsted confirms this in regards to running JAPE-2 standalone (i.e. without Elmo), relating that most of WordNet's lexical items are too "obscure" to create meaningful puns [personal communication, May 13, 1996]. JAPE-2, therefore, requires manual filtering of its output, using the MRC psycholinguistic database. Puns with words scoring low on concreteness, familiarity, or imageability are weeded out by hand. In addition, Binsted reports that JAPE-2 is much slower than JAPE-1 [personal communication, June 20, 1996]. This alone might have slowed Elmo down past expected response times. In sum, JAPE-2 with WordNet is too far-ranging, and too slow.

We have so far discussed attempts at improving relevance in Elmo's humor. Let us now discuss our approach for improving his conversational flow.

As mentioned, our approach was to convert JAPE's riddle surface structure to a simple declarative. Recall that Elmo gave the following response to a request for help:

"No more help for me?
You say, "No more help for me?"
Elmo says, "I don't know what else to say, Dan."
Elmo says, "What do you call a useless assistant?"
Elmo says, "a lemon aide."
Recall also that we manually manipulated the above riddle to the somewhat more fluid response:

"That's all the help you have?  
You say, "That's all the help you have?"  
Elmo says, "Sorry I'm not more helpful"  
Elmo says, "I guess all I can offer you is 'a lemon aide.'"

The question is: can we change the former riddle to the latter "somewhat human" response automatically, without human intervention? As a test, we extracted the pre-surface form of JAPF's riddle, getting the "pieces" of the joke. For instance, the important pieces of the above pun are useless assistant and lemon aide. Then, noting that Elmo's Prolog parser is reversible, we simplified part of it to generate a simple declarative out of the pieces of the joke:

"That's all the help you have?  
You say, "That's all the help you have?"  
Elmo says, "Sorry I'm not more helpful"  
Elmo says, "A useless assistant is a lemon aide."

Again, our ideal was to have Elmo say, "I guess all I can offer you is a lemon aide." Our actual reply was the simple "A useless assistant is a lemon aide". While the latter has a zen-like, fortune-cookie sound to it, it does not quite work. The reason, we feel, is the structure. A simple declarative is not sophisticated enough to handle all responses. But how to know when to use a declarative, or an imperative, or an interrogative? This would require more sophisticated rule-based algorithms than mere sentence generation. While this would be a fruitful area of research (and may already have been solved), we did not at this time pursue it further.

6 CONCLUSIONS

Binsted and Ritchie have shown that it is feasible to automatically generate a certain type of humor: the punning riddle. Integration of this pun generator with a natural language interface, while successful, revealed some difficulties in producing relevant, fluid humor or arbitrary user input. This study attempted to improve relevance by casting a broader net around both the words in the discourse and the words available to the humor generator, hoping to widen the scope of possible responses. Unfortunately, the net proved to be too large and too slow, producing even more irrelevant responses at speeds slower than normal conversational rates. Also failing was a simple attempt at improving fluidity by replacing raw puns with declaratives, in the hopes that these would weave more deftly into the conversation. Although somewhat better, declaratives still fell short of the mark. In sum, our goal of more natural humor from the robot is unrealized.

Some improvements over our current methods come to mind. One stumbling block was the length of time it took to search Moby's part-of-speech list. Being linearly organized, it is extremely inefficient for searching. A tree implementation of the list would speed things up exponentially as [Loehr 96b] has shown in a different domain (that of quickly searching large dictionaries to correct Optical Character Recognition errors). Or, a more sophisticated part-of-speech tagger might be fast enough, rather than searching a linear list.

We mentioned also the problem of our "net" catching too many irrelevant concepts. We could perhaps build a "smarter net". For instance, we used the example of insect being returned as related to keyword bite. The original sentence also had another keyword: dog. It might be better to have our net return words related to at least two of our original keywords (i.e. both bite and dog), instead of one, in the hopes that extraneous items like insect might slip through the net. (Dogs do have fleas, of course, so this algorithm is not fool-proof).

Finally, it would be beneficial to explore some sort of rule-based system for patterning response structures, as mentioned above. For example, one might match questions with declaratives, or imperatives with "OK..." or "I'm sorry, but..." or other more sophisticated ways of framing the humorous content.

There is one final, almost gestural technique to show that Elmo has a sense of humor, one which lends itself well to text-based communication. That is the ubiquitous "smiley", as in:

"Goodbye, Elmo.  
You say, "Goodbye, Elmo."  
Elmo says, "Goodbye, Dan :-)

REFERENCES

Manuscript, available by HTTP: http://www.dai.ed.ac.uk/students/kimb/papers.tar.gz, file ijcai.ps


Loehr, Dan (1996b). Approximate fixed-length string matching for OCR context. Manuscript


Humorous analogy: Modeling The Devil's dictionary*

Cameron Shelley
Department of Philosophy
University of Waterloo
Waterloo, ON
Canada N2L 3G1
cpsbelle@watarts.uwaterloo.ca

Toby Donaldson
Department of Computer Science
University of Waterloo
Waterloo, ON
Canada N2L 3G1
tjdonald@neumann.uwaterloo.ca

Kim Parsons
Department of Computer Science
University of Waterloo
Waterloo, ON
Canada N2L 3G1
kparson@neumann.uwaterloo.ca

Wit, n. The salt with which the American humorist spoils his intellectual cookery by leaving it out.

Ambrose Bierce, THE DEVIL’S DICTIONARY

ABSTRACT

The use of analogy in definitions of Ambrose Bierce’s The Devil’s dictionary is examined by applying the multiconstraint theory of analogy and the General Theory of Verbal Humor. This examination suggests several generalities about the use of analogy in short, humorous texts. A constraint-based approach is discussed for realizing humorous analogies linguistically, focusing on concept selection and lexical choice.

1 HUMOR AND ANALOGY

The idea that humor employs a kind of logic is not readily apparent, but a careful investigation shows that humor is not simply a random or unpredictable facet of human nature. Rather, being humorous is usually a purposeful activity that takes place in many ways and by many different means. Humor may be captured in pictures or in sounds, in long texts or in short. One master of the short, humorous epigram is American author Ambrose Bierce. Particularly outstanding in this regard is his collection of satiric lexical definitions, The Devil’s dictionary [3]. In each definition, Bierce displays both a great economy and crispness of expression, and a variety of linguistic and conceptual devices including puns, metaphors, and analogies. For example, consider the following definition:

Here Bierce constructs a fairly simple analogy between a tailor, who sews clothing with needle and thread, and a blacksmith, who produces iron armor with hammer and tongs. Much of the humor of this definition comes from considering the structural similarity of the activities of tailors and blacksmiths as Bierce describes them. In fact, Bierce is clearly using the analogy as the means of moving the joke along: He primes the reader with the term “armor,” alludes incongruously to the needle and thread of the tailor and finishes with the hammer and tongs of the blacksmith. This sort of threefold pattern, setup, development, punchline, is a well-known structure for jokes [1, pp. 60–107].

There are many more examples in The Devil’s dictionary in which analogy is used to convey humor. This suggests that a study of similar definitions from the dictionary should reveal interesting aspects of the connection between humor and analogy. Such a study is presented in this paper. It is fortunate that, at this time, theories about humor, analogy, and computational modeling have all developed a compatible set of tools for the job. The examination of some of Bierce’s definitions therefore draws on the methods provided within each of these fields of research. Section 2 supplies an overview of the multiconstraint theory of analogy [7]; section 3 demonstrates the application of the multiconstraint theory to the analysis of several examples of Bierce’s definitions; section 4 discusses a computational model of the results of section 3; section 5 presents the summary and points to directions for future work. The theory of humor employed here [12] is explained as required throughout the paper.

*Thanks to Paul Thagard, Chris Eliasmiss, and the two anonymous reviewers for their comments on earlier versions of this paper.
<table>
<thead>
<tr>
<th>Vision</th>
<th>Audition</th>
</tr>
</thead>
<tbody>
<tr>
<td>eye, seeing</td>
<td>ear, hearing</td>
</tr>
<tr>
<td>instrument-of(eye,seeing)</td>
<td>instrument-of(ear,hearing)</td>
</tr>
</tbody>
</table>

Figure 1: A multiconstraint representation of the analogy "An eye is to seeing as an ear is to hearing." Domain titles appear in boldface above their respective domains. Subscripts distinguish occurrences of otherwise identical predicates.

2 THE MULTICONSTRAINT THEORY OF ANALOGY

The multiconstraint theory of analogy [7] provides an explicit and systematic means of identifying conceptual patterns shared between two concepts or domains of knowledge. Parallels, or mappings, between such domains can be described and juxtaposed, allowing the "fit" of an analogy to be assessed. A familiar form of analogy consists of the "proportional" analogy often found in IQ test questions, such as "An eye is to seeing as an ear is to what?" This can be represented symbolically with the notation "A:B::C:X", where "X" stands for the missing term "hearing" [7, p. 28]. This notation relies heavily on the relative spatial arrangement of letters and colons to represent analogical relationships. The analogical rule in this case, that the left-hand side of each domain identifies the instrument of the sense named on the right-hand side, remains implicit in this notation.

The multiconstraint notation, however, relies on making such rules explicit. Thus, the eye-ear analogy can be represented as shown in figure 1. The top box (below the boldface domain titles) gives the basic components of the analogy: eye, ear, seeing, and hearing. Each mapping occupies a single row in the figure. The mappings between these basic elements are called attribute mappings, since they represent the similarity between simple attributes of each domain [7, pp. 24–6]. The next box down in figure 1 gives the rule that relates the attributes, the same rule being present in both domains. Each parallel between such rules is called a relational mapping because it describes relations among the attributes [7, pp. 26–8]. Together, these attribute and relational mappings seem to capture everything significant about the eye-ear analogy.

But the multiconstraint theory is also able to describe more complex analogical structures. Consider the analogy presented in the following verse by Sir John Davies (1570–1626):

Wedlock, indeed, hath oft compared been
To public feasts, where meet a public rout,—
Where they that are without would fain go in.
And they that are within would fain go out.

The analogy suggested here can be represented as shown in figure 2.

There are several things to note about this example. First, a third category of mapping has been introduced in the bottom box of figure 2. This category describes system mappings, so-called because they embody systemic relations between propositions in each domain, rather than relations between attributes [7, pp. 28–31]. System mappings typically identify the reasons why events occur or why relations are true; thus they are often embodied by predicates such as cause, because, and enable. Second, marriage is suggested as the target domain of the analogy, because the term "wedlock" occurs as the topic of Davies’ sentence—that is, sentence-initially. By the same token, feasting is suggested as the source domain because it occurs as the comment of the sentence—that is, non-initially [9, pp. 500–11]. These textually-based assignments are borne out by the purpose of the analogy, which is ostensibly to inform the reader about marriage by comparison with feasting. Third, the similarity of the two domains is most noticeable in the system mappings where identical predicates, namely because, are placed in correspondence. This fact indicates that the analogy in Davies’ verse operates at a sophisticated level of knowledge concerning the two domains. Certainly, the analogy is more sophisticated than the eye-ear analogy presented in figure 1.

In the multiconstraint theory, the quality of an analogy is assessed on the following three criteria [7, pp. 22–38]:

1. Semantic similarity: corresponding predicates are close in meaning, e.g., near neighbors in a semantic network;
2. Structural isomorphism: each predicate corre-
<table>
<thead>
<tr>
<th>Eating</th>
<th>Marriage</th>
</tr>
</thead>
<tbody>
<tr>
<td>feast</td>
<td>wedlock</td>
</tr>
<tr>
<td>rout</td>
<td>*people</td>
</tr>
<tr>
<td>*participants</td>
<td>*couples</td>
</tr>
<tr>
<td>*onlookers</td>
<td>*singles</td>
</tr>
<tr>
<td>*full</td>
<td>*weary</td>
</tr>
<tr>
<td>*hungry</td>
<td>*hopeful</td>
</tr>
<tr>
<td>part-of_{0}(participants,rout)</td>
<td>part-of_{1}(couples,people)</td>
</tr>
<tr>
<td>part-of_{2}(onlookers,rout)</td>
<td>part-of_{3}(singles,people)</td>
</tr>
<tr>
<td>go-out(participants,feast)</td>
<td>end(couples,wedlock)</td>
</tr>
<tr>
<td>become_{0}(participants,full)</td>
<td>become_{1}(couples,jaded)</td>
</tr>
<tr>
<td>go-in(onlookers,feast)</td>
<td>begin(singles,wedlock)</td>
</tr>
<tr>
<td>feel_{0}(onlookers,hungry)</td>
<td>feel_{1}(singles,hopeful)</td>
</tr>
<tr>
<td>because_{0}(go-out,become_{0})</td>
<td>because_{1}(end,become_{1})</td>
</tr>
<tr>
<td>because_{2}(go-in,feel_{0})</td>
<td>because_{3}(begin,feel_{1})</td>
</tr>
</tbody>
</table>

Figure 2: Davies’ comparison of wedlock to a feast. Identical predicates are distinguished by subscripts, and interpolated attributes are preceded by an asterisk.

sponds with one other predicate, and respective slotholders of corresponding predicates are themselves corresponding predicates elsewhere in the analogy;

3. Purpose: predicates in each domain, particularly at the system level, contribute to the solution of the problem that the analogy is intended to address.

The fitness of an analogy is measured by how well it meets these criteria. Each of these criteria are considered soft constraints, so that the best analogy meets them as well as possible, but may not meet them completely. The analogy shown in figure 2 meets these criteria well, although the assessment is made uncertain by textual considerations, i.e., the number of predicates that must be interpolated from context. (Interpolated attribute mappings are shown with an asterisk in figure 2.)

Davies’ verse forms a suitable primer for Bierce’s definitions. Ostensibly, Davies presents a clarification of the nature of marriage, almost a definition of “wedlock,” in fact. But his true purpose is apparently a display of wit, if not gentle mockery. Bierce’s definitions are very similar in form and purpose, although Bierce is often much more sardonic in tone. Like The Devil’s dictionary definitions, Davies’ verse explicitly conveys just enough information to ensure that the analogy is properly understood, and without going into detail that would make the humor tedious and ineffective. It is also worth noting that an important component of Davies’ analogy is the comparison between marriage—a cultural institution—and eating—a biological function. As such, the verse can also be understood as presenting a parody of marriage: Feasts retain the form of marriage (by virtue of analogical construction) while debasing its content (mapping the operation of a social institution to appetite).¹

As section 3 makes clear, this is an important component of Bierce’s humor as well.

3 Example definitions

This section presents analyses of a number of definitions from The Devil’s dictionary, definitions which exemplify the humorous use of analogy. Before proceeding with the analyses, it is appropriate to establish the analytic parameters to be used. The General Theory of Verbal Humor (GTVH) [1, pp. 222–9] (an extension of the Semantic Script Theory of Humor [12]), proposes the following parameters for understanding humorous texts:

1. Language (LA): the way in which the text is verbalized;
2. Narrative Strategy (NS): the narrative organization of the text;
3. Target (TA): the “butt” of the joke perpetrated by the text;

¹In comedic terms, a parody retains the form of its object while debasing its content. A travesty, conversely, retains the content while debasing the form, as in a burlesque. See [9] for further discussion.
4. **Situation** (SI): the setting and "props" referred to in the text;

5. **Logical Mechanism** (LM): the means by which the sense of the text is developed;

6. **Script Opposition** (SO): the oddity or incongruity used in the text to establish its humor.

The remainder of this section presents a number of analyzed examples from *The Devil's dictionary*, forming a kind of inductive argument about how these parameters, mainly LM and SO, may be applied. To avoid confusion, however, it is worth stating some of the basic conclusions at the outset, so that the presentation can be made as briefly as possible. Therefore, we begin by stating the following four conclusions: Other things being equal,

1. The textual topic (see section 2) identifies the *target* domain of the analogy;

2. The target domain of the analogy is identical with the target (TA) of the humor;

3. Analogical parody functions by
   - emphasizing an important similarity between two domains, or
   - ironically emphasizing an important *dissimilarity* between two domains;

4. The presence of irony excludes the target assignment given in rule 1, effectively assigning the textual topic to the *source* domain of the analogy.

### 3.1 **Academé**

For the first example, consider the following pair of definitions of *academé* and *academy*:

**Academé, n.** An ancient school where morality and philosophy were taught.

**Academy, n.** (from academé). A modern school where football is taught.

This is probably the most explicit analogy Bierce constructs, comparing his terms one after the other, in syntactically similar expressions. The analogy is described in figure 3.

The basic import of the analogy is fairly clear: The modern academic establishment is held to ridicule as a pale reflection of an ancient model. Taken literally, the two definitions seem merely to be true statements about typical schools from the ancient and modern eras. However, the reader is primed for an ironic interpretation of the definitions by the parenthetical note "from academé." In a normal dictionary this note is a conventional device to indicate etymological descent. Here the note simply allows Bierce to pretend that the two definitions are juxtaposed naturally, rather than by his design. The resulting potential for ironic comment [15, pp. 237–43] is exploited to make the academy the true target of the analogy and the humor.

This interpretation is supported by noting the use of the analogy to emphasize a *dissimilarity* between the academé and the academy. Morality and philosophy are largely intellectual disciplines, whereas football is a combative, physical sport. There is the additional difference that the academé is an ancient institution, whereas the academy is a modern one. The historical replacement of an intellectual pursuit with a physical sport may provide the opposition between domains required by Raskin's theory [12, p. 99], perhaps as a species of the *expected/unexpected* type of opposition. Indeed, the joke is reminiscent of cold-war humor about the U.S. president...
and Soviet chairman playing ping-pong to determine domination of the world. 2

Two further issues about the opposition between the source and target domains are worth pursuing here. First, the source and target domains are opposed, or at least importantly dissimilar, and Bierce's text presents both of them. However, Raskin's theory also requires that the text be ambiguous with respect to which domain is being described. In the current example, this ambiguity is supplied by the ambiguity about whether academia belongs to the source or target domain. Were it not for the irony, this ambiguity would not exist. There is the potential for trouble in applying Raskin's theory in that case. The second issue concerns the metaphorical nature of the opposition between the source and target domains. The academia involves mental activity, whereas the academy most prominently involves physical activity. A common metaphor for describing the mind is as a physical, mechanical system, as in the expression "I'm having trouble switching gears" to describe difficulty in thinking about one thing after having thought about another. The metaphorical relationship between philosophy and football is somewhat similar; at least, that is Bierce's suggestion. The fact that the source and target domains share this relationship gives the analogy an added appearance of being well-motivated—that is, non-arbitrary—and sophisticated. Although the presence of such a metaphoric relation between source and target does not seem to be a requirement of humorous analogy [12, p. 138], it is a common property of Bierce's wit.

3.2 BELLADONA

A similar example to the academe/academy example (section 3.1) is the definition of belladona:

**Belladona, n.** In Italian, a beautiful lady; in English, a deadly poison. A striking example of the essential identity of the two tongues.

The analogy is given in two short, parallel clauses in the first sentence and is shown in figure 4.

The humor of the definition is fairly clear, although quite a bit of material is left implicit: Bierce is aiming a sardonic barb at women. Taken literally, the two remarks are true statements about the meaning of the word belladona in English and Italian, respectively. (Belladonna is also an Italian term for the deadly nightshade and subsequently borrowed into English as such.) The target of the humor is identified by the initial occurrence of "belladona" as a beautiful lady, the Italian meaning of the term. The source is subsequently identified as the English meaning, a deadly poison. The first sentence leaves open the possibility of an ironic interpretation, since the senses of belladona are noted to be related etymologically, like academe and academy. The second sentence, however, appears to cancel any such interpretation, by emphasizing the similarity of the two domains in question. The Italian sense is therefore confirmed as the target of the humor and the analogy.

The humor consists in the implicit comparison between the deleterious effects of poison and beautiful ladies. We know that deadly poisons act to kill people if ingested. Similarly, it is suggested, beautiful ladies act ill effect on their lovers. This situation raises the difficulty noted in section 3.1, namely that there is little ambiguity about the assignment of source and target domains from the text. Since the fit of the analogy appears to be strong, it is difficult to specify in what way the two domains could be considered

<table>
<thead>
<tr>
<th>English</th>
<th>Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>deadly (poison)</strong></td>
<td><strong>beautiful (lady)</strong></td>
</tr>
<tr>
<td>*live (people)</td>
<td>*happy (men)</td>
</tr>
<tr>
<td>*dead (people)</td>
<td>*miserable (men)</td>
</tr>
<tr>
<td>ingest (people, deadly)</td>
<td>love (men, beautiful)</td>
</tr>
<tr>
<td>become (live, dead)</td>
<td>become (happy, miserable)</td>
</tr>
<tr>
<td>cause (ingest, become)</td>
<td>cause (love, become)</td>
</tr>
</tbody>
</table>

Figure 4: Bierce's analogy between poison and beautiful ladies.

---

2 If the ironic use of "from academe" is passed up, then the humor also works reasonably well with the ancient academe as the target. In this case, the humor comes from perceiving philosophy as being importantly similar to football, and therefore not actually very refined.
<table>
<thead>
<tr>
<th>St. John</th>
<th>Commentators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revelation</td>
<td>*commentaries</td>
</tr>
<tr>
<td>St-John knowledge</td>
<td>commentators</td>
</tr>
<tr>
<td>conceal(St-John,knowledge)</td>
<td>nothing</td>
</tr>
<tr>
<td>compose₀(St-John,Revelation)</td>
<td>reveal(commentators,nothing)</td>
</tr>
<tr>
<td>know₀(St-John,knowledge)</td>
<td>compose₁(commentators,commentaries)</td>
</tr>
<tr>
<td>enable₀(know₀,conceal)</td>
<td>know₁(commentators,nothing)</td>
</tr>
<tr>
<td>in-order-to₀(compose₀,conceal)</td>
<td>enable₁(know₁,reveal)</td>
</tr>
<tr>
<td></td>
<td>in-order-to₁(compose₁,reveal)</td>
</tr>
</tbody>
</table>

Figure 5: Bierce’s analogy between St. John and his commentators.

opposite.

The *academy/academe* definition involves a comparison between very closely related domains: schools of different eras. The *belladona* definition provides an analogy with more diverse domains: biochemistry and women. Nevertheless, the analogy is quite coherent, with identical predicates appearing at the relational and system mapping levels. The coherence of the analogy makes its two domains appear quite compatible, which gives the definition one of the requisite conditions for being humorous.

Finally, the definition of *belladona* shares roughly the same metaphorical organization as the definitions of *academe* and *academy*. Love is generally held to exist in the mental or spiritual realm, whereas the action of poison definitely exists in the mechanical realm of chemistry. But love may be metaphorically described as a physical effect, as when one “falls” in love, and lovers may be said to share a certain “chemistry.” This relationship clarifies the identity of the source and target domains, and makes the comparison appear more apt and sophisticated.

3.3 REVELATION

A somewhat different example of a humorous metaphor is given in the following definition:

Revelation, n. A famous book in which St. John the Divine concealed all that he knew. The revealing is done by the commentators, who know nothing.

This definition compares the knowledge and actions of St. John with the knowledge and actions of later readers of his work and is shown in figure 5.

Initially, this definition explains why the name Revelation is appropriate for St. John’s book—that it conveys secret information to the prepared reader. Thus, it appears that the target of the analogy is the book of Revelation. But the final sentence focuses on a significant difference between St. John and the commentators, that the former is knowledgeable whereas the latter are ignorant. This emphasized difference motivates an ironic interpretation of the analogy conveyed by the text, effectively making the commentators the target of the analogy and of the joke. Thus, it is the commentators who are parodied here by Bierce, and their revelations come off the worse for comparison with the originals of St. John.

This definition does show more signs of textual ambiguity than the previous two. The non-ironic interpretation is kept active until the last three words of the final sentence, when the reversal of source and target domain assignments is suddenly introduced. As with the *academe/academy* example (section 3.1), this use of irony could be used to argue that the definition displays an opposition of the expected/unexpected variety. Perhaps, though, this use of irony to exploit ambiguity in the assignment of source and target domains deserves separate recognition, at least where analogical humor is concerned.

Finally, it is worth noting that Bierce employs a comparison between two domains that are metonymically related. In the source domain, St. John is said to have concealed knowledge within his book. In the target domain, commentators end up doing nothing more than giving the appearance of uncovering this knowledge. The substitution of appearance for content occurs in such metonymic expressions as calling a marked police car a “black-and-white” or calling a chair with a rounded back a “roundabout.” Similarly, the source domain of the definition deals with the substantive knowledge that St. John concealed in the book of Revelation, whereas the target domain deals with the mere appearance of knowl-
<table>
<thead>
<tr>
<th>Court</th>
<th>Heaven</th>
</tr>
</thead>
<tbody>
<tr>
<td>jesters</td>
<td>*clergy</td>
</tr>
<tr>
<td>motley</td>
<td>canonicals</td>
</tr>
<tr>
<td>Court</td>
<td>Heaven</td>
</tr>
<tr>
<td>wear&lt;sub&gt;0&lt;/sub&gt;(jesters,motley)</td>
<td>wear&lt;sub&gt;1&lt;/sub&gt;(clergy,canonicals)</td>
</tr>
<tr>
<td>amuse(jesters,Court)</td>
<td>praise(clergy,Heaven)</td>
</tr>
<tr>
<td>laugh-at&lt;sub&gt;0&lt;/sub&gt;(Court,jesters)</td>
<td>laugh-at&lt;sub&gt;1&lt;/sub&gt;(Heaven,clergy)</td>
</tr>
<tr>
<td>signify&lt;sub&gt;0&lt;/sub&gt;(wear&lt;sub&gt;0&lt;/sub&gt;,amuse)</td>
<td>signify&lt;sub&gt;1&lt;/sub&gt;(wear&lt;sub&gt;1&lt;/sub&gt;,praise)</td>
</tr>
<tr>
<td>cause&lt;sub&gt;0&lt;/sub&gt;(amuse,laugh-at&lt;sub&gt;0&lt;/sub&gt;)</td>
<td>cause&lt;sub&gt;1&lt;/sub&gt;(praise,laugh-at&lt;sub&gt;1&lt;/sub&gt;)</td>
</tr>
</tbody>
</table>

Figure 5: Bierce's analogy between jesters and the clergy. Compare with his definition of *jester*.

3.4 Canonicals

Another definition with a religious theme is as follows:

**Canonicals, n.** The motley worn by jesters of the Court of Heaven.

This definition contrasts the worldly office of jesters with the supposedly spiritual office of the clergy. The analogy is shown in figure 6.

The comparison is highly condensed in the text; not only is the jibe conveyed by a single clause, but the actual target of the analogy is identified solely by the head of the definition, i.e., canonicals, and the final word Heaven. Yet, Bierce presents a straightforward comparison between the garb of clergymen and jesters. Of course, the purpose of the analogy is not completely served by simply acknowledging these attribute level mappings. Rather, from general knowledge, we know that the function of jesters is to amuse the court, and that they wear ridiculous costumes in order to signify this function. These facts are explicitly mentioned in Bierce's definition of *jester* elsewhere in the dictionary, so this knowledge is very likely what he had in mind for the current example. Similarly, the function of the clergy is to praise Heaven and God, and we know that they wear canonicals to signify this office. In suggesting that the clothing of clergymen and jesters are similar, Bierce is suggesting that their functions are similar as well. Because of this emphasis on the similarity of the source and target domains, it is difficult to say exactly how they are opposed. Possibly, the very comparison of ostensibly different domains implies that the opposition is of the expected/unexpected type.

Determining the type of opposition present is also made difficult by the fact that the text is not very ambiguous. There is a formal ambiguity in interpreting the final expression "Court of Heaven." Read as a possessive genitive, the expression simply refers to a court within, belonging to, the heavenly community. As such, the expression simply completes the analogy provided by the definition. But, read as an appositive genitive, the expression identifies Heaven as a court. The strong identification of the source and target domains provided by this reading better supports Bierce's insistence on their similarity, rather like the last sentence in the *belladonna* example (section 3.2). However, this ambiguity is not strongly activated by the text. But it is not clear from Raskin's theory how active, in the psychological sense, the required ambiguity must be.

Since both *Revelation* and canonical concern religious themes, it is not surprising that both definitions convey a similar relationship between their source and target domains. In the case of *Revelation*, the comparison between St. John and his commentators was metonymic, between substance and the mere appearance of substance. The definition of canonicals also focuses on the relation between substance and appearance. The motley of the jesters correctly signifies their function, whereas the canonicals of the clergy falsely signifies theirs. In the former case, this contrast is evidenced by the attribute mapping between knowledge and nothing (see figure 5). In the latter case, the evidence for a metonymic relation is given by the system mapping between cause<sub>0</sub> and cause<sub>1</sub>. Thus, although the definition of canonicals is similar to the definition of *Revelation*, it
<table>
<thead>
<tr>
<th>Slavery</th>
<th>Commerce</th>
</tr>
</thead>
<tbody>
<tr>
<td>chain-and-whip</td>
<td>debt</td>
</tr>
<tr>
<td>slave-driver</td>
<td>*creditor</td>
</tr>
<tr>
<td>*slave</td>
<td>*debtor</td>
</tr>
<tr>
<td>*labor</td>
<td>*money</td>
</tr>
<tr>
<td>wield(slave-driver,c-a-w)</td>
<td>assess(creditor,debt)</td>
</tr>
<tr>
<td>compel(c-a-w,slave)</td>
<td>oblige(debt,debtor)</td>
</tr>
<tr>
<td>perform(slave,labor)</td>
<td>earn(debtor,money)</td>
</tr>
<tr>
<td>benefit(labor,slave-driver)</td>
<td>enrich(money,debtor)</td>
</tr>
<tr>
<td>because0(wield,compel)</td>
<td>because1(assess,oblige)</td>
</tr>
<tr>
<td>cause0(compel,perform)</td>
<td>cause1(oblige,earn)</td>
</tr>
<tr>
<td>enable0(perform,benefit)</td>
<td>enable1(earn,enrich)</td>
</tr>
</tbody>
</table>

Figure 7: Bierce’s analogy between slavery and commerce. Compare with his definition of *creditor*.

3.5 Debt

In addition to academia and religion, Bierce liked to parody finance and commerce:

**Debt, n.** An ingenious substitute for the chain and whip of the slave-driver.

This definition elaborates on the same idea denoted by the term “debt-slavery.” The analogy conveyed by the definition is described in figure 7.

The definition makes use of a fair amount of implicit information concerning slavery and commerce. The target, debt, is identified strait away and a comparison of it with slavery is pursued. As far as a debtor is concerned, commerce consists of a debt (a way of objectifying a monetary obligation), the money to be repaid to defray the debt, and the creditor to whom the money is owed. A similar scheme can be realized for the slavery domain, with the very concrete chain and whip substituted for the abstract notion of a debt. The opposition expressed in the definition is straightforward. The emphasis here is clearly on the unsavory similarity between aspects of commerce and slavery. There is little that can be said concerning any opposition between these two domains, especially since they are conventionally related, e.g., in expressions like “debt-slavery.”

The issue of ambiguity also presents difficulties. As with the definition of *canonicals* (section 3.4), there is a formal ambiguity concerning what kind of substitute debt is for slavery. On the one hand, the definition could be understood as making a primarily historical claim about debt evolving from slavery as a form of compulsion. On the other hand, the definition could be understood as making a primarily factual claim about debt serving the same purpose as slavery. In the first case, debt and slavery are similar for historical reasons; whereas in the second case, debt and slavery are similar simply because they are analogous. Unfortunately, it is not clear that the humor turns crucially on proceeding from the first interpretation to the second, or that the ambiguity needs to be resolved at all.

A distinguishing feature of the definition of *debt* is that it provides a parody of commerce rather than religion or mind. Nevertheless, the relationship between source and target domains can be represented along similar lines. In this case, certain aspects of commerce are compared to the use of physical force. A common metaphor for describing commerce is in terms of fluid dynamics, as in the expressions “cash flow” or “economic pressure.” Thus it is quite appropriate to compare commerce with the physical coercion of slavery. This metaphorical relation between domains may provide a substitute for an opposition between domains as such.

3.6 Palm

For a final example, we analyze a more extensively articulated analogy:

Palm, n. A species of tree having several varieties, of which the familiar “itching palm” (*PalmæÂohaÂominis*) is most widely distributed and sedulously cultivated. This noble vegetable exudes a kind of invisible gum, which may be detected by applying

3 Interestingly, a poem Bierce attached to the definition of *debt* compares the debtor to a fish in an aquarium.
<table>
<thead>
<tr>
<th>Vegetable</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>palmₐ</td>
<td>palm₁</td>
</tr>
<tr>
<td>invisible(gum)</td>
<td>*hidden(greed)</td>
</tr>
<tr>
<td>bark</td>
<td>*skin</td>
</tr>
<tr>
<td>gold-or-silver</td>
<td>*money</td>
</tr>
<tr>
<td>*fruit</td>
<td>*profit</td>
</tr>
<tr>
<td>bitter(fruit)</td>
<td>*charity</td>
</tr>
<tr>
<td>*growers</td>
<td>*profiteers</td>
</tr>
<tr>
<td>cultivate₀(growers,palm₀)</td>
<td>grasp(palm₁,profit)</td>
</tr>
<tr>
<td>produce(palm₀,fruit)</td>
<td>collect(profiteers,profit)</td>
</tr>
<tr>
<td>harvest(growers,fruit)</td>
<td>part-of₁(charity,profit)</td>
</tr>
<tr>
<td>part-of₀(bitter,fruit)</td>
<td>manifest(palm₁,hidden,skin)</td>
</tr>
<tr>
<td>exude(palm₀,invisible,bark)</td>
<td>apply₁(money,skin)</td>
</tr>
<tr>
<td>apply₀(g-o-s,bark)</td>
<td>remain-on(money,skin)</td>
</tr>
<tr>
<td>adhere-to(g-o-s,bark)</td>
<td>dislike₁(profiteers,charity)</td>
</tr>
<tr>
<td>dislike₀(growers,bitter)</td>
<td>give-away₁(profiteers,charity)</td>
</tr>
<tr>
<td>give-away₀(growers,bitter)</td>
<td>because₀(cultivate₀,produce)</td>
</tr>
<tr>
<td>because₀(cultivate₀,produce)</td>
<td>enable₀(produce,harvest)</td>
</tr>
<tr>
<td>enable₀(produce,harvest)</td>
<td>cause₀(apply₀,adhere-to)</td>
</tr>
<tr>
<td>cause₀(apply₀,adhere-to)</td>
<td>enable₀(exude,adhere-to)</td>
</tr>
<tr>
<td>enable₀(exude,adhere-to)</td>
<td>because₀(give-away₀,dislike₀)</td>
</tr>
<tr>
<td>because₀(give-away₀,dislike₀)</td>
<td>because₁(cultivate₁,grasp)</td>
</tr>
<tr>
<td>because₁(cultivate₁,grasp)</td>
<td>enable₁(grasp,collect)</td>
</tr>
<tr>
<td>enable₁(grasp,collect)</td>
<td>cause₁(apply₁,remain-on)</td>
</tr>
<tr>
<td>cause₁(apply₁,remain-on)</td>
<td>enable₁(manifest,remain-on)</td>
</tr>
<tr>
<td>enable₁(manifest,remain-on)</td>
<td>because₁(give-away₁,dislike₁)</td>
</tr>
</tbody>
</table>

Figure 8: Bierce’s analogy between a palm tree and the palm of a hand.

to the bark a piece of gold or silver. The metal will adhere with remarkable tenacity. The fruit of the itching palm is so bitter and unsatisfying that a considerable percentage of it is sometimes given away in what are known as “benefactions.”

This definition constructs an elaborate analogy between fruit-trees and profiteers, and is shown in figure 8.

This analogy is far more complex than the previous examples because Bierce has more than one objective in mind. In fact, a proper understanding of his meaning requires some investigation outside the definition itself.

Up to the final sentence, the humor of the joke is fairly clear. Bierce plays on two puns in order to motivate a comparison between palm trees and human greed. The word “palm” itself is polysysemous and may refer either to a kind of tree, or to the inside surface of a hand. The word “cultivate” literally refers to the farming of plants, but can also be extended to refer to promotion of certain character traits, e.g., greed. Also, the term “itching palm” is a standard image for greed (see *Julius Caesar*, Act IV, Scene iii), which firmly establishes the respective domains of the analogy. The next two sentences exploit the analogy by comparing greed to the stickiness of tree sap.

The final sentence, however, changes the import of the definition. Instead of focusing on greed as a proclivity for making profit, the last sentence focuses on greed as the reluctance to give up profit already made. The occurrence of the word “benefactions” makes Bierce’s objective in this regard clear. The result of this change of emphasis is some textual incoherence. Based on the first part of the definition, we would expect the word “fruit” to refer to the entire output of cultivation. But then we would expect the fruit to be *sweet* to the greedy person, rather than bitter. The mention of food to designate an unwilling expenditure of profit can be seen in following additional definitions from *The Devil’s dictionary*:

**Air, n.** A nutritious substance supplied by a bountiful Providence for the fattening of the poor.

**Hovel, n.** The fruit of a flower called the Palace.

Both epigrams describe results of an extreme reluctance of the well-to-do to part with any portion of their wealth. In the second case, “fruit” is the term used to describe the outcome of this reluctance, as it is for *palm*. Since this fruit is
not willingly given up. Parting with it can be described as "bitter."

As with the definition of academe (section 3.1), the definition of palm uses a (bogus) etymological note to help establish an ironic reading of the joke. The term "palm" is ambiguous, but the first sense used is that of the species of tree. Immediately after this usage, however, the Latin expression palma hominis is given to designate this species. Of course, palma hominis is nothing of the kind, as it really means "human palm." Use of this device motivates an ironic interpretation of the definition, thus making the human palm the target of the humor. As with the definitions of academe and Revelation (section 3.3), this fact implies that the opposition between the domains may be best understood as simply one of irony.

Bierce does make a show of the putative ambiguity about which sense of palm he is using. In particular, the puns on the terms "palm" and "cultivate" are used to put on a tongue-in-cheek display of ambiguity. As with the previous examples studied in this section, this ambiguity is not relied upon for humorous effect in the way that Raskin specifies. In fact, some of the humorous quality of the definition seems to come from flaunting the literal ineffectiveness of the puns in disguising the preferred interpretation of the text.

The issue of metaphorical relationships between the source and target domains is fairly complex in the current example since several points of comparison are touched upon. Adherence is a common metaphor to describe commitment to an intention, as in the expressions "adhering to a course of action" or "sticking with a plan." The "remarkable tenacity" with which coins adhere to the gum of the palm is therefore well compared to the sedulously cultivated greed of profiteers. The final sentence of the definition introduces the metaphor of taste for feeling or affect, as in the expression "a bitter loss" or "the sweet taste of victory." The bitterness of the benefactions is caused by an unwelcome deviation from the acquisition of profit already alluded to. It is an unusual property of this analogy that its source and target domains are simultaneously compatible with two different metaphorical relationships, a property which makes the definition seem all the more ingenious, perhaps too much so.

3.7 THE ANALOGICAL PATTERN

As a logical mechanism, analogy clearly has some interesting properties when applied to create humor. The examples in this section demonstrate that analogy interacts in complex ways with the other parameters of humor, as laid out by the GTVH. Of particular interest is the effect that analogy may apparently have on the selection of target (TA) and script opposition (SO).

The use of analogy as a logical mechanism neatly raises the issue of the relation between the target (TA) of the joke and the target domain of the analogy. It can be considered an advantage of the analogical mechanism for joke construction that the target (TA) of the humor is assigned to the target domain of the analogy. This situation makes some intuitive sense, since both jokes and analogies may seek to inform us about their respective targets. It is unlikely that Bierce's humor would be as effective if the choice of source and target domains were reversed. Consider a reversal of the definition of canonicum (section 3.4):

Motley, n. The canonicums worn by Clergy of the king's Court.

If intended to mock the clergy, this joke is weak in the extreme. For the most part, the use of an analogical mechanism implies that the target of the joke occurs as the target domain of the analogy.

Bierce's use of analogy for the purpose of humor can be described as parody. In the sense discussed immediately above, this parody serves to bring some domain "down a notch" by an apt comparison with a less elevated domain. One domain is used, as it were, to drag down another by virtue of their analogical similarity. An alternative strategy is to show that one domain falls critically short of being aptly compared to a more elevated domain. In these cases, Bierce gives early indications of one assignment to the target domain, but then employs an ironic device to cancel those indications and adopt the opposite assignment. The reader is left with the spectacle of one domain measuring up badly to another, perhaps due to the semantic distance between philosophy and football, or knowledge and ignorance. In an ironic parody, semantic dissimilarity between domains can count in favor of the fitness of the analogy.

The use of analogy also has interesting implications for the script oppositions (SO) discussed by Raskin in his semantic theory of humor [12, p. 99]:

A text can be characterized as a single-joke-carrying text if both of the conditions [below] are satisfied:

- 182 -
1. The text is compatible, fully or in part, with two different scripts;
2. The two scripts with which the text is compatible are opposite.

Compliance with the first condition simply follows from the use of a fit analogy, with a source and target domain. Each source and target domain studied in section 3 represents information that is commonly known, or explicitly stated, about some subject. Examples include typical knowledge about schools, priests, and trees. This property of analogical domains is shared with Raskin’s [12, pp. 80–5] definition of a script. As a logical mechanism for humor, analogy is particularly well-suited to compliance with the first condition stated above.

The presence of two different domains in the definitions is also sometimes indicated by the use of puns, as in the belladonna and palm examples (sections 3.2 and 3.6). Puns typically preserve two readings of a text, even against people’s natural ability to disambiguate the sense of utterances [1, p. 183]. For analogies in which the source and target domains are semantically diverse, puns stand out in the multiconstraint framework by providing a mapping between low-level, identical predicates. The possibility for constructing a pun can therefore be considered as a positive constraint on the selection of source and target domains for an analogy.

The difficulty in applying Raskin’s second condition has already been noted throughout the discussion of Bierce’s examples. This difficulty follows from the fact that good analogies present two domains that are similar, rather than opposite in the obvious sense. As noted above, dissimilarity can play an important role in the cases where irony is used. As such, the parameter SO might be expanded to include an ironic assertion of similarity between two scripts. It is more difficult to see what can be done in non-ironic cases. Perhaps the notion of script opposition can be expanded to include the metaphorical and metonymic relationships identified in Bierce’s humorous analogies. Bierce shows a tendency to select domains that bear a metaphorical or metonymic relation to one another. Because his usual purpose is to parody humans, their ways and institutions, he usually selects a human target. But there are many metaphors available for understanding human activity in terms of physical mechanisms, chemical reactions, etc. This fact provides for an interesting way of constraining the selection of source domains. By appealing, albeit implicitly, to such metaphorical mappings, Bierce is free to compare life to domains which don’t, on the surface, bear much resemblance to it. This is not to say that Bierce always operates in this manner, or that he is compelled to do so, but it does provide him with the opportunity to make apt and clever comparisons.

Finally, it is worthwhile making a comment on the language (LA) of Bierce’s definitions. Even a casual glance through the figures of section 3 shows that many attributes of the target domain are left unexpressed in the text of the definitions. It is quite likely that the number of implicit attributes (and other mappings) in the target domain of a humorous analogy is never less than the number of implicit attributes in the source domain. Undoubtedly, this asymmetry is an effect of the need to mislead the hearer about the intended interpretation of the joke until the last possible moment. Bierce generally shows a great economy in the way he reveals the intended target of each definition. This suggests that one heuristic for generating good joke texts from prospective analogies is to express as little information about the target as is possible without damaging the comprehensibility of the text. Further consideration of language-generation (LA) issues is provided in section 4.

4 A COMPUTATIONAL MODEL

The analyzed examples from section 3 demonstrate the existence of a common pattern of construction. In order to better understand, and eventually simulate, Bierce’s brand of humor, this pattern is re-examined in this section from a computational point of view. Section 4.1 presents possible variations for the verbal realization of humorous analogies; section 4.2 describes a constraint-satisfaction approach to evaluating possible variations; section 4.3 touches on the issue of lexical choice in verbalizing humorous analogies.

4.1 REALIZING HUMOROUS ANALOGIES

The linguistic realization of humorous analogies (or any analogies, for that matter), presents an interesting challenge for natural language generation (NLG). For the most part, NLG is concerned with the problem of transforming a single source of knowledge \( K \) into a linguistic object \( L \) (e.g., a text, or an utterance in a conversation) [2]. By definition, analogies involve two pools of knowl-
edge: a source domain $S$, and a target domain $T$. The linguistic realization of an analogy can thus be seen as the problem of generating a single linguistic object $L$ from two sources of knowledge, $S$ and $T$. Along with the many other issues with which NLG systems must contend [8], an analogy generator must also decide what knowledge from $S$ and $T$ ought to appear in the final text. In this section, a simple constraint-satisfaction approach is outlined for choosing the relevant concepts from $S$ and $T$. We ignore for the moment the problem of constructing the analogies, instead starting from a complete analogy and focusing on the problem of which concepts should be explicitly mentioned in the linguistic realization of the analogy.

As Bierce’s work demonstrates, often only a few carefully selected words are required to fully realize a humorous analogy. Consider the belladonna analogy of section 3.2. Bierce’s realization explicitly mentions (in the first sentence) “poison” and “deadly”, and “lady” and “beautiful.” However, there are other ways of realizing this analogy, for example (items from the source and target domains are explicitly given beneath each text):

1. In English, that which causes the living to become dead. In Italian, a beautiful lady.
   $S_1 = \{live, dead, become, cause\}$
   $T_1 = \{beautiful, lady\}$

2. In English, a poison that affects living people. In Italian, a lady who affects men.
   $S_2 = \{poison, live, people\}$
   $T_2 = \{lady, men\}$

3. In English, a deadly poison. In Italian, that which causes happy men to become miserable.
   $S_3 = \{deadly, poison\}$
   $T_3 = \{happy, miserable, become, cause\}$

The problem for a humorous analogy generator is to choose the “funniest” realization of the analogy. For simplicity, attention is restricted to a linguistic form that closely follows the multi-constraint representation of analogies. Given a source $S$, a target $T$, a source name $N_S$ (e.g., English), and a target name $N_T$ (e.g., Italian), the problem is to generate a description of $S$ and a description of $T$, under the constraint of realizing the analogy in a humorous way. Note that the belladonna and academy/academe analogies both follow this pattern.

4.2 ANALOGY REALIZATION AS CONSTRAINT SATISFACTION

Realizing a humorous analogy in the best possible way requires careful selection of the information to be rendered explicitly and the information to be understood implicitly. The following six (ranked) heuristics, or preferences, characterize what should be true of the best verbal realization of domains $S$ and $T$ (see section 3.7):

1. Prefer overall shorter realizations to longer ones;\(^4\)
2. Say more about $S$, less about $T$;
3. Prefer attributes/objects to relational mappings, and relational mappings to systemic mappings;
4. For analogically paired concepts, prefer antonymic concepts;
5. Otherwise, prefer semantically distant (e.g., metaphorically related) concepts;
6. When possible, make a pun.

The best verbal realization of a humorous analogy is the one that most satisfies these constraints.

Assuming the number of items in the domains $S (|S|)$ and $T$ equals $n$, a brute-force algorithm would simply calculate an evaluative score for all pairs of $<S', T'>$, where $S' \subseteq S$, $T' \subseteq T$. Using the above heuristics, a score can be calculated for each pair, and the pair with the highest score is chosen. Practically, this would require calculating the score for $2^n$ pairs, so the space of possible candidates must be trimmed. To this end, the second and third heuristics can be formulated as hard constraints, so pairs $<S', T'>$ where $|S'| < |T'|$ or $S' \cup T'$ contains any systemic relations can be discounted.

As an example of how these heuristics can be applied, consider the three alternate realizations of belladona given in section 4.1, along with Bierce’s original realization: $S_B = \{deadly, poison\}$, $T_B = \{beautiful, lady\}$. If, as suggested above, the second and third heuristics are treated as hard constraints, then two of the four pairs can be discarded: $<S_1, T_1>$ and $<S_3, T_3>$ both contain explicit systemic relations (and $|S_1| < |T_1|$). For the other two pairs, the following facts are used to determine their final score:

\(^4\) "Brevity is the soul of wit." Hamlet, Act II, Scene ii.
\(<S'_2, T'_2>\)

- \(|S'_1| + |T'_2| = 5\)
- \(|S'_1| \geq |T'_2|\)
- No systemic or relational concepts appear in \(S'_2 \cup T'_2\)
- Lady and poison are semantically distant

\(<S_B, T_B>\)

- \(|S_B| + |T_B| = 4\)
- \(|S_B| \geq |T_B|\)
- No systemic or relational concepts appear in \(S_B \cup T_B\)
- Deadly and beautiful are semantically distant, and lady and poison are semantically distant

Clearly, Bierce's realization will score better than \(<S'_2, T'_2>\), and the best weightings for each heuristic should be determined empirically through experimentation. Depending upon how much computational effort one wishes to expend, scores may be defined ahead of time such that any pair that scores higher than this limit will be kept. Undoubtedly, it is also useful to allow constraints to be relaxed; it may be that no pair meets all the constraints, and so some can be removed to make the problem easier to satisfy. Such a procedure can be implemented using standard constraint satisfaction techniques, such as those provided by the SCREAMER constraint programming package [14].

The constraints given here are, in general, not constraining enough. To see this, suppose \(S\) and \(T\) both have \(n\) items, and half are systemic relations (judging from the examples in section 3, an overly generous assumption), so \(n\) could be replaced by \(n/2\), i.e., we need only consider \(2^n\) relations. Of these \(2^n\) remaining pairs, we can discard about half treating heuristic 2 as a hard constraint. The other heuristics are ordering constraints, and do not directly reduce the number of pairs that must be considered. Thus, we must still consider an exponentially large number of pairs. Either more constraints, or a more intelligent search method, are needed to make this particular approach computationally feasible.

### 4.3 Analogies and Lexical Choice

As mentioned above, economy of realization is an important feature of Bierce's definitions, especially the simpler ones. For this class of examples, a simulation of Bierce's mode of expression must deal primarily with the issue of lexical choice rather than rhetorical structure. As the heuristics listed in section 4.2 determine what information should be expressed, the next step is to choose exactly how to express it. Theories of lexical choice that apply to more general domains also apply here. The analogy-maker must adopt a theory of lexical choice, such as using pattern matching for choosing verbs [10], or using lexical preferences for choosing nouns [13]. Any theory of lexical choice requires that the lexicon be given the semantic information required to differentiate between similar words.

At least as important as choosing the right noun from a set of nouns, or the right verb from a set of verbs, is the task of determining the most salient concept. Whereas the heuristics of section 4.2 identify the relevant concepts for realization, the most salient concepts must also be determined in order to achieve the best linguistic realization. The salient concept is the one that will be most prominent when the analogy is realized. Issues relevant to determining salience include:

- **Heuristics**: e.g., actions are more salient than state changes, or state changes are more salient than constant states [11]. Since conceptual salience originates (in part) from the degree of unusualness or unexpectedness of a concept, using an "odd" salience in the analogy may contribute to its humorous quality. Thus a method is required for determining when a particular choice of salient concept is odd or unusual. Also, to maintain coherence through the process of realizing an analogy, analogically paired concepts should both be salient.

- **Descriptive Methods**: e.g., lexical cooccurrence relations, that govern which words should be realized together in a single text [15]. The use of lexical cooccurrence relations requires a lexicon with substantial information for each entry in the lexicon, along with definitions of permissible lexical relations.

At least two other issues must be considered when generating the final form of the analogy. Any constraint placed on the text generated for one domain should likely be placed on the text generated for the other domain as well, and the final sentences produced for each domain should be grammatically parallel.
For simple examples, a reasonable method of text generation is template filling [4], where templates give the text its structure, and the main problem is to fill in the templates with the most appropriate words. Such templates might have the following structure:

Domain 1, part of speech. description of domain 1.

Domain 2, part of speech. description of domain 2.

5 SUMMARY AND FUTURE WORK

By now it is clear that even apparently simple examples of analogical humor, such as the definitions analyzed above, conceal quite complex ideas and relationships. A broad range of approaches must be taken in order to understand, and eventually simulate, the kind of humor offered in plenty by figures like Bierce. Given the wide range of information necessary to get at the humorous qualities of definitions from The Devil's dictionary, it is not surprising that the computational method proposed here is based on various forms of constraint satisfaction. Such methods are well suited to handle highly disparate forms of information within a single framework and for a single purpose. As of yet, no implementation of the proposed model of humorous analogy generation has been completed. However, the techniques discussed here have been implemented, each on its own, by other researchers. Therefore, the outlook for success in constructing a computational model is promising.

The analysis of analogical humor given in this paper also raises a number of challenges which could not be fully addressed in the space available. We conclude this paper by summarizing the most immediate challenges below:

- This work focuses on small texts, a few sentences in length. Larger texts should also be considered, as well as non-linguistic analogies [6].

- Effects of irony on the understanding of analogy-bearing texts are noted. This interaction between irony and analogy merits further investigation in its own right.

- The script opposition (SO) required by the GTVH is not readily applied to the examples in section 3. Either the SO must be refined, or the interaction between the LM and the SO requires further scrutiny.

- Bierce's analogies show evidence of metaphorical relationships between domains. These relationships are not readily statable within the multiconstraint framework.

- We have not investigated methods for computationally creating humorous analogies—that is, for selecting the source and target domains and the relevant mappings. It would be interesting to determine in what way the creation of a humorous analogy differs from the creation of analogies for other purposes.

- Computational models of humorous analogies would benefit from being integrated appropriately with practical applications, such as user interfaces and machine translation systems.

The very fact that these challenges exist indicates that humorous analogies will continue to reward both theoretical and computational research.

REFERENCES


More on Humor Act: 
What Sort of a Speech Act is the Joke?

Michal Ephrat
Department of Hebrew Linguistics
University of Haifa, Israel
M.eprhat@UVM.HAIFA.A

ABSTRACT

Jokes are not modular syntactic or semantic structures, but manifestations of a global linguistic-pragmatic device. The paper sketches various aspects concerning jokes being a canonical means for performing a humor act. After describing the properties of the humor act as an illocutionary act, the two means for performing an illocutionary act, i.e., direct and indirect, are discussed regarding the humor act. We then look at the set of possible combinations of humor act, direct or indirect, with direct and indirect non-humor acts.

For computational humor, being a means for our better understanding of humor as a cognitive-human skill, to be fit for providing a true simulation and a conclusive reflection of the humor behaviour, it must treat jokes as a global construction, taking into account the various matters discussed above.

1 COMPUTATIONAL LINGUISTICS

Language is still seen as a unique human capability par excellence. Scholastic occupation with Computational linguistics (being a major part of artificial intelligence) could therefore be motivated by two different sources: (1) procedure oriented: use of our control over the computer’s hardware and software features as a means for simulating human-cognitive behavior to achieve a better understanding of the procedure, and (2) output oriented: the will to obtain results-products which are customarily produced by humans using their cognitive capabilities in a more efficient, rapid and accurate manner calls for the search for non-human substitutes, i.e., the computer.

2 THE THEORY OF SPEECH ACTS

Using words for saying vs. using words for doing
In the realm of semantics and the evolving science of pragmatics, scholars have observed the difference between utterances that serve to assert a state of affairs in the world and utterances that serve to create state of affairs in the world. Within the speech act theory, the latter have been named ‘illocutionary acts’. The theory of speech acts then explains (and predicts) the use of sentences for doing (rather than for saying) (Austin, 1962; Searle, 1969).

As cases of illocutionary acts one can mention, for example, sentences such as “I order you to P”, “I refuse P”, or “I promise you P”, “I apologize for P”, “I greet you for P”, etc.

The illocutionary act is listener centered. Roman Jakobson describes six autonomous components that participate in communication. He then correlates each of these components with a language function which is motivated by that particular component (Jakobson, 1960). In that model, the conative function is listener (addressee) centered (Jakobson, 1960:335). Regarding situational content, speech acts come under what Jakobson has termed the ‘conative function’.

Conditions for successful illocutionary acts
Each illocutionary act is defined and characterized both by its form and by its pragmatic content (situation). These include conditions of propositional content, preparatory conditions regarding speaker and listener and sincerity conditions (see Searle, 1969:54-71). Canonical utterances of (direct) illocutionary acts have the following full morpho-syntagmatic form: they contain a main clause and an embedded clause that indicate the propositional content. The subject
of the main clause canonically indicates the actor (formed using first person), and its deed (formed using a verb in first person present tense, see Searle, 1969:30, and the above examples). Conditions of the propositional content of P vary according to the specific act.

3 THE HUMOR ACT

Humor is customarily seen as another characteristic unique to humans (Raskin, 1985:14). Verbal jokes (opposed to 'practical jokes', see Grotjahn, 1987:40-43 - which are not telephone gags) are linguistic entities. A joke is a text. Raskin (1985:29) describes a joke as a "very short funny story". Being funny is not a result of a referential function performed using a joke. A joke is then not a text for asserting propositions but a text for creating laughter. It is an illocutionary act: a humor act (Raskin, 1985:1, passim; Bradshaw, 1976; Larudee, 1974).

It is interesting to note that Searle, counting on sincerity (Grice's boza-fide), thought jokes (as well as play-acting) to be a parasitic form of communication which falls outside the interest of speech act theory (see Bradshaw, 1976:62).

The humor act is listener centered

Clearly the act performed by a humor act is amusing to the listener (the audience). The humor act (as a special case of illocutionary speech act) is listener centered. Jokes are a matter of interaction. Says Martineau: "Humor is conceived generically to be any communicative instance which is perceived as humorous by any of the interacting parties. The humorous communicative instance becomes a vehicle or a social mechanism employed for interaction." (cited in Zajdman, 1992:357).

A joke situation necessitates a speaker and a listener (in a generic use: as writer-reader). Being listener centered, a joke teller needs a partner: the listener. A person can not tell a joke to himself. Says Fine (1983:159): "Just as one cannot tickle oneself, so, too, one can hardly tell oneself a joke or play a prank on oneself". When Raskin (1985:3) explains why the first factor of the humor act is that "there should be human participants in the act", says that: "It is the perceiver's presence, of course, which makes the humor act a humor act, simply because it is the perceiver who laughs".

Conditions for a successful humor act

The set of conditions for a humor act is by now a well established part of the current humor paradigm (see Raskin, 1985:55). Bradshaw (1977:63) lays out the following set of rules:

"Intention: A intends to achieve X, where X = amuse B, state p about C ... by means of a joke, J.

Occasion: A knows that B can hear him and is paying attention. A has reason to believe that B is not likely to be distracted and that B is not unwilling to be amused.

Audience: A believes that B can 'take a joke' (and that B can recognise a joke-commencement cue such as 'have you heard ...?') A believes that, on recognising the cue, B will remain silent and attentive until JU (joke utterance) is uttered, and that B will suspend judgment until JU is uttered.

Joke: A believes that B has not heard J before or recently, will not find J trivial, and will prefer his telling J to his not telling J. A believes that B has sufficient intelligence to guess M (meaning) from JU, has sufficient knowledge of C, and has sufficient knowledge of the idiom or figure of speech particular to J. A believes that B will not be merely embarrassed by J.

Skill: A knows that if the foregoing rules are observed and if his beliefs are justified he has the skill to utter J in such a manner as to stimulate B's interest about C, while preventing him from guessing M until JU is complete.

Laughter: A knows that he may laugh if B laughs, provided that M is implicit and not explicit.

Fulfillment: When all the above rules are observed, where relevant, the utterance of JU is the clue to M and hence is the token of A's intention to achieve X. The intention is fulfilled if B's laughter is triggered by JU and not adventitiously."

For a humor act to succeed, the listener must go through the following three sequential stages:

1. the listener must identify the utterance of the speaker as a humorous one; and then,
2. the listener must resolve the humor; and finally
3. the listener must show his enjoyment caused by the joke.

Computational Humor

When we examine which of the two listed motivations for computational linguistics (procedure oriented or output oriented) is applicable in the
case of computational humor, we find that only the former is applicable. We can use our control of the computer's components and procedures to learn about the strategies and capabilities we use for producing jokes and for understanding them: identifying an utterance as a joke, and parsing it as such (Bradshaw, 1977:62; Ephratt, 1990). Using different script type data as trials cases, we can create a simulation program for checking rules assumed for a humor act (e.g. Bradshaw's set, cited above). Yet, by definition of the joke situation as a case of a humor act: doing by saying, computers lacking the needed qualities of a perceiver in the humor act (the one who laughs) excludes the output oriented approach regarding a humor act. Moreover, the rationale for the choice of a computer as a tool in the output oriented approach is the need to produce results efficiently, accurately and quickly. The result of a successful humor act is invoking laughter in the listener. As explained, such results cannot be obtained by the computer, and even if they could, clearly they are not the sort of results for which we need efficiency, accuracy or speed. Regarding the humor act then the output oriented motivation is meaningless.

What we can expect from computational humor is that the computer would serve as a device for simulating and thus reflecting the behavior (procedures) that underlie the human humor act. This could be achieved if we assume the set of formal and situational conditions and the formal linguistic structures involved for a humor act to occur. Then we could formulate them as a computer algorithm, be it a deterministic algorithm or fuzzy decision device or a connectionist model enforcing or inhibiting humor readings, to check and clarify by simulation the assumed procedure.

4 DIRECT AND INDIRECT ILCUTIONARY ACTS

Canonical illocutionary acts are characterized by rules relating to their structure and their logical form. Being canonical (idiomatic), a listener who shares the linguistic competence marks this sentence as a case of an illocutionary act, parses it, and acts accordingly.

Linguists went on to show that among utterances that are used for performing an illocutionary act, one can identify two different means: direct (DIA), and indirect (IDIA) (see Searle, 1975; Davison, 1975). Indirect illocutionary speech acts lack the formal properties listed for their direct counterparts. Thus, they share their situational conditions.

Searle starts his paper that is devoted to 'indirect speech acts' in stating that:

"The simplest cases of meaning are those in which the speaker utters a sentence and means exactly and literally what he says. ... But notoriously, not all cases of meaning are this simple ... One important class of such cases is that in which the speaker utters a sentence, means what he says, but also means something more. ... In such cases a sentence that contains the illocutionary force indicators for one kind of illocutionary act can be uttered to perform, in addition, another type of illocutionary act" (Searle, 1975:59, and see Davison, 1975: especially 157).

5 DIRECT AND INDIRECT HUMOR ACTS (DHA, IDHA)

The direct humor act, as described above, is an explicit joke utterance. The speaker (joker) explicitly includes in the joke's setting an indication such as e.g., "I am now telling a joke", or "Here's a joke". We should comment here regarding Zajdman's (1991:26) statement that such an indication ('password' as she terms it) is external to the text of the joke, that it is indeed if it is an idiomatic way of introducing a joke, not at all external to the humor act. On the contrary, it is a necessary formal indicator for performing a direct humor act.

An indirect humor act is an implicit joke. Being indirect, it lacks some of the formal structural properties of the canonical direct joke yet keeps the logical propositional requirements (Davison, 1975:158). As in all case of an indirect illocutionary act, it demands of the listener to first identify the surface utterance as a means for an indirect act. It is claimed that in all cases of indirect humor, the listener does not identify the utterance as a joke right at its start. Even this initial stage is post factum. The successful fulfillment of the act is the is listener's laughter.

The humor act is a subset rather than a disjoint set of illocutionary acts. Non-humor acts are a heterogeneous group containing all illocutionary acts except for the humor act. The separation between humor and non-humor illocutionary acts is then not a categorical one but a matter of conve-
nience when concentrating here on the humor act. Regarding humor act and non-humor acts, two phases are possible: first, when the humor act is the indirect one: achieving the humor act by a direct non-humor act (DNHA), and second, when the non-humor act is the indirect one: achieving a non-humor act by a direct humor act.

As stated explicitly in the citation from Searle, an indirect illocutionary act is performed by means of a direct speech act (not necessarily illocutionary). It is always the case that the indirect act combines with the direct one. Its meaning is added to, rather than replaces, the meaning of the direct one (Searle, 1975; Davidson, 1975:143-149). It then follows that the meaning of any such utterance in which a humor act participates, directly or indirectly, is the sum of the direct and the indirect acts: a humor act and an additional act. If successful, it brings together laughter as well as another act.

The numerous social studies of humor have listed endless by-products as if perlocutionary acts that may result from the amusement and laughter invoked are a successful fulfillment of a humor act. Freud, many years prior to the evolution of speech act theory, surveyed the therapeutic motives and effects of making jokes and of hearing them. Alongside the primary enjoyment of jokes expressed by laughter (which constitute the success of the humor act), other specific effects such as refinement of the propositional content, relaxation, closer familiarity, and leaving a way out, follow.

These special soothing qualities of the humor act, and the formal possibility (explained in the speech act theory) of its combination with another illocutionary act, make jokes (direct and indirect) a useful means of communication. Cicero (1970:150-151) claimed that “it certainly becomes the orator to excite laughter; ... chiefly because it mitigates and relaxes gravity and severity, and often, by a joke or a laugh, breaks the force of offensive remarks, which cannot easily be overthrown by arguments.” Yet, he is well aware that the appropriateness and effectiveness of a joke in each situation is to be considered pendent on the degree and circumstances in which it is told. He then advises the orator that “A regard, therefore, to proper times, moderation and forbearance in jesting, and a limitation in the number of jokes, will distinguish the orator from the buffoon” (page 153, and see Fine (1983:173) regarding groups). It is usually the case that any successful humor act results in non-humor acts as the humor’s by-products. Having these characteristics is its important social power (see Raskin, 1985:140-144). Indirect manner compared to its direct counterpart, too, might carry an effect of playing down. This is why, when processing cases where both a humor act and (another) indirect illocutionary act occur, caution must be taken to successfully determine whether a particular effect is a result of the humor act or of the stylistic preference for using an indirect rather than a direct manner (see Davison, 1975). We also alert the reader, that by no means should an indirect humor act be confused with what Attardo (1988:359-362) terms ‘second degree’ jokes (see paragraph preceding figure 2).

For a joke to constitute a performance of an indirect non-humor illocutionary act, it is crucial that nothing be added to the joke. The text of the joke, (explicitly introduced or implicitly) combined with the relevant situational conditions, must be sufficient for the successful performance of the indirect non-humor illocutionary act. This is why, it seems to me that of the four means which Zajdman (1991:34) uses to describe how jokes can be inserted into a conversation, only her last type, the merge type (or appropriate idiomatic introductory indication, ‘password’ + merge) constitute an indirect non-humor illocutionary act. In her three former types, if an act is performed it is not done so by the joke but via other (direct or indirect) means.

6 Examples

We wish to show how a (surface) direct humor act can in fact be meant as a different (deep) non-humor indirect speech act (=IDNHA). How can one request, promise, refuse, greet or apologize, without performing their corresponding direct act, but by telling a joke? (Davison, 1975). The reverse type: performance of a direct non-humor act and meant as an indirect humor act is not dealt with here, since it is dealt with extensively elsewhere (see for example, Zajdman, 1992, detailed account of the role bona-fide and non-bona-fide plays in jokes, see also Raskin, 1985). We start off with an example of a visiting card which reads as follows:

Using any script model for a humor act would reveal two separate scripts: the upper two thirds of the card is in complete accord with visiting cards, stating addresses and phone numbers. Visiting cards are literary instances of a script of a professional act of introduction.
The lower third of the card does not fit into any of the expected slots of an introduction script. Motivated by the relevance and quantity maxims of Grice's cooperation convention (Grice, 1975, and see Raskin, 1985:103), one might assume here that the habit which currently spreads in e-mail, of adding a slogan at the end of the message alongside the signature and address, has also reached visiting cards. Having resolved and achieved a successful act of introduction, one could have stopped here ignoring the joke (see following 6 in figure 2).

Yet, going further with Grice's cooperation convention, the reader may raise the question 'why did the writer choose to add a slogan?' Careful reading of that addition would reveal script discrepancy, where Don's acting according to Joyce's advice would in fact result in destroying Don's initial intention (original script) of respecting John and Ann's right not to be disturbed during certain hours.

Should persons who read this visiting card identify and resolve these script discrepancies as shown, they would identify these lines as a joke and a successful humor act would be achieved (see above, citation of Bradshaw's rules). Owing to the fact that there is no formal explicit indication on the coming humor act, such as the title 'Joke:', this is a case of an indirect humor act (see following 2 in figure 2). Having now achieved a direct introduction act and an indirect humor act (see following 5 in figure 2), the reader can now stop here. Successfully amusing the interlocutor is a means for making him/her feel good, relaxed, and in the context of introduction (by a visiting card) is a means for increasing their desire to want to further interact with you.

Yet, taking Grice's maxim of relevance even further, the reader may now wonder, why did the writer choose this particular joke? One could have arrived at both the humor act and its by-product effects by many other direct or indirect jokes.

If the reader refers now (applying Grice's maxim of relevance) to the content of the joke, he identifies an indirect illocutionary act: a request not to phone between 2-4.

Such a request could have been achieved by either a direct request act (see following number 6 in figure 2): if the writer had chosen to occupy the lower third of the card by a clause such as: "Please avoid calling between 2 and 4", or an indirect act if she would have printed there: "We would much appreciate not having to answer the phone between 2 and 4". Yet, by choosing a second degree indirect speech act (see following 8 in figure 2), the writer provides the reader with a joke which functions as a polite subtle, not too aggressive request (but see Davison, 1975:149-156). Stating this request as an indirect joke leaves the listener an escape way out: if he/she does phone between the above mentioned hours, it does not necessarily indicate that they purposely refused to respect the request. Other possibilities may still hold: either the humor act was not successful, or the indirect request was not transferred (see Davison, 1975:153).

As explained, the humor act and the non-humor acts do not exclude each other. These two groups yield inclusive interactions between direct and indirect humor and non humor acts. Since the humor act is of one nature where as non-humor acts are heterogeneous, a combination of a direct humor act and an indirect non-humor act (or vice versa) are not excluded from appearing in a single utterance.
On the other hand, a case of a joke that is the subject matter of a meta-joke (see Attardo, 1988; Zajdman, 1992:363-366) always results in one and the same act: the humor act, and so does not constitute a separate category.

We then describe these possibilities as a matrix (Figure 2) consisting of twelve possibilities:

We shall now exemplify each of these groups:

Groups 1-2 consist of only humor acts:

1. The first group stands for direct explicit jokes. These are the canonical cases of a humor act. Such an example is, when in a gathering of economics students one announces: "Here's a joke: 'There are direct and indirect taxes' stated the lecturer. 'Could you provide an example of a direct tax?' he asked the students, 'yes' replied one of the students, 'logs' tax.' 'What do you mean?' asked the lecturer, 'The dog itself does not have to pay the tax.' ".

For such a sole direct joke to achieve its humor act the act's parsing but not identification conditions must all be fulfilled. It is claimed that since the crucial factor in the achievement of the humor act is the listener willingness to participate in the act and make it succeed, using the canonical direct way enables the speaker (be it any speaker or more so in the case of a professional comedian) to bring the listener into the right atmosphere (see Raskin, 1985:12). Notice also that the subject matter of the joke is not at issue here. In such a frame work, one can also tell a joke about speech acts, yet, it would still be considered a pure instance of a direct humor act. For example, when with friends one says "And the next joke goes like this: once in a restaurant, the waiter spilled soup over the client's trousers - and apologized. Replied the client: 'never mind colleague' asked the waiter: 'say, you too are a waiter?' replied the waiter: 'no, but neither are you!' ". The joke is about apologizing and on mocking, though the manner in which it is told here is not an apologizing neither a mocking act, but only a humor act.

The same goes for a joke about an indirect speech act. When in that same gathering one says, "And there's another joke: says one to the other: 'In order to enter the underground movement you have to kill a traitor and a dog' asks the other 'Why a dog?' replies the former 'You're accepted!' ". Identifying the acts hidden in the joke and parsing them as such requires one to invoke various indirect speech act rules some of which are not at all trivial. Yet, the only act performed here is a direct humor act. Similar strategies operate in the joke " 'Are you a psychiatrist?'; 'Why do you ask?'; 'You are a psychiatrist!' " (Raskin, 1985:254).

Notice that in this group only one speech act takes place: the direct humor act. For this to be true we must treat the effects of the humor act as its by-products, and not as separate autonomous acts obtained indirectly.

2. The second group differs from the previous one only in the formal characteristics: the jokes in this group are presented in an indirect rather than in a direct manner. The explicit statements such as "Here's another joke" or "And the next joke goes like this", are missing, compelling the listener to first identify the utterance as a joke.

3. The third group is beyond our scope of interest since no illocutionary act takes place (no humor act and no other non-humor acts).

Groups 4-6 contain cases of direct non-humor acts:

4. The fourth group brings together two simultaneous direct illocutionary acts, one of which is a humor act and the other is a non-humor act. The case described above of an explicit joke on a visiting card, which has no further indirect meaning, is an example of such a combination. Explicit cases of 'wh-jokes' also belong to the dual simultaneous acts: humor and question (or rhetoric question). For example, "Answer this joke: what did the horse-radish say to the fridge – Keep me cool, I'll stay hot".

5. Here, too, the direct illocutionary act is a non-humor one. The humor act follows from it indirectly. Except for the non-explicit joke counterparts of 4, we can include here two additional major groups:

a. Non-humor direct illocutionary acts in which the act itself (change of state of affairs in the world) results in a practical joke (on practical jokes, see Grotjahn, 1957:40-43). To take a simple example: an older brother says to his younger sister: "I ask you to bring me some bread" (a direct request). When she opens the cupboard to reach for the bread, a cat jumps out of the cupboard. If she takes this as funny this would constitute a practical joke resulting in a humor act.

b. A different group, is that of explicit advertisements (direct advise or persuasion act) that embody a non-explicit joke. Should the potential consumer identify the joke (being indirect) and laugh as a result, a humor act would have been achieved. The peanuts advertisement could
be such a case. In the ad. one sees an elderly man (could be the grandparent) who while holding a can of peanuts, says to a child (could be the grandchild): "I bet you can’t eat one". If the child keeps reaching for more and more peanuts, while laughing at the logical puzzle, then both acts appear to have been fulfilled.

6. This group of non-humor illocutionary acts is the canonical case dealt with by Austin (1962) and Searle (1969). They include direct directives, and commissives such as apology, greetings, declaring, etc. "I hereby pronounce you man and wife.", "I promise P.", etc.

Notice, however, that here only one direct non-humor act takes place. No indirect acts are in effect. There are acts or situations where this is the only possible way for a successful performance of the specific propositional content. Many of the declarative acts are such that an explicit direct performance is a necessary condition for their successful completion. No changes such as omission or addition are allowed. Any such change would result in an unsuccessful termination of the act. The priest would not say "May I hereby pronounce you man and wife."; or, for an act of condescension to be successful, an (indirect) sarcastic (not bona-fide) meaning of 'beloved' in a saying such as 'I am sorry to hear of the death of your beloved mother-in-law.', is ruled out. Where idiomatizability is part of the conditions for the specific illocutionary act, an indirect humor act is categorically ruled out. These may be the only cases where Bradshaw's claim (1977:63) that jokes are "pseudo speech acts, parasitic on the corresponding true acts" may be justified.

**Groups 7-9 deal with indirect non-humor speech acts**

7. In this group we present cases where a direct humor act is meant as an indirect non-humor act. Such is the case when in the course of a lecture on speech acts, the professor explains why a speech act is listener centered. He then says: "A joke goes like this: 'A Jew of German origin bought a train ticket and asked the conductor for a seat facing the direction in which the train was heading. When he entered the carriage, he noticed that the seat allocated to him was facing the wrong direction. When this passenger arrived at his destination, he went to the station-manager's office and complained. The manager accepted his complaint but asked curiously, 'why didn't you ask one of the passengers to switch seats?'; 'I couldn't' he replied, 'the carriage was empty, so there was no one to ask.' ". The direct act performed by the professor is a humor act. Yet, not being a comedian on stage, this was not his main aim. The primary indirect acts performed here are exemplifying, and prove that speech acts are listener centered, and that one has to request something of somebody.

Non-humor speech acts which are not idiomatic, (thus causing the exclusion of humor, see 6b. above) can be performed indirectly by a direct humor act. Jokes causing laughter refine other acts by playing down their emotional impact. According to Fine (1983:170), one of the five mechanisms through which humor enters is that "The humorous remark must be functional. It must support a goal toward which members are striving". This is why beyond the linguistic characteristics of the above mentioned cases, special pragmatic situations might favor or disfavor the insertion of humor. A subtle trade-off relation holds between the emotional or other degree of the non-humor act and between the choice of whether to use a joke to obtaining the non-humor act.

Where severe hierarchical distance holds and the speaker strongly disapproves his interlocutor's deed, it is quite unfit to use a joke for telling off. Yet, when relations are less formal, and dispute is more subtle, a joke can be a polite acceptable way
of telling off, promising, refusing or even threatening. It is different for a soldier to promise his officer to be on time for the next drill than for a busy husband to promise his wife that this time he would fulfill the promise he had repeatedly made to her. For example, after asking her husband over and over to pull out the tree in the garden path or alternately let her order a gardener to do the job, they sit one evening watching the moon from the terrace, and she asks him would he fulfill this request during the coming vacation. He says: “I just came across a joke. It goes like this: ‘Marcia, the lawyer’s wife was complaining: ‘Just look at our house, Bart! Its disgraceful, and our neighbors are talking about us. The furniture is old, the carpets are worn, the house needs painting. It’s a mess!, she whined. ‘We need a complete re-decoration’. ‘Honey, honey, don’t worry, just be patient’, Bart pleaded. ‘I just got a new divorce case today, and as soon as I break up their home, I can fix ours.”’ (Behrman, 1991:55).

8. The eighth group lacks any direct illocutionary speech act, yet it combines two indirect acts: a humor indirect act and a non-humor indirect one. It is in fact the indirect counterpart of groups 4 and 7 above.

The threat and its intensity are different if a person just pardoned for tax evasion asks the tax commissioner whether he should go straight now, or if he asks his companion the same question. The latter might respond: “It only depends whether you want to eat well during the day, or to sleep well during the night”.

The choice of expressing disapproval by a humor act, too, is a matter of intensity trade-off. For example, a waitress can say to a client who is going out of the restaurant without having left a tip: “Should you lose your wallet, remember that you did not take it out here”. Likewise, refusal, such as when the teacher turns to a pupil and asks him to awaken the pupil sitting next to her, instead of carrying out the request, she says: “you made him fall asleep, so you should awake him”.

We should comment here that this group is the only group presented here with two indirect speech acts. As already mentioned, humor acts are a subset rather than a disjoint set of illocutionary acts. The non-humor acts are a heterogeneous group containing all the illocutionary acts apart from humor. The question then arises whether an indirect humor act can combine with an indirect non-humor act, other indirect non-humor acts can combine with each other to yield a second degree indirect act? This seems to be a separate issue much beyond humor acts. Our very preliminary observation shows that whereas as a humor act can combine with different autonomous acts by way of mutual subject matter (but see reservations in 6b above), all cases that appeared as if they consist of two distinct non-humor illocutionary acts turned out to be cases that share essential conditions, and the second illocutionary act is in fact a perlocutionary act of the former. A frequent example is the combination of threat and mockery, or of advice and persuasion. In fact, these cases are of the same relationship described above for the soothing by-products effect of the humor act.

9. This group consists of the pure indirect non-humor illocutionary acts. Some such expressions behave as idioms for performing an illocutionary act. A frequently cited example is “I am freezing” as means for requesting someone to close the window, put on the heating, or fetch a sweater as the case may be. (Searle, 1975)

Groups 10-12 deal with the combination of direct and indirect non-humor speech acts in a single surface utterance

10. Three acts combine in this group: a direct humor act and a direct non-humor act that entail a third (different) indirect non-humor act. The visiting-card example with an explicit joke typed as a slogan that entails a request, is an example of such a combination.

11. Here too three acts are combined: the direct act here is a non-humor one. This in turn entails two separate indirect acts: an indirect humor act, and a non-humor indirect act (which must be different than the one performed directly).

To continue the previous example (see 9 above), John asks Daniel to close the window because it is cold outside. Instead of doing as requested he says: “And if I shall close the window, will it become warm (outside)?” The direct question provokes (indirectly) both a joke and a refusal (if he did not accompany his reply with the performance of the expected deed). Even a mild playdown threat can be achieved in this way. For example, the instructor in a small workshop, can no longer bear the on going talk of a group of participants, so he turns in their direction, and asks: “Could you tell me how many students would remain for the class break after I confine three of you?”.

12. An incidence of a direct non-humor act that stands for another non-humor act does not include humor. The classic example of this is “Can
you pass the salt?” uttered around the table and directed towards a person sitting close to the saltcellar (this too is an illocutionary idiom).

7 EPILOGUE

The fact that jokes are the canonical means for achieving a humor act is well established. Looking at jokes in general, and more specifically looking at them within the framework of computational humor, it appears that explicit jokes and more so implicit ones are not restricted to the humor act. Their social effects make them, in particular social settings, a most suitable means for obtaining indirect non-humor acts.

We set out here to outline the nature of the interaction between the jokes as means of the humor act and the non-humor acts. We have seen various possibilities of such interactions: jokes are indeed used for achieving the humor act. Some of these jokes make fun of speech acts: they involve an unexpected turn in a speech act script; others demand mastering speech act competence for the identification, or parsing of the joke. Yet, others add an additional layer to the humor act obtained by the joke: an indirect non-humor act.

If we see language as including more than a modular based mechanism of syntax, semantics and phonology, then a detailed account and prediction model of these possibilities (explicitly including enumeration of impossibilities) is essential in any attempt to understand language behavior.

Dealing with humor within the computational humor paradigm or otherwise, apart from identifying direct speech acts (humor or other), the program would now have to be able to identify and parse also indirect speech acts in which humor is involved. The computer’s algorithm would consist of the set of rules for differentiating and identifying each of the possibilities. It would then have to characterize the structures involved and relate each such property to the specific act that it performs.

Such understanding would shed light on the role of jokes within a micro- or macro society and among societies. It would make us better understand why an attempt to manually or automatically translate a joke is sometimes blocked not by semantic incompatibility (lack of identical lexical entries) but by indirect illocutionary act incompatibilities between the original and the target society, or social setting.

REFERENCES


Zajdman A., 1992, Did you mean to be funny? well, if you say so..., Humor, Vol. 5(4), pp. 357-368.
Parlevink
The Parlevink project is a language theory and technology project of the University of Twente. Starting-point is a (software) engineering approach to natural language and natural language processing systems. Special attention is paid to possible interactions with theoretical computer science.

Research Topics
In subprojects research is devoted to topics as syntax, semantics and pragmatics. Dialogue modeling is also part of the project, just as connectionist language learning and processing. Support research is done in the areas of formal languages and neural networks. Integration of the different topics takes place in the SCHISMA subproject: design and realization of a prototype natural language accessible theater information and booking system (joint with KPN Research). Other topics that play a role in this integration project are the embedding of such a system in a 'digital city' and World Wide Web, speech processing, natural language technology assessment and societal aspects. Another integration project that will start in 1996 is the European funded 21-project on multimedia information retrieval. Other funded projects that have started in 1996 are on neural networks and on societal effects of scientific research.

Ph.D. Research
In 1996 four to five Ph.D. students will perform their research in this project, partly in cooperation with other projects (robust language analysis, pragmatics, dialogue modeling and analysis, design and specification of NLP systems). In the spring of 1995 the Ph.D. thesis 'Little Linguistic Creatures' appeared. In 1997 a Ph.D. thesis on pragmatics in language technology is expected.

Students
It is expected that in the forthcoming years about thirty students will do their M.Sc. research in the Parlevink project. In addition there are many students who take a practical term outside the university (language and neuro engineering).

Workshops
Twice a year a Twente Workshop on Language Technology (TWLT) is organized. Starting as a local event these workshops have now become meetings of international specialists on topics from language engineering. In 1996 two workshops are organized. The first one, in June, will be on “Dialogue Management in Natural Language Systems” (TWLT11); the second one, in September will be on “Computational Humor: Automatic Interpretation and Generation of Verbal Humor” (TWLT12).
Twente Workshops on Language Technology

The TWLT workshops are organised by the PARLEVINK project of the University of Twente. The first workshop was held in Enschede, the Netherlands on March 22, 1991. The workshop was attended by about 40 participants. The contents of the proceedings are given below.

### Proceedings Twente Workshop on Language Technology 1 (TWLT 1)

**Tomita's Algorithm: Extensions and Applications**
Eds. R. Heemels, A. Nijholt & K. Sikkel, 103 pages.

**Preface and Contents**

A. Nijholt (University of Twente, Enschede). (Generalised) LR Parsing: From Knuth to Tomita.
G.J. van der Steen (Vleermuis Software Research, Utrecht). Unrestricted On-Line Parsing and Transduction with Graph Structured Stacks.
T. Vosse (NICI, Nijmegen). Detection and Correction of Morpho-Syntactic Errors in Shift-Reduce Parsing.
R. Heemels (OCÉ Nederland, Venlo). Tomita's Algorithm in Practical Applications.
M. Lankhorst (University of Twente, Enschede). An Empirical Comparison of Generalised LR Tables.
K. Sikkel (University of Twente, Enschede). Bottom-Up Parallelization of Tomita's Algorithm.

The second workshop in the series (TWLT 2) has been held on November 20, 1991. The workshop was attended by more than 70 researchers from industry and university. The contents of the proceedings are given below.

### Proceedings Twente Workshop on Language Technology 2 (TWLT 2)

**Linguistic Engineering: Tools and Products.**

**Preface and Contents**

A. Nijholt (University of Twente, Enschede). Linguistic Engineering: A Survey.
B. van Bakel (University of Nijmegen, Nijmegen). Semantic Analysis of Chemical Texts.
T. Vosse (ICLI, Nijmegen). Detecting and Correcting Morpho-syntactic Errors in Real Texts.
A. van Rijn (CIAD/Delft University of Technology, Delft). A Natural Language Interface for a Flexible Assembly Cell.
J. Honig (Delft University of Technology, Delft). Using Delta in Natural Language Front-ends.
D. van den Akker (IBM Research, Amsterdam). Language Technology at IBM Nederland.
The third workshop in the series (TWLT 3) was held on May 12 and 13, 1992. Contrary to the previous workshops it had an international character with eighty participants from the U.S.A., India, Great Britain, Ireland, Italy, Germany, France, Belgium and the Netherlands. The proceedings were available at the workshop. The contents of the proceedings are given below.

Proceedings Twente Workshop on Language Technology 3 (TWLT 3)

Connectionism and Natural Language Processing
Eds. M.F.J. Drossaers & A. Nijholt, 142 pages.

Preface and Contents
L.P.J. Veelenturf (University of Twente, Enschede). Representation of Spoken Words in a Self-Organising Neural Net.

P. Wittenburg & U. H. Frauenfelder (Max-Planck Institute, Nijmegen). Modelling the Human Mental Lexicon with Self-Organising Feature Maps.


W. Daelemans & A. van den Bosch (Tilburg University, Tilburg). Generalisation Performance of Back Propagation Learning on a Syllabification Task.

E.-J. van der Linden & W. Kraaij (Tilburg University, Tilburg). Representation of Idioms in Connectionist Models.

J.C. Scholtess (University of Amsterdam, Amsterdam). Neural Data Oriented Parsing.


M.F.J. Drossaers (University of Twente, Enschede). Hopfield Models as Neural-Network Acceptors.


R. Reilly (University College, Dublin). An Exploration of Clause Boundary Effects in SRN Representations.

S.M. Lucas (University of Essex, Colchester). Syntactic Neural Networks for Natural Language Processing.

R. Mikkulainen (University of Texas, Austin). DISCERN: A Distributed Neural Network Model of Script Processing and Memory.

The fourth workshop in the series has been held on September 23, 1992. The theme of this workshop was "Pragmatics in Language Technology". Its aim was to bring together the several approaches to this subject: philosophical, linguistic and logic. The workshop was visited by more than 50 researchers in these fields, together with several computer scientists. The contents of the proceedings are given below.

Proceedings Twente Workshop on Language Technology 4 (TWLT 4)

Pragmatics in Language Technology

Preface and Contents

D. Nauta, A. Nijholt & J. Schaake (University of Twente, Enschede). Pragmatics in Language technology: Introduction.

Part 1: Pragmatics and Semiotics

J. van der Lubbe & D. Nauta (Delft University of Technology & University of Twente, Enschede). Semiotics, Pragmatism, and Expert Systems.

F. Vandamme (Ghent). Semiotics, Epistemology, and Human Action.

H. de Jong & W. Werner (University of Twente, Enschede). Separation of Powers and Semiotic Processes.
Part 2: Functional Approach in Linguistics
C. de Groot (University of Amsterdam). Pragmatics in Functional Grammar.
E. Steiner (University of Saarland, Saarbrücken). Systemic Functional Grammar.
R. Bartsch (University of Amsterdam). Concept Formation on the Basis of Utterances in Situations.

Part 3: Logic of Belief, Utterance, and Intention
J. Schaalke (University of Twente, Enschede). The Logic of Peirce's Existential Graphs.
H. Bunt (Tilburg University). Belief Contexts in Human-Computer Dialogue.

The fifth workshop in the series took place on 3 and 4 June 1993. It was devoted to the topic "Natural Language Interfaces". The aim was to provide an international platform for commerce, technology and science to present the advances and current state of the art in this area of research.

Proceedings Twente Workshop on Language Technology 5 (TWLT 5)
Natural Language Interfaces
Eds. F.M.G. de Jong & A. Nijholt, 124 pages.

Preface and Contents
F.M.G. de Jong & A. Nijholt (University of Twente). Natural Language Interfaces: Introduction.
L. Boes (University of Nijmegen). Spoken Language Interfaces.
J. Nierent (University of Groningen). NL Interfaces and the Turing Test.
D. Speelman (University of Leuven). A Natural Language Interface that Uses Generalised Quantifiers.
W. Menzel (University of Hamburg). Title.
G. Neumann (University of Saarbrücken). Design Principles of the DISCO system.

The sixth workshop in the series took place on 16 and 17 December 1993. It was devoted to the topic "Natural Language Parsing". The aim was to provide an international platform for technology and science to present the advances and current state of the art in this area of research, in particular research that aims at analysing real-world text and real-world speech and keyboard input.

Proceedings Twente Workshop on Language Technology 6 (TWLT 6)
Natural Language Parsing: Methods and Formalisms
Eds. K. Sikkel & A. Nijholt, 190 pages.

Preface and Contents
A. Nijholt (University of Twente). Natural Language Parsing: An Introduction.
V. Manca (University of Pisa). Typology and Logical Structure of Natural Languages.
The seventh workshop in the series took place on 15 and 16 June 1994. It was devoted to the topic "Computer-Assisted Language Learning" (CALL). The aim was to present both the state of the art in CALL and the new perspectives in the research and development of software that is meant to be used in a language curriculum. By the mix of themes addressed in the papers and demonstrations, we hoped to bring about the exchange of ideas between people of various backgrounds.
C. Schwind (University of Marseille, France). Error analysis and explanation in knowledge based language tutoring.
J. Thompson (CTI, Hull, United Kingdom/EUROCALL). TELL into the mainstream curriculum.
M. Zoek (Limsi, Paris, France). Language in action, or learning a language by watching how it works.

The eighth workshop in the series took place on 1 and 2 December 1994. It was devoted to speech, the integration of speech and natural language processing, and the application of this integration in natural language interfaces. The program emphasized research of interest for the themes in the framework of the Dutch NWO programme on Speech and Natural Language that started in 1994.

Proceedings Twente Workshop on Language Technology 8 (TWLT 8)
Speech and Language Engineering
Eds. L. Boves, A. Nijholt, 176 pages.

Preface and Contents
Chr. Dugast (Philips, Aachen, Germany). The North American Business News Task: Speaker Independent, Unlimited Vocabulary Article Dictation
J.M. McQuean (Max Planck Institute, Nijmegen, The Netherlands). The Role of Prosody in Human Speech Recognition.
L. ten Bosch (IFO, Eindhoven, the Netherlands). The Potential Role of Prosody in Automatic Speech Recognition.
M.F.J. Drossaers & D. Dokter (University of Twente, Enschede, the Netherlands). Simple Speech Recognition with Little Linguistic Creatures.
H. Helbig & A. Mertens (FernUniversität Hagen, Germany). Word Agent Based Natural Language Processing.
Geunbae Lee et al. (Pohang University, Hyoja-Dong, Pohang, Korea). Phoneme-Level Speech and natural Language Integration for Agglutinative Languages.
G. Veldhuijzen van Zanten & R. op den Akker (University of Twente, Enschede, the Netherlands). More Efficient Head and Left Corner Parsing of Unification-based Formalisms.
G.F. van der Hoeven et al. (University of Twente, Enschede, the Netherlands). SCHISMA: A natural Language Accessible Theatre Information and Booking System.
G. van Noord (University of Groningen, the Netherlands). On the Intersection of Finite State Automata and Definite Clause Grammars.
R. Bod & R. Scha (University of Amsterdam, the Netherlands). Prediction and Disambiguation by Means of Data-Oriented Parsing.

The ninth workshop in the series took place on 9 June 1995. It was devoted to empirical methods in the analysis of dialogues, and the use of corpora of dialogues in building dialogue
systems. The aim was to discuss the methods of corpus analysis, as well as results of corpus analysis and the application of such results.

**Proceedings Twente Workshop on Language Technology 9 (TWLT 9)**
*Corpus-based Approaches to Dialogue Modelling*

Preface and Contents

**N. Dahlbäck** (*NLP Laboratory, Linköping, Sweden*). Kinds of agents and types of dialogues.


**J. Alexandersson & N. Reithinger** (*DFKI, Saarbrücken, Germany*). Designing the dialogue component in a speech translation system - a corpus-based approach.

**H. Aust & M. Oerder** (*Philips, Aachen, Germany*). Dialogue control in automatic inquiry systems.

**M. Rats** (*ITK, Tilburg, the Netherlands*). Referring to topics - a corpus-based study.

**H. Dybkjær, L. Dybkjær & N.O. Bernsen** (*Centre for Cognitive Science, Roskilde, Denmark*). Design, formalization and evaluation of spoken language dialogue.

**D.G. Novick & B. Hansen** (*Oregon Graduate Institute of Science and Technology, Portland, USA*). Mutuality strategies for reference in task-oriented dialogue.

**N. Fraser** (*Vocalis Ltd, Cambridge, UK*). Messy data, what can we learn from it?

**J.A. Andermarch** (University of Twente, Enschede, the Netherlands). Predicting and interpreting speech acts in a theatre information and booking system.

The tenth workshop in the series took place on 6-8 December 1995. This workshop was organized in the framework provided by the Algebraic Methodology and Software Technology movement (AMAST). It focussed on algebraic methods in formal languages, programming languages and natural languages. Its aim was to bring together those researchers on formal language theory, programming language theory and natural language description theory, that have a common interest in the use of algebraic methods to describe syntactic, semantic and pragmatic properties of language.

**Proceedings Twente Workshop on Language Technology 10 (TWLT 10)**
*Algebraic Methods in Language Processing*
Eds. A. Nijholt, G. Scollo and R. Steetskamp, 263 pages

Preface and Contents

**Teodor Rus** (*Iowa City, USA*). Algebraic Processing of Programming Languages.

**Eelco Visser** (*Amsterdam, NL*). Polymorphic Syntax Definition.


**Teodor Rus & James, S. Jones** (*Iowa City, USA*). Multi-layered Pipeline Parsing from Multi-axiom Grammars.

**Klaas Sikkul** (*Sankt Augustin, D*). Parsing Schemata and Correctness of Parsing Algorithms.

**François Barthélemy** (*Paris, F*). A Generic Tabular Scheme for Parsing.

**Frederic Tendeau** (*INRIA, F*). Parsing Algebraic Power Series Using Dymanic Programming.

**Michael Moortgat** (*Utrecht, NL*). Mutimodal Linguistic Inference.

**R.C. Berwick** (*MIT, USA*). Computational Minimalism: The Convergence of the Minimalistic Syntactic Program and Categorial Grammar.

**Annius V. Groenink** (*Amsterdam, NL*). A Simple Uniform Semantics for Concatenation-Based Grammar.

**Grzegorz Rozenberg** (*Leiden, NL*). Theory of Texts (abstract only).
Jan Rekers (Leiden, NL) & A Schürr (Aachen, D). A Graph Grammar Approach to Graphical Parsing.
Sándor Horvath (Debrecen, H). Strong Interchangeability and Nonlinearity of Primitive Words.
Theo M.V. Jansen (Amsterdam, NL). The Method of ROSETTA, Natural Language Translation Using Algebras.
Pál Dömsö (Kussuth University, H) & Jürgen Duske (University of Hanover, G). Subword Membership Problem for Linear Indexed Languages.
Vincenzo Manca (Pisa, I). A Logical Formalism for Intergrammatical Representation.

The eleventh workshop in the series took place on 19-21 June 1996. It focused on the task of dialogue management in natural-language processing systems. The aim was to discuss advances in dialogue management strategies and design methods. During the workshop, there was a separate session concerned with evaluation methods.

Proceedings Twente Workshop on Language Technology 11 (TWLT 11)
Dialogue Management in Natural Language Systems
Eds. S. LuperFoy, A. Nijholt and G. Veldhuijzen van Zanten, 228 pages

Preface and Contents
Pierre Nuges, Christophe Godéreux, Pierre-Olivier and Frédéric Revolta (GREYC, F). A Conversational Agent to Navigate in Virtual Worlds.
Anne Vilnat (LMSI-CNRS, F). Which Processes to Manage Human-Machine Dialogue?
Joris Hulstijn, René Steetskamp, Hugo ter Doest (University of Twente, NL), Stan van de Burgt (KPN Research, NL) and Anton Nijholt (University of Twente, NL). Topics in SCHISMA Dialogues.
Gavin Churcher, Clive Souter and Eric S. Atwell (Leeds University, UK). Dialogues in Air Traffic Control
Elisabeth Maier (DFKI, D). Context Construction as Subtask of Dialogue Processing -- the VERBMOBIL Case.
Wieland Eckert (University of Erlangen, D). Understanding of Spontaneous Utterances in Human-Machine-Diab.
TWLT 12 took place on 11-14 September 1996. It focussed on 'computational humor' and in particular on verbal humor. TWLT 12 consisted of a symposium (Marvin Minsky, Douglas Hofstadter, John Allen Paulos, Hugo Brandt Corstius, Oliviero Stock and Gerrit Krol as main speakers), an essay contest for computer science students, two panels, a seminar organized by Salvatore Attardo and Wladyslaw Chlopicki and a two-day workshop (Automatic Interpretation and Generation of Verbal Humor) with a mix of invited papers and papers obtained from a Call for Papers.

Proceedings Twente Workshop on Language Technology 12 (TWLT 12)

Automatic Interpretation and Generation of Verbal Humor

Eds. J. Hulstijn and A. Nijholt, 208 pages

Preface and Contents

Oliviero Stock 'Password Swordfish': Verbal Humor in the Interface.


Akira Ito and Osamu Takizawa Why do People use Irony? - The Pragmatics of Irony Usage.

Akira Utsumi Implicit Display Theory of Verbal Irony: Towards a Computational Model of Irony.

Osamu Takizawa, Masuzo Yanagida, Akira Ito and Hitoshi Isahara On Computational Processing of Rhetorical Expressions - Puns, Ironies and Tautologies.

Carmen Curcă Relevance Theory and Humorous Interpretations.

Ephraim Nissan From ALIBI to COLOMBUS: The Long March to Self-aware Computational Models of Humor.

Salvatore Attardo Humor Theory beyond Jokes: The Treatment of Humorous Texts at Large.

Bruce Katz A Neural Invariant of Humour.

E. Judith Weiner Why is a Riddle not Like a Metaphor?

Tone Veale No Laughing Matter: The Cognitive Structure of Humour, Metaphor and Creativity.

Tony Veale and Mark Keane Bad Vibes: Catastrophes of Goal Activation in the Appreciation of Disparagement Humour and General Poor Taste.

Kim Binnend and Graeme Ritchie Speculations on Story Fun.

Dan Loehr An Integration of a Pun Generator with a Natural Language Robot.

Cameron Shelle, Toby Donaldson and Kim Parsons Humorous Analogy: Modeling 'The Devils Dictionary'.

Michal Ephrat More on Humor Act: What Sort of Speech Act is the Joke?

The proceedings of the workshops can be ordered from Vakgroep SETI, Department of Computer Science, University of Twente, P.O. Box 217, NL-7500 AE Enschede, The Netherlands. E-mail orders are possible: twlt_secr@cs.utwente.nl. Each of the proceedings costs Dfl. 50.