

TWLT20
The April Fools' Day Workshop
on Computation Humour

PROCEEDINGS OF THE TWENTIETH
TWENTE WORKSHOP ON LANGUAGE TECHNOLOGY

AN INITIATIVE OF HAHACRONYM
European Project IST-2000-30039

APRIL 15-16, 2002
TRENTO, ITALY

Oliviero Stock, Carlo Strapparava,
and Anton Nijholt (eds.)

CIP GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Stock O., Strapparava C., Nijholt A.,

The April Fools' Day Workshop on Computational Humour
Proceedings Twente Workshop on Language Technology 20
An initiative of HAHAcronym. European Project IST-2000-30039
O. Stock, C. Strapparava, A. Nijholt (eds.)
Italy, Trento, ITC-irst

ISSN 0929-0672

trefwoorden: computational humor, humor theory, computational linguistics, emotion modelling,
irony modelling, pragmatics, embodied agents

© Copyright of the papers remains with the authors

Book orders:
Ms. C. Bijron
University of Twente
Dept. of Computer Science
P.O. Box 217
NL 7500 AE Enschede
tel: +31 53 4893680 – fax: +31 53 4893503
Email: bijron@cs.utwente.nl

Druk- en bindwerk: Reprografie U.T. Service Centrum, Enschede

Preface

These are the proceedings of the international *April Fools' Day Workshop on Computational Humour* that, despite its name, has been a serious and successful event. It was held in Trento on April 15-16, 2002¹. The workshop follows the successful International Workshop on Computational Humour held in Enschede in 1996 and is an initiative of HAHAcronym² (IST-2000-30039), the first European project about computational humour.

The aims of the workshop were bringing together research results in the area of computational humour, with an emphasis on the computational interpretation and generation of verbal humour. Scientifically, the field is a young area of investigation that poses some very challenging questions: what 'intelligence' is required for the skillful use of humour? How can modelling humour contribute to modelling intelligence? What role can humour play in making systems look more intelligent and enjoyable?

A number of research results on humour can suggest a computational treatment of the topic.

One assumption that underlies the research on computational humour is that in future human-machine interaction, humans will demand a naturalness and effectiveness that requires also the incorporation of models of possibly all human cognitive capabilities, including the handling of humour. We believe there are many practical settings where humour will add value. Among them there are: business world applications (such as advertisement, e-commerce, etc. . .); general computer-mediated communication and human-computer interaction; increase in the friendliness of natural language interfaces, edutainment and autonomous agents systems.

The program of the workshop has included a selection of six accepted papers, proposed by researchers in Europe, North America, Japan; six invited talks by distinguished scientists in the field; an intriguing after dinner talk by Aaron Ben-Ze'ev, Rector of the University of Haifa; a panel concerned with the relation between modelling of emotions and modelling of humour; demonstrations of running systems.

The Program Committee of the workshop was formed by:

- Salvatore Attardo (Youngstown University, USA)
- Kim Binsted (I-Chara, Japan)
- Douglas Hofstadter (Indiana University, USA)
- Anton Nijholt (University of Twente, The Netherlands)
- Andrew Ortony (Northwestern University, USA)
- Victor Raskin (Purdue University, USA)
- Willibald Ruch (Queens University, Belfast, UK)
- Oliviero Stock (ITC-irst, Italy) (chair)
- Carlo Strapparava (ITC-irst, Italy)

Our thanks go to all the participants, the people that submitted papers, and the prestigious invited speakers, that have enthusiastically accepted the invitation - and even sent in some written contribution in time - and foremost to the members of the Program Committee.

At the end, humour is something we need for our survival. For surviving with computers they will have to demonstrate some humour capability themselves. This workshop has been an initial contribution in that direction.

Oliviero Stock, Carlo Strapparava and Anton Nijholt

Trento, April 2002

¹April 1st was not possible, as that day in 2002 is both Easter Monday and a day of Passover

²<http://haha.itc.it>

Previous TWLT workshops

Previous TWLT workshops were

- TWLT1, *Tomita's Algorithm: Extensions and Applications*. 22 March, 1991.
- TWLT2, *Linguistic Engineering: Tools and Products*. 20 November, 1991.
- TWLT3, *Connectionism and Natural Language Processing*. 12 and 13 May 1992.
- 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.
- TWLT11, *Dialogue Management in Natural Language Systems*. 19-21 June, 1996.
- TWLT12, *Automatic Interpretation and Generation of Verbal Humor*. 11-14 September 1996.
- TWLT13, *Formal Semantics and Pragmatics of Dialogue, Twendial'98*. 13-15 May 1998.
- TWLT14, *Language Technology in Multimedia Information Retrieval*. 7-8 December 1998.
- TWLT15, *Interactions in Virtual Environments*. 19-21 May 1999.
- TWLT16, *Algebraic Methods in Language Processing (AMiLP2000)*. 20-22 May 2000.
- TWLT17, *Learning to Behave: Interacting Agents (CEvoLE1)*. 18-20 Oct 2000.
- TWLT18, *Learning to Behave: Internalising Knowledge (CEvoLE2)*. 22-24 Nov 2000.
- TWLT19, *Information Extraction in Molecular Biology*. 11-14 Nov 2001.

For the contents of the previous proceedings, please consult the last pages of this volume.



ITC
irst

CENTRO PER LA RICERCA
SCIENTIFICA E TECNOLOGICA

Contents

Invited Speakers

<i>Formalizing Humor Theory</i>	1
Salvatore Attardo (Youngstown University, USA)	
<i>Optimal Innovation and Pleasure</i>	11
Rachel Giora (Tel Aviv University, Israel)	
<i>Frame Blends, Analogies, and Humor</i>	29
Douglas Hofstadter (Indiana University, USA)	
<i>Computational Humor and Ontological Semantics</i>	31
Victor Raskin (Purdue University, USA)	
<i>The Structure of Forced Reinterpretation Jokes</i>	47
Graeme Ritchie (Edinburgh, Scotland)	
<i>Computers with a Personality? Lessons to be Learned from Studies of the Psychology of Humor</i>	57
Willibald Ruch (Queens University, Belfast, UK)	

Invited After Dinner Speaker

<i>Humour in Online Romantic Relationship</i>	71
Aaron Ben-Ze'ev (University of Haifa, Israel)	

Regular Talks

<i>Generation of Idiom-based Witticisms to Aid Second Language Learning</i>	77
Justin McKay (Binsted-McKay, Tokio, Japan)	
<i>Real and Fictional Ridicule. A Cognitive Approach for Models of Humour</i>	89
Francesco Mele (Istituto di Cibernetica, CNR, Napoli, Italy)	
<i>Embodied Agents: A New Impetus to Humor Research</i>	101
Anton Nijholt (University of Twente, The Netherlands)	
<i>A Formalism for Misapprehended Identities: Taking a Leaf Out of Pirandello</i>	113
Ephraim Nissan (University of Greenwich, UK)	
<i>Humorous Agent for Humorous Acronyms: The HAHAAcronym Project</i>	125
Oliviero Stock, Carlo Strapparava (ITC-irst, Trento, Italy)	
<i>The Verbal/Non-verbal Distinction: Still and O-pun Question</i>	137
Chris Vernour, Michael Levinson, Greg Lessard (Queen's University, Canada)	

Formalizing Humor Theory

Salvatore Attardo
Youngstown State University
Youngstown, OH 44555, USA
attardo@cc.ysu.edu

Abstract

This paper assumes familiarity with the Semantic Script Theory of Humor (SSTH; Raskin 1985) and the General Theory of Verbal Humor (GTVH; Attardo and Raskin 1991). All unexplained acronyms are defined therein. It also assumes a general knowledge of the research on humor in linguistics, as reviewed in Attardo (1994).

1 WHY HAVE A FORMAL THEORY?

In Raskin and Attardo (1994), we introduced the argument that humor theory would benefit from being confronted with computational treatments because providing a formal, algorithmic treatment of the subject forced clarity, well-definiteness, and in general all implicit parts of an argument to the surface, so to speak. More in general, computability approximates quite well formality, in the meta-epistemological sense intended. In what follows I assume without further argument this point.

It should be noted that I am not claiming the equivalence of computability and formality, for the obvious reason that some formal systems are non-computable, for example because they involve recourse to infinity.

2 HOW FORMAL IS HUMOR THEORY?

The research on the linguistics of humor has come under fire for not being sufficiently formal(ized) to be effectly evaluable. While in some cases this charge is correct, in others it is equivalent to the proverbial throwing away the baby with the bath water. A survey of the research on humor in linguistics will be found in Attardo (1994).

In this paper I will argue that significant parts of the linguistics of humor are sufficiently formal for any empirical verification, while stopping short of axiomatization (mostly on historical grounds). Specifically, I will look at two issues: the crucial definition of “script” in Raskin (1985) and the attempt at formal description of a class of logical mechanisms (LMs) in Attardo *et al.* (forth.). I will also look briefly at the first implementation of Jape (Binsted and Ritchie 1997), in light of the previous discussion. A longer paper, including discussion of other implementations of Jape and of other computational humor systems, is in preparation.

3 SCRIPTS

Cognitively a script is an organized complex of information “about” something (typically, a lexical item, but not exclusively, since there are obviously non-lexicalized concepts). There is a significant literature on scripts, frames, and other such semantic “objects” reviewed in Raskin (1985) and Attardo (1994, 2001), among others. We will not refer to it, except to note that a script, in the sense used in this paper, is roughly similar to a frame.

A script is defined in mathematical terms as a domain of an unlabeled directed graph (Raskin 1985b: 308). Technically, the semantic network of script-based semantics is a multigraph (i.e., several edges may link two vertices). No theoretical problem comes from this simplification. To

be fair, very little actual work in script-based semantics has been done in the formalism of graph theory, although ontological semantics (Nirenburg and Raskin *forth.*) is expressed in a formalism defined by a Bachus-Naur form. Most day-to-day work in script-based semantics is done using a simplified, easy to process, slot-and-filler formalism, reminiscent of lexical-functional grammar (e.g., Raskin 1985a), or in LISP pseudocode (Raskin *et al.* 1994a/b). On more work in graph-theoretic terms, see of course section (4), below.

In other words, if every simple element of meaning (I am evasive on the exact nature of the simplex semantic “atoms” because much inter-theoretical strife has gone into identifying them; for lack of a better term “concepts,” “lexemes,” or “semantic features” may be used) is represented as a vertex in a graph, and a semantic relationship between any two simplex elements is represented as the two elements being adjacent in the graph, then a script is the resulting set of vertices and hedges connecting the adjacent vertices. Hedges can be labeled so as to emphasize certain semantic relationships, typically used in script/frame theory, such as ISA, PARTONYMY, etc.

Under this definition, the obvious issue becomes the delimitation of each individual script: in other words, since every concept is somehow related to every other (indirectly, to be sure!) since it is intelligible, theoretically a script for DOG would include information on MANDOLIN. Raskin (1985: 81) notes that a script is a “domain” within the general graph. Not much more definition is necessary. One can advocate, as I have done (Attardo 1994), a spreading activation model which will decrease the activation of nodes of the graph as one moves progressively away from the central node (the lexical “handle” of the script, for those which have one) and it appears the psychologically at least, spreading activation is a passable model of actual human processing. However, the solution to the apparent problem is much simpler: in reality, the “boundaries” of the “individual” script are uninteresting because we “reify” scripts for pedagogical purposes. In actual understanding/processing we operate with dynamic, constantly changing, often “ad hoc” concepts (consider what the script for “occasional Monday night football pizza” must look like). Formally, graph theory takes care of things very elegantly. Stipulation that only directly linked nodes may be included in the script for a given “handle” will immediately determine the size/shape of the domain. Larger domains may be carved out by stipulating n-degrees of linking: if V is the set of vertices, $V : \{a, b, c\}$ and E the set of edges, $e = ab \in E$ (1-degree of linking), if $bc \in E$, but ac is not, then a-c is a 2 degree link, etc.

There are many issues in script-theory that we cannot address, due to lack of space, such as the different kinds of scripts, the different ways that scripts can be activated during processing, and many others. Raskin (1985) and Attardo (1994, 2001) provide discussion of several of these issues. However, we will briefly address the issue of saliency in scripts.

3.1 SALIENCY ISSUES IN SCRIPTS

In Attardo *et al.* (*forth.*), we argue in some detail *against* the idea that scripts are sets (although one may choose to represent them as such for illustrative reasons) for the simple but important reason that scripts are not random collections of elements (as sets are). Scripts, quite to the contrary, are organized. A common way to conceptualize one aspect of the organization of scripts is a slot-filler pairs, with a (small) number of “service” or “metalinguistic” slots (such as “agent,” “time,” “location,” etc.) which provide the conceptual organization of the script. For example, a verb-like concept will have an “agent” slot, a “time” slot, and possibly a “patient” or “recipient” slot, etc.

The organization of scripts does not end here: further ordering is introduced by the presence of events and subevents, presupposed events, goals, etc. Consider the script for ALIMONY which presupposes the following, obviously organized, events: marriage, divorce, court order, payment from the agent to the former spouse. In other words, (some, mostly verbal) scripts “tell a story” (Raskin *et al.* 1994a/b) which is organized along a timeline.

Even further organization is provided, on a different plane, by saliency effects within the script. Raskin (1985: 82) notes that some parts of the script domain are “emphasized” to indicate their centrality. In Attardo *et al.* (*forth.*), we use the notion of “foregrounding” which we represent as a proper subset of the script, or more effectively as links to a special “foreground” node.

Foregrounding or saliency is particularly important in humor theory, because I have argued that it may be the key to better understanding the complex issues surrounding the notion of “oppositeness” in the SSTH. I will return briefly to this in section (5), below. On saliency and humor, see Giora (1991).

3.2 DISCOVERY PROCEDURES FOR SCRIPTS

In contrast to most “formal” theoretical linguistic work, script-based semantics has actual implementable discovery procedures for its constructs, and does not have to rely on unconstrained and notoriously unreliable “native speaker intuition.” Script-based semantics considers a script as an hypothesis on the semantic content of a lexical item (ignoring non-lexical scripts, i.e., scripts that do not correspond to any lexical item in a given language, for simplicity in this context). In a very Popperian fashion, the current state of the script is considered merely as not (yet) proven false by evidence, which will come from further uses of the lexeme in a given text. When the speaker (or the NLP system) encounters a sentence containing the targeted lexeme its script is activated and reassessed on the basis of any new information provided by the new context. If the script is found to be defective (i.e., missing some information, or containing information that is faulty) it is updated as needed. Alternatively, the speaker (or the NLP system) may reject the new information (i.e., decide that the information embodied in the script is more reliable than the new, contrasting information).

From a different, but complementary perspective, the following is Raskin’s original formulation of this principle:

The inclusion of an element of semantic information in a script is justified if a sentence can be found, such that it contradicts this element and is semantically deviant for this reason alone (1985b: 311)

Thus far we have shown that script-theory is falsifiable, with respect to the contents of the scripts. How formal is script-theory, then? Formalization is of course an independent metric from falsifiability (it is trivial to construct formal but unfalsifiable theories). Certainly script-theory, with very few exceptions, one of which discussed immediately below, is not presented in any axiomatic form. The reasons for not doing so, strike me as follows: 1) there exists a risk of premature formalization, the history of syntactic theory in the past 30 years is a looming caveat to the effect that elaborating a sophisticated formalism is no guarantee that one’s theory will succeed even in describing an artificially delimited subdomain of the field, let alone provide a complete description of its object. 2) in a sense, the axiomatic method (provide axioms and rules for the generation of theorems) is not very appropriate to the discovery procedures of the lexicon. The lexicon is more like the set of axioms that are fed to the generative process. 3) in another sense, the lexicon is a very large set of items, whereas typically axioms are kept to a small number and rules generate large amounts of theorems.

Of course, a theory can be formal, while not being formalized. What matters is of course that one’s theory be formalizable in principle. Under this light, the SSTH is definitely a formal theory of semantics. Some of its aspects may still need work, to be sure, but which theory doesn’t? Having established that formal(izable) nature of the SSTH, let’s turn to its descendent, the GTVH.

4 THE GTVH IN GRAPH THEORY

The GTVH assumes and incorporates the SSTH, while significantly expanding it. Some aspects of the GTVH may be completely intractable formally, at least at present (consider that the GTVH includes narrative aspects, situational aspects, and aggression-targeting aspects, for example). However, the two higher-level knowledge resources, script opposition and logical mechanism, may be described more formally. Attardo *et al.* (forth.) endeavors to do so. What follows is a (much) shortened version of the argument presented therein.

As the reader of Raskin (1985) will recall, the SSTH claims that a joke presents two features:

- a full or partial overlap between two scripts, and

- that the two scripts are (locally) antonymous (“opposed”)

We now proceed to formulate these concepts in the theory of graphs.

4.0.1 SCRIPT OVERLAP

Two scripts (graphs) S_1 and S_2 are said to be overlapping iff they share at least an edge (e) and two vertices adjacent to e . Or, in other words, if there exists a subgraph G , such that $G \subset S_1 \& S_2$ or to put it differently $G = S_1 \cap S_2$.

We now need to introduce a specification, to disallow trivial overlaps. Let us consider the infamous doctor’s wife joke (1):

- (1) “Is the doctor at home?” the patient asked in his bronchial whisper. “No,” the doctor’s young and pretty wife whispered in reply. “Come right in.” (Raskin 1985: 100)

Since both the patient visiting the doctor and the lover visiting his wife are humans and they both are agents of the actions they respectively perform, according to the definition above, the two scripts would overlap. In fact, any scripts about any human doing something would trivially overlap with them. Even worse, scripts come with “service” slots and subslots, which for example, record coreferences between slots, time-stamp actions, etc. All scripts share them and therefore all scripts would trivially overlap.

We know, however, that scripts come with foregrounded parts. So all we need to do is restrict the definition above so that the subgraph $G = S_1 \cap S_2$ include a path which includes at least one “focus” vertex (we can think of this as a label or as an actual vertex which links to the focused vertices). Figure (1) shows a fragment of the graph, with the overlapping parts, some of which are in focus and some of which are opposed (semantically).

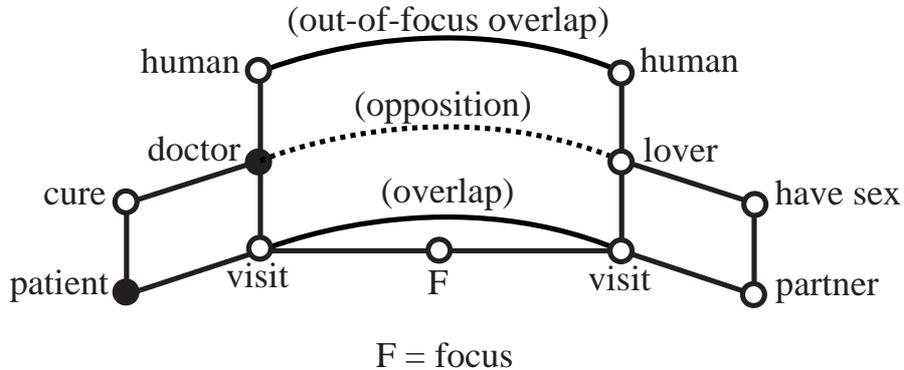


Figure 1: A fragment of the graph for joke (1). Only the vertices (scripts) are labeled.

An issue that should be kept in mind is that overlap can be considered as a mapping of a vertex onto itself (i.e., a loop). We can avoid using loops, by representing twice the vertex, or by representing the vertex in the intersection part of a Venn diagram (in other words, the vertex occurs in both graphs (sets)).

4.1 LM

We can now move on to the LM of a joke. We can describe it as an isomorphic mapping between two graphs $G_1 \rightarrow G_2$. Significantly, the isomorphic mapping may *not* be present in the graphs, but be nonetheless “implied” by it or “forced” upon it. Consider the following example, which is the title of Devlin (1993): *The Joy of Sets*. Let us represent as a simplified graph Devlin’s title as a chain (Bollobás 1998: 80) or path (Harary 1969: 13), in which each of the words of the title is a vertex (Figure 2) and in parallel we represent the title of Comfort (1972) *The Joy of Sex*.

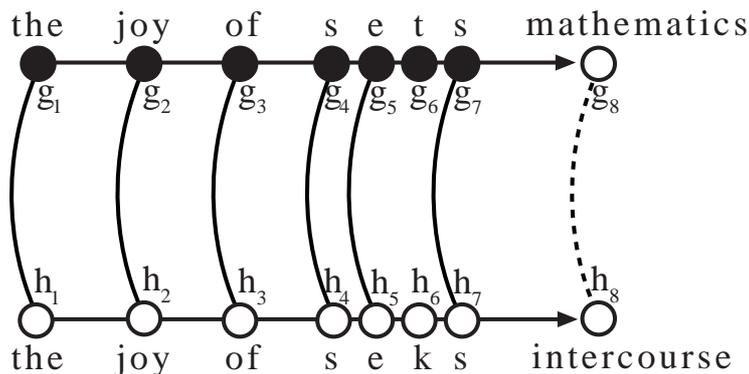


Figure 2: Chains for *The Joy of Sets/Sex*

Assume that the words of the titles (the analysis could be done at the morphemic or phonemic level, of course) are themselves subgraphs ($g_1, g_2, g_3, \dots, g_n$) of a graph G (*The Joy of Sex*) and H (*The Joy of Sets*). It is clear that the parallelism between the two paths imposes a mapping $G \rightarrow H$ or $H \rightarrow G$.¹ However, the mapping stops short at the vertices $g_6 \not\rightarrow h_6$.² We claim that it is precisely the spurious extension of the mapping to $g_8 \rightarrow h_8$ that constitutes the (local logic, playful) resolution of the incongruity of a parallelism between set theory and sexual intercourse. The reader may be puzzled by our choice to illustrate the notion of mapping LM by a spurious example, but he/she need only recall that LMs are not necessarily *real* resolutions of the incongruity, but may be merely playful, pretend resolutions. The next section presents examples of non-entirely spurious mappings.

4.1.1 FURTHER EXAMPLES OF LMS REPRESENTED IN GRAPHS

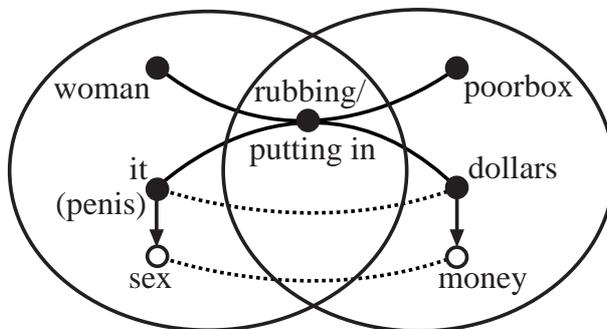


Figure 3: Graph for joke (2): false analogy LM.

False analogy Consider joke (2)

- (2) A married man goes to confessional and tells the priest, “I had an affair with a woman - almost.” The priest says, “What do you mean, ‘almost?’” The man says, “Well, we got undressed and rubbed together, but then I stopped.” The priest replies, “Rubbing together is the same as putting it in. You’re not to go near that woman again. Now, say five Hail

¹In fact, the two graphs overlap, see above.

²Actually, it continues up until the third phoneme (g/h_5), and is reestablished in the fourth (g/h_7), but we choose to ignore this last fact. We simplified the phonetic representation, for ease of drawing. The dotted lines represent the inferred links to the scripts MATHEMATICS and INTERCOURSE, as well as the parallelism-imposed mapping SETS \rightarrow SEX.

Mary’s and put \$ 50 in the poor box.” The man leaves confessional, goes over and says his prayers, then walks over to the poor box. He pauses for a moment and then starts to leave. The priest, who was watching him, quickly runs over to him and says, “I saw that. You didn’t put any money in the poor box!” The man replied, “Well, Father, I rubbed up against it and you said it was the same as putting it in!” (randomjoke)

which can be represented (in part) in figure (3). In (3) we see two scripts INTERCOURSE and PENANCE (in this case, instantiated as the offering of money to the poor). The scripts are not fully represented (since that would be unwieldy), but rather they are alluded to³ by the two circles (the left circle stands for the INTERCOURSE script, while the right one stands for PENANCE. The intersection/overlap of the two scripts are the nodes RUBBING and PUTTING IN which (the joke tells us) are locally⁴ equivalent. We then get a mapping of WOMAN onto POORBOX *via* the isomorphism between POORBOX, DOLLARS, and RUBBING / PUTTING IN (not represented in the picture) and WOMAN, IT = PENIS and RUBBING / PUTTING IN (also, not represented in the picture). The picture does, however, represent the mapping WOMAN → POORBOX, and IT = PENIS → DOLLARS. Note that RUBBING / PUTTING IN is mapped onto itself. Picture (3) also represents the opposition between IT = PENIS and DOLLARS and (by extension, indicated by the arrows) between SEX and MONEY. Thus (3) represents the two scripts overlapping, in opposition, and their logical mechanism.

Parallelism From the preceding discussion, it follows that the parallelism LM is the prototypical case of mapping LMs. Indeed this is the case. Consider joke (3).

(3) A wife is like an umbrella. Sooner or later one takes a cab. (Freud 1905: 93)

We can represent the parallelism as a directed graph and if we introduce the convention that empty circles are scripts not occurring in the text (PRIVATE and PUBLIC, which are arrived at inferentially), the result is figure (4), in which we see the two parallelizing predications of the text and the resulting “triangulation” which allows to deduce the double parallelism of the joke: A wife is to a prostitute, what an umbrella is to a cab, i.e., the private form of a commodity. Therefore, it follows that a wife is like an umbrella (and a prostitute is like a cab).

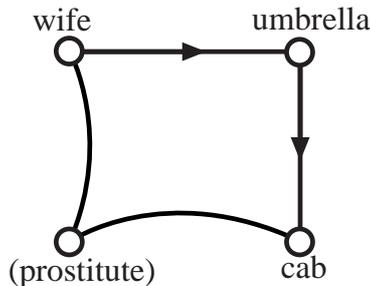


Figure 4: Graph for joke (3): parallelism LM.

Space prevents further exemplification and discussion, but I think that it should be clear that (some) LMs are amenable to formalization. Given their abstract nature and their complexity, this bodes well for the ultimate goal of formalizing the GTVH entirely.

5 JAPE

We now turn extremely briefly to a concrete example of application of humor theory of computational matters. Jape is a punning riddle generator (Binsted and Ritchie 1997). It generates riddles of the kind:

³The reader will recall that scripts can be represented as sets for illustrative purposes.

⁴Cf. Raskin’s “local” antonymy (Raskin 1985: 108) and Attardo (1997) for discussion.

(4) What do you get when you cross breakfast food with a murderer? A cereal killer.

The components that make up Jape are:

1. The lexicon: “store(s) general (non-humor oriented) syntactic and semantic information” (Binsted and Ritchie 1997: 44) it consists of 59 words and 21 NPs (Binsted and Ritchie 1997: 58).
2. Homophone base: list of homonyms and homographs.
3. six schematas (more semantic) “The fact that there is anything amusing or bizarre about the question-answer pair is caused by the choice of lexemes related in particular ways, and this is the responsibility of the schema” (Binsted and Ritchie 1997: 40)
4. fourteen templates (more syntactic) “There is no essentially humorous knowledge encoded in the set of templates.” (Ibid.)
5. “a post-production checker, which applies some simple heuristics to sift out *some of the more obvious non-jokes*” (my emphasis, SA) (Binsted and Ritchie 1997: 43).

Binsted and Ritchie (1997: 68-69) are fairly negative (but not completely so) about the relevance of humor research for their computational modeling of humor. Mostly they object of the lack of formalization of the resources postulated by humor theory and to their unconstrained “AI-complete” nature (which makes actual implementation impossible or impractical). Both of these claims have some elements of truth. However, the arguments presented in this paper should have allayed at least some of these objections.

Let us turn to the relevance of Jape toward humor theory. Binsted and Ritchie (1997: 70) note that their “model does not emphasize the same features of jokes structure” as the GTVH. Specifically, they claim SO “is not even represented in [their] model” (Ibid.). They conclude: “Either script opposition is not actually very important in question-answer punning riddles, or one script opposition is implicit and fixed for all such jokes” (Ibid.).

So, which is it? A few lines before the discussion of SOs and Jape quoted above, Binsted and Ritchie state the in their examples “there is no strong evidence for the two scripts to be opposed—they need only be different” (Binsted and Ritchie: 1997: 70). This is a telling statement. In Attardo (1997) and Attardo *et al.* (forth.), I have argued that the complex issue of defining local antonymy (Raskin’s 1985 definition of “oppositeness”) should be handled along the lines of a “semantic axis.” This is not the place to go into the details of this issue (see Attardo *et al.* forth.), but suffice it to say that there are foregrounded and backgrounded aspects of the script/frame.

Now, according to Binsted and Ritchie (1997: 61), Jape’s output suffered significantly from failing to account for this fact (the volunteers who produced the lexical entries being untrained in lexical semantics, presumably). In their example, the lexical entry for BOAR includes as a link the ACT-VERB SNIFF which while true (apparently) is not foregrounded (this resulted in a humorous text that was judged poorly formed: “what do you use to sniff a drilling tool: a wild bore” [bore/boar]). Incidentally, Binsted and Ritchie (1997: 61) reveal even less knowledge about *Sus Scrofa* than the present author has, when they claim that boars do not “typically” sniff. I am pretty sure that the typical boar sniffs, having a snout, more or less as much as it breathes, or eats. The problem of course is not typicality, but foregroundedness: sniffing is a possible physical activity of most animals with breathing apparatus and is not especially foregrounded, the way length of the neck is foregrounded (at default) for giraffes. Presumably boars have other hog-like foregrounded aspects of their being or behavior.

To be fair, this is a very complex area, as foregrounding is not a static affair: work on saliency (e.g., Giora 1991) has shown that contextual pressure may alter (even radically) the foregrounded parts of a script/frame. I am afraid we are still far away from a complete understanding of these issues. However, it is already clear that, whatever was wrong in the failed joke quoted in Binsted and Ritchie (1997: 61) had to do with a faulty SO: by establishing an opposition between non-foregrounded parts of the scripts, the SO fails to function (possibly, because “sniffing” simply fails to activate the script for “boar” at the right time).

Furthermore, provided that some sort of “semantic axis” requirement is respected (i.e., that the opposition is between foregrounded or salient portions of the scripts/frames, at the time of the opposition) I tend to think that indeed mere difference is all that is required semantically for incongruity (i.e., being the complementary set, cf. Attardo 1997). So, summing up, Jape is in substantive agreement with the SO definition and in fact it seems likely that implementing a more specific SO mechanism would have prevented some of Jape’s misfires.

Let us conclude with an observation that seems to have escaped Binsted and Ritchie: namely that all question-answer punning riddles share a LM, namely the Cratylistic association between signifier and signified that is present in all homophonic or paronymic puns (see Attardo 1994: ch. 4 for details). Thus we see that Jape, far from being at odds with the GTVH, is in fact largely congruent with it.

References

- Attardo, Salvatore. (1994). *Linguistic Theories of Humor*. Berlin: Mouton De Gruyter.
- . (1997). “The semantic foundations of cognitive theories of humor.” *Humor: International Journal of Humor Research*. 10-4. 395-420.
- . (2001). *Humorous Texts*. Berlin: Mouton De Gruyter.
- Attardo, Salvatore and Victor Raskin. (1991). “Script theory revis(it)ed: joke similarity and joke representation model.” *Humor: International Journal of Humor Research*. 4-3/4. 293-347.
- Attardo, Salvatore, Christian F. Hempelmann and Sara Di Maio. (Forth.) Script oppositions and logical mechanisms: modeling incongruities and their resolutions. *HUMOR: International Journal of Humor Research*.
- Binsted, Kim and Graeme Ritchie. (1997). Computational rules for generating punning riddles. *HUMOR: International Journal of Humor Research*. 10:1. 25-76.
- Bollobás, Béla. (1998). *Modern Graph Theory*. New York, NY/Berlin: Springer.
- Comfort, Alex (ed). (1972). *The Joy of Sex; a Cordon Bleu guide to lovemaking*. New York, Crown.
- Devlin, Keith J. (1993). *The Joy of Sets : Fundamentals of contemporary set theory*. New York : Springer. Rev. ed. of: *Fundamentals of contemporary set theory*. 1979. New York: Springer.
- Diestel, Reinhard. (1997). *Graph Theory*. New York, NY: Springer.
- Giora, R. (1991) On the cognitive aspects of the joke. *Journal of Pragmatics*, 16:5. 465-485.
- Harary, Frank. (1969). *Graph Theory*. Reading, MA: Addison Wesley.
- Raskin, Victor. (1985a). *Semantic Mechanisms of Humor*. Dordrecht: D. Reidel.
- . (1985b). Script-based semantics: a brief outline. *Quaderni di Semantica*. 6:2. 306-313.
- Raskin, Victor and Salvatore Attardo (1994). Non-literality and non-bona-fide in language. An approach to formal and computational treatments of humor. *Pragmatics and Cognition*. 2:1. 31-69.
- Raskin, Victor, Salvatore Attardo and Donalee H. Attardo. (1994a). Augmenting Formal Semantic Representation for NLP: The story of SMEARR” *Machine Translation*. 9, pp. 81-98. Reprinted in Bert Peeters (ed.) *The lexicon-encyclopedia interface*. Amsterdam/New York: Elsevier, 2000. As Augmenting linguistic semantics descriptions for NLP: Lexical knowledge, encyclopedic knowledge, event structure. 463-481.

Raskin, Victor, Salvatore Attardo and Donalee H. Attardo. (1994b). "The SMEARR Semantic Database: An Intelligent and Versatile Resource for the Humanities" in D. Ross and Brink (eds.), *Research in Humanities Computing*. 3, Oxford: Clarendon Press, pp. 109-124.

Optimal Innovation and Pleasure

Rachel Giora*
Linguistics
Tel Aviv University
Tel Aviv 69978
Israel
giorar@post.tau.ac.il

1 ON THE ROLE OF SALIENCE AND NOVELTY IN AESTHETICS: THE OPTIMAL INNOVATION HYPOTHESIS

When asked which of the stimuli in (context-less) quartets like (1-4) they liked best, subjects selected the third option as most pleasing. Next came the first, followed by the second. Least likable was the fourth (Kronrod, 2001; Kronrod et al. 2000):

- (1) Body and soul.
- (2) Bodies and souls
- (3) Body and sole
- (4) Bobby and Saul

When asked which of the pairs (in 5) they liked best, subjects rated (5a/c) as more pleasing than (5a/b) (Elnatan, 2000, 2002; Giora, Kronrod, Elnatan and Fein, 2001):

- (5a) You don't know your right from left?
- (5b) *The Comprehensive Lexicon* will teach you whatever you don't know
- (5c) Buy *The Comprehensive Guide for the Political Factions in Israel*.

How do (1-4) and (5a/b-a/c) differ? What could account for the gradability in their likability and pleurability? As shown by Kronrod (2001), Kronrod et al. (2000), and Giora et al. (2001), and see also Giora (in press, Chapter 7), the factor that accounts for their pleurability is optimal innovation: novelty that allows for the recoverability of the familiar. Pleasure, it seems, hinges on recognizing the familiar in the novel (see also Freud, 1905; Mukarofsky, 1932/64, 1978; Shklovsky, 1917/1965).

Indeed, the various stimuli vary on a familiarity scale: (1) was rated as most familiar: it is a common, fixed expression. (2) was rated as second most familiar: it is a variant version of (1), but such that preserves its meaning. Third came (3), which will be termed 'optimally innovative': like (2), it feeds on the familiar (1), but it also conveys an extra, unfamiliar sense. The item in (4), however, was rated least familiar: it is, in fact, entirely new. Though akin to (1) in rhythm, shape, and sound, meaning-wise it is a drastic diversion. Unlike (2) and (3), it does not echo the familiar stimulus¹. Similarly, (5a/b) was rated as most familiar: it instantiates the salient, idiomatic interpretation; (5a/c), however, was rated as less familiar: it is optimally innovative (see 7): it invites a less salient, literal interpretation, without dispensing with the

*I am grateful to Mira Ariel, Salvatore Attardo, Albert Katz, Orna Peleg, and Noa Shuval for their very helpful comments.

¹Indeed, the quartets were constructed in such a way so as to contain a familiar source (1) and variations (2-4) diverging from the source in varying degrees. As confirmed by readers' ratings, (2) is the variation most similar to (1), involving just a minimal change. Mid-position is occupied by (3). Least similar to (1) is (4), involving two changes (see Kronrod, 2001; Giora, Kronrod et al., 2001)

salient interpretation. Contra the received view (Brinker, 1988), then, it is not pure novelty (4) that accounts for pleasure, but optimal innovation–innovation that resides in familiarity (3, 5a/c):

- (6) The optimal innovation hypothesis

If a stimulus is optimally innovative it would be rated as more pleasurable than either a familiar stimulus or a purely innovative stimulus.

- (7) Optimal innovation

What stimulus would count as optimally innovative?

A stimulus would be optimally innovative if it involves

- (a) a novel response to a familiar stimulus,

but

- (b) such that would also allow for the automatic recoverability of a salient response related to that stimulus so that the similarity and difference between the novel and the salient would be assessable (see also Giora, in press, Chapter 7).

For a response (e.g., a meaning) to be salient, it should be foremost on our mind due to factors such as experiential familiarity, frequency, conventionality, or prototypicality (see Giora, 1997, in press, Chapter 2). Salient responses get activated automatically upon encounter of a familiar stimulus, regardless of context fit. To the extent that a linguistic innovation (*body and sole*) allows the comprehender an insight into some salient meanings (*body and soul*) while promoting new ones, it is optimally innovative.

By definition, then, the notion of optimal innovation excludes familiar stimuli (1, 5a/b): familiar stimuli do not meet any of the requirements in (7). However, it also excludes innovations such as variants of such stimuli (2) as well as pure innovations (4). Variations and variants (*A single piece of paper*, which stems from *A piece of paper*) do not meet the first requirement (7a above). Though they involve a slight change, this modification does not result in a novel response (as opposed to e.g., *A peace of paper*). Neither are pure innovations optimal, as they do not meet the second requirement (7b above): no familiar response is recoverable so as to become instrumental in constructing the novel response (*Y2K, fax, grok; spandy-wear*).

As illustrations of optimal innovations, consider the following examples. *Sofa so good*—the name of a London shop—constitutes a literal pun. It introduces a novel meaning that recommends its goods (furniture). Yet it does not dispense with the salient meaning of the string (So far so good). Similarly, the stimulus *Her wedding ring is a "sorry we're closed" sign* conveys a novel, metaphoric meaning (non-receptiveness to courtship). This, however, involves the salient, literal meaning that gets activated automatically (Pexman, Ferretti & Katz, 2000). A novel irony such as *read my lipstick* is optimally innovative too. While allowing for a new meaning to emerge, ridiculing the speaker on account of her femininity, it still invokes the salient (suspicion/lie) meaning of the familiar irony *Read my lips*. Jokes such as *How do you get holy water? Boil the hell out of it* are also optimally innovative. They are instances of novelty that resides in salience. They harp on old, salient strings (curse), involving a slight twist (see Attardo, 2000) that results in a novel interpretation of the salient (nonliteral) meaning. In the same way, the interplay between the salient (*body and soul*) and the innovative (*body and sole*) accounts for the optimal innovation. Or consider a visual example: Ariel Sharon's portrait ((8) by Piven (1999), carving the current Israeli prime minister (salient response) out of bloody minced meat (novel response), thus alluding to his murderous and voluptuous nature.

Note that according to the optimal innovation hypothesis, it is neither literality nor figurativeness that accounts for the pleasure induced, but some salience imbalance: the surprising discovery of the novel in the salient or the salient in the novel. It is not a sheer surprise, then, that is pleasing, but a somewhat novel response that could evoke a salient response or could be assigned to a salient response though it did not come to mind immediately (see also Giora, 1991).



(8) "Arik Sharon": An optimally innovative visual stimulus

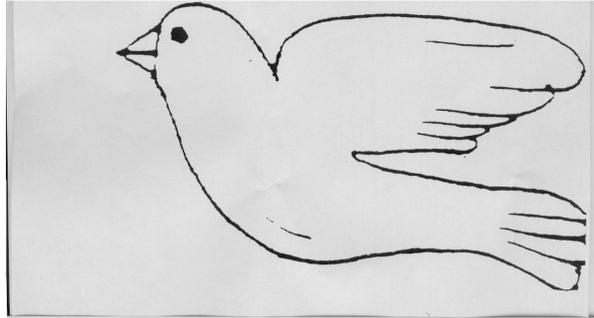
2 TESTING THE OPTIMAL INNOVATION HYPOTHESIS

In our lab, we further examined aspects of the optimal innovation hypothesis (Elnatan, 2002; Giora et al., 2001; Kronrod et al., 2000). We demonstrated that stimuli rated most pleasurable (3) were indeed optimally innovative: they involved processing the familiar stimuli from which they derive (7b) and more (7a). Using reading times, we showed that familiar stimuli (1) took less time to read following the most pleasurable innovation (3) than following the least pleasurable innovation (4), suggesting that the most pleasurable innovations meet requirement (7b) while least pleasurable innovations do not, attesting to the role of salience in inducing pleasure. Complementarily, we showed that innovative interpretations (9d) rated online as highly pleasurable following the idiomatic expression (9a) also took longer to read following this context than following a context that does not require such an interpretation (9b), but somewhat shorter than following an unrelated control context (9c). Compared to deriving the salient interpretation only, the innovative interpretation (9d) is more effortful, apparently because it involves accessing the salient response and more, thus meeting the optimal innovation criteria (7a-b). It is, however, less effortful than trying to make sense of an interpretation that is incoherent and does not involve any familiar meaning (9c) and does not induce pleasure:

- (9) a. You don't know your right from left?
- b. If you want to familiarize yourself with the political situation in Israel,
- c. If you are the type that sings in the shower,
- d. Buy *The Comprehensive Guide for the Political Factions in Israel*.

In Giora, Zur & Fein (in prep), we tested the optimal innovation hypothesis with regard to visual stimuli. We showed that visual stimuli rated most pleasurable (11) occupied mid position on the familiarity scale and involved the familiar (7b) and more (7a). Least pleasurable were stimuli rated as least familiar (12). Mid position on the pleurability scale was occupied by stimuli rated as highly familiar (10). It is salience then that accounts for pleurability, and recognizing the salient in the innovative is most pleasurable.

Taken together, such findings support the view that it is neither pure innovation nor familiarity alone that account for aesthetic judgments. Rather, it is optimal innovation–novelty that allows an insight into some salient response—that is most pleasurable. Indeed, optimal innovations (3, 9a/d, 11) are more pleasurable than familiar stimuli (1-2, 9b/d, 10). Though they share similar familiar responses, they also involve unfamiliar responses. Optimal innovations (3, 11) are also



(10) Familiar visual stimulus



(11) Optimally innovative visual stimulus



(12) Highly innovative visual stimulus

more pleasurable than pure innovations. Though both involve novel responses, only optimal innovations further involve familiar responses. Pleasure, then, is a function of both salience and innovativeness. It is the surprise experienced in suddenly discovering some novelty where it is least expected, or the gratification in discovering the familiar in the novel.

Indeed, familiarity (though less so excessive familiarity) is known to be a factor in pleasure or liking (on the various aesthetic effects of mere exposure, see Bornstein & D'Agostino, 1992; Harrison, 1977; Kunst-Wilson & Zajonc, 1980; Zajonc, 1968, 1980, 2000, and see also our findings above). It is not the most familiar, then, that is least enjoyable, but rather the most novel that is least pleasing. Pleasure, however, resides half way between high salience and high novelty.

3 ON RECOGNIZING THE NOVEL IN THE SALIENT: THE CASE OF JOKES

Jokes are considered paradigmatic in inducing affect. Are they optimally innovative? Consider the following example:

(13) War doesn't determine who's right but who's left.

The text in (13) meets the requirements in (7). It involves an innovative sense (just-survive) of the collocation *right-left* while invoking salient meanings (just, political orientations). It is quite plausible to assume that upon encountering *right*, the comprehender accesses the salient meaning (just) which, when *left* is encountered, is reinterpreted as political orientation. Eventually, the less salient meaning of *left* (survive) of *right-left* is accessed, ridiculing the previous responses.

The structure of most jokes is such that it keeps us attending to the salient response until the punchline point where a reversal is enforced allowing for the recognition of the novel. The pleasure derivable from this joke hinges on recognizing the innovative in the salient.

3.1 THEORETICAL BACKGROUND

Jokes are probably among the most amusing, funny, and laughable texts. Most theories of joke comprehension focus on one type of jokes—semantic jokes that involve double entendre². Looking into this genre, most theorists discussed jokes in terms of contextual incompatibility, which triggers a search for the resolution of that incongruity. According to Attardo (1994, 1996a,b, 2000, 2001), Attardo, Attardo, Baltes and Petray (1994), processing a joke begins with processing one sense (the 'just' meaning of the ambiguous word *right*) which is retained up to the point (*left*) at which the initial interpretation no longer makes sense. At this point, termed the disjunctur, the interpretation process is disrupted. If the disruption is minimal, it affects reinterpretation of the initial sense, resulting in a different/opposed sense (which would now be 'a political orientation and then 'just and alive'). If it is more than minimal, the comprehender may dismiss the text as ill formed. Attardo and his colleagues assume Raskin's (1985) view of jokes as centering on opposition relations obtaining between two senses/scripts.

For Coulson and Kutas (1998a,b, and see also Coulson, 2001: 77-82), it is script-shifting (termed 'frame-shifting') that accounts for joke comprehension (see below). Compared to alternative surprise endings that do not involve script or frame shifting, jokes' endings are costly processing-wise precisely because of the need to switch from one script or frame to another.

Curcó (1995, 1996a,b, 1998) assumes a different processing model whereby a key assumption that is weakly manifest (involving the 'political orientation' meaning of *right & left*), becomes strongly manifest upon encounter of the 'target assumption' (involving the 'remain alive/'survive' meaning of *left*). The incongruity between the two assumptions triggers this shift from weak to strong manifestation and makes the joke work. Such a view thus assumes that the initial interpretation of the ambiguity involved is retained for contrastive purposes.

De Palma and Weiner (1992), Giora (1991), and Weiner (1996) posit that the structure of jokes is categorial (à la Rosch, 1973), starting with the most prototypical (salient) member of the set (of things that war determines such as who wins - the right or the wrong) and ending with a marginal (less salient) member (who remains alive). Such a view suggests that a humorous 'surprise' does not rely on a stark difference between two interpretations. Rather, it depends on the likelihood of the less accessible/less-salient interpretation to be relevant to and included within the category proposed initially. A humorous surprise, then, does not constitute an entirely unthinkable option (Giora, 1991), but only a less salient one—one that should not but which nevertheless has escaped our attention because of our salience prone mind.

Giora (1991, 1995, in press, Chapter 6), however, further proposed that joke comprehension involves attenuation of the salient, but contextually incompatible meaning (the 'political orientation' interpretation of *right & left*) following the disjunctur point. For the joke to get through, the (salient) meaning activated initially should be suppressed, at least momentarily, and replaced by the (less salient) appropriate sense (The suppression hypothesis; see Gernsbacher & Robertson, 1995 for a similar view, and see Colston, Giora & Katz, 2000, and Vaid, Hull, Gerkens, Heredia, & Martinez, 2000 for supportive results discussed in Appendix 5).

3.2 THE ROLE OF SALIENCE AND CONTEXT IN JOKE COMPREHENSION

Most jokes make up a discourse that best exposes our tendency to opt for the salient, most probable concept first. Note how we are caught entirely off guard when we realize that a person who has a *drinking problem* is, in fact, one who spills liquid all over himself while drinking (*Flying High* by Abrahams, Zucker & Zucker, 1980). Jokes' pleurability, then, hinges on discovering the

²The jokes dealt with here either include polysemous or ambiguous interpretations. Polysemies involve related meanings; ambiguities involve unrelated meanings. According to Attardo et al. (1994), such jokes are the most common ones and highly typical among verbal jokes (as opposed to referential jokes, the most common type of joke).

novel in the familiar. The stimulus invites the comprehender to process a more salient, familiar meaning first ('drinking alcohol excessively') only to make her discover that a less salient, seemingly unthinkable meaning makes more sense.

To let our salience-prone mind go astray, the initial context of a joke is usually unambiguous, compatible with the salient meaning, so that this interpretation is retained exclusively up until the punchline, at which point a sudden incongruity forces a reinterpretation. The following joke (taken from Coulson & Kutas, 1998b) may serve to illustrate the collaboration between salience and context:

- (14) By the time Mary had had her fourteenth child,
she'd finally run out of names to call
her husband.

The salient meaning of *names* is associated with 'proper names'. In addition, the most plausible interpretation of *name calling* in the context of childbirth is also associated with 'proper names'. Furthermore, the most plausible interpretation of 'running out of *names*' in the context of the birth of a fourteenth child is further supportive of the 'proper names' interpretation. Salience and contextual information harmonize here: contextual information is highly supportive and predictive of the salient meaning of the polysemous word. The punchline, therefore, comes as a 'total' surprise.

Revisitation of salient meanings—discovering novelty in the familiar (7a-b)—rather than mere surprise ending (7a) may indeed be a major factor in accounting for the largely acknowledged difficulty and pleasurability of joke interpretation. Indeed, as shown by Coulson and Kutas (1998b), other forms of discourse involving less salient, unexpected items, which do not require reinterpretation of salient meanings, have been found to be less demanding and less funny. For instance, in their study, nonjoke texts (15a), having an equally unpredictable (low cloze probability) ending as jokes (15b), diverging from jokes in that they do not involve reinterpretation (of salient concepts), were faster to read than jokes (15b) and judged as less pleasurable:

- (15) By the time Mary had had her fourteenth child, she'd finally run out of names to call her
(a) Nonjoke ending: offspring
(b) Joke ending: husband

Crucially, then, for a joke to be enjoyed (understood as a joke), it should involve recognizing novelty in salient concepts³.

In sum, most semantic jokes make up an optimal innovation. They involve recognizing the novel in the familiar. The assumption that in addition, they further involve suppressing the familiar should wait further research (see Appendix 5).

4 ON RECOGNIZING THE NOVEL IN THE SALIENT: THE CASE OF NONCONVENTIONAL IRONY

Given the optimal innovation hypothesis (6), only novel or less-familiar ironies will be discussed here, since only they meet the requirements for optimal innovativeness (7a-b). By contrast, conventional ironies such as *wise guy*, or *tell me about it*, embedded in irony-inviting contexts, would not affect humor or pleasure, because they do not meet the requirement in (7a): no novelty is involved in their decoding. Indeed, with the exception of Giora (in press) and Giora & Fein (1999b), most of the literature on humor dealt with less familiar irony only.

Unlike jokes, irony relies entirely on context for its optimal innovativeness. The various theories on irony would, therefore, diverge with regard to the role context plays in irony comprehension.

³Not all jokes conform to the salient-less salient course of interpretation. When they don't, the contextually appropriate meaning derived initially needs to be revisited at the punchline position. It is hypothesized, however, that such reinterpretation will be less surprising as a result.

4.1 THEORIES OF IRONY COMPREHENSION

The field of irony research has received an immense boost in the last decade. For an extensive review of the field (which is beyond the scope of this chapter) see Attardo (2000) and Giora (1998, 2001, in press). Here I briefly sketch contemporary views of how we understand irony vis à vis the context in which it is embedded. Three major approaches dominate the field: (i) 'the direct access view', which counters (ii) 'the traditional, standard pragmatic model', and (iii) 'the graded salience hypothesis' that counters both. It is only according to the graded salience hypothesis and standard pragmatic model, however, that irony is optimally innovative and hence pleasing. According to the direct access view, however, irony should be treated as pure innovation whose decoding is entirely context dependent and involves only its innovative sense, and hence less pleasing.

4.1.1 THE INTERACTION-BASED, DIRECT ACCESS VIEWS

The interactionist, direct access view assumes that a constraining context affects comprehension primarily: a strong context governs processing significantly, affecting linguistic processes very early on. Consequently, in a rich and supportive context, irony comprehension need not involve a contextually incompatible (e.g., literal) stage at all. Rather, context should activate the contextually appropriate (ironic) interpretation exclusively so that only that meaning becomes available for further processes. In a strong context, then, irony comprehension should proceed seamlessly, and should not differ from processing equivalent literal language (see 'the relevance-theoretic account' and 'the allusional pretense view'). Equal reading times of items embedded in literally and ironically biasing contexts support this view (see Gibbs, 1986, but see Giora, 1995 for a critique; Giora, Fein & Schwartz, 1998; Schwoebel, Dews, Winner and Srinivas, 2000 for different findings). According to the direct access view, then, the sarcastic target in (16) should not be more pleasurable than the nonsarcastic target in (17):

- (16) Billy and Joe were long-time pals. But one time when Billy was away on a business trip, Joe slept with Billy's wife, Lynn. When Billy found out about it afterwards, he was upset. He confronted Joe and said to him:
[Sarcastic target] "You are a fine friend."
- (17) Billy and Joe were long-time pals. One time Billy was in desperate need of money. His car had broken down and he needed \$300 to fix it. So, he asked Joe for a loan. Joe said he could lend Billy the money. This made Billy happy and he said to Joe,
[Nonsarcastic target] "You are a fine friend."

4.1.1.1 The relevance theoretic account

The relevance theoretic account of irony (Gibbs, 1986; Jorgensen, Miller & Sperber 1984; Sperber & Wilson 1981, 1986/95; Sperber 1984; Wilson & Sperber 1992) is consistent with some aspects of the direct access hypothesis. Basically, it assumes that utterance interpretation relies on context for its decoding. According to Sperber and Wilson, context is generally not given in advance, but is searched for as part of the interpretation process. As a result, processing irony (19) need not differ from processing a nonironic utterance (18). Specifically, irony is an "echoic interpretive use in which the communicator dissociates herself from the opinion echoed with accompanying ridicule or scorn" (Wilson & Sperber 1992: 75). While its processing involves disengaging from an echoed opinion, or from what it assumes (Curcó, 2000), processing an equivalent nonironic utterance involves endorsing an echoed opinion. In both, then, there is an echoic allusion to be picked up which is invited by context (Sperber & Wilson 1986/95: 239):

- (18) a. He: It's a lovely day for a picnic.
[They go for a picnic and the sun shines.]
b. She: (happily): It's a lovely day for a picnic, indeed.
- (19) a. He: It's a lovely day for a picnic.
[They go for a picnic and it rains.]
b. She: (sarcastically): It's a lovely day for a picnic, indeed.

4.1.1.2 The allusional pretense view

Kumon-Nakamura, Glucksberg and Brown (1995) proposed a more general theory of discourse irony. According to the allusional pretense view, irony alludes to or reminds the addressee of what should have been—an expectation or a norm that went wrong (see also Kreuz & Glucksberg, 1989). To enable the addressee to appreciate the allusion, irony involves pragmatic insincerity, thus allowing various speech-acts to be ironic. For example, when a car driver says *I just love people who signal when turning* when the car ahead of her makes a turn without signaling, the speaker alludes to a social norm or expectation to signal upon turning, while simultaneously pretending to compliment the errant driver. Such view of irony assumes that irony comprehension involves activating the linguistic meaning of what is said in order to assess its sincerity and derive the ironic or nonironic interpretation, depending on the context. Thus, in any given situation, there is a decision to be made whether what is said is intended sincerely (i.e., literally) or insincerely (i.e., ironically). In this way, ironic and literal interpretations are equally affected by contextual information (Glucksberg, 1995).

The assumption that context plays a primary role in comprehension so that processing does not involve an incompatible phase views irony as pure innovation which does not expect to induce affect.

4.1.2 THE STANDARD PRAGMATIC MODELS

According to the various versions of the standard pragmatic model, processing involves an initial literal stage which, at times, will have to be revisited by late context effects (as in the case of irony). Such an approach might be more consistent with the optimal innovation hypothesis, provided the literal response to be revised is a familiar response.

4.1.2.1 The traditional view

The traditional view, notably, Grice (1975) and Searle (1979), assumes restricted contextual effects. According to this view, context affects comprehension at a late stage, only following linguistic processes, and only if these processes do not result in context fit. Thus, if I say *What a lovely day for a picnic* on a stormy day, my addressee would first compute the literal meaning of the statement, then reject it as the intended meaning and replace it with an alternative, contextually appropriate meaning. According to the traditional view, then, understanding nonliteral language involves a sequential process. The first stage is literal and obligatory, and the second stage is nonliteral and optional, induced by contextual information. In Gricean (1975, pp. 46ff) terms, the initial process involves a breach of a norm (primarily the Quality maxim). According to Attardo (2000), the violation should be minimally disruptive, though perceivable as disturbing contextual appropriateness. The overt, least disruptive violation is a signal for the addressee to reject the literal meaning as the intended meaning and derive the speaker's intention (dubbed 'particularized conversational implicature') which should replace the inappropriate interpretation. In contrast, literal language involves initially just one process. Therefore, understanding literal and nonliteral language should differ, with nonliteral language requiring a double take. Longer reading times found for utterances embedded in ironically vs. literally biasing contexts support the standard pragmatic model (Giora et al., 1998; Schwoebel, Dews, Winner & Srinivas, 2000).

Consistent with the modular view (Fodor, 1983), which proposes that the contextually inappropriate meanings activated during the first initial access phase should be suppressed only subsequently, the standard pragmatic model assumes that the contextually incompatible literal meaning of irony should be suppressed and replaced by a contextually compatible ironic meaning. Thus, *What a lovely day for a picnic* said on a stormy day is rejected as contextually incompatible and replaced by its approximate opposite—*What a lousy day for a picnic*. In this view, then, irony comprehension involves a suppression process at the second integration phase. It therefore differs from processing its equivalent literal interpretation. This suppression assumption, however, has not gained empirical support (see Giora, in press; Giora & Fein, 1999b; Giora et al., 1998).

4.1.2.2 The relevant inappropriateness assumption

Attardo (2000) provides a more general account of irony than the traditional view by going beyond the maxim violation condition posited by Grice (1975). According to Attardo, irony need not violate any maxim. Rather, while assuming the Relevance maxim (for the second, integration phase), it should breach contextual appropriateness ostensibly at the initial phase, so that the comprehender may detect the intended violation and derive the ironic interpretation. For example, when, in a drought-stricken area, one farmer says to another *Don't you just love a nice spring rain?* the utterance may be true, yet inappropriate, given the situation of utterance (it is not raining). According to Attardo, violation of contextual appropriateness includes violation of both sincerity and cultural norms or expectations (assumed necessary for irony interpretation by the allusional pretense, see above) and more (e.g., deictic inappropriateness). Indeed, optimal innovations involve a minimal disruption or change—one that does not obscure the familiar response while inviting a novel response.

4.1.2.3 The joint pretense view

The joint pretense view (Clark & Carlson, 1982; Clark & Gerrig, 1984; Clark, 1996) is also inspired by the Gricean view (Grice, 1978). It assumes a speaker who pretends "to be an injudicious person speaking to an uninitiated audience; the speaker intends the addressee of the irony to discover the pretense and thereby see his or her attitude toward the speaker, the audience, and the utterance" (Clark & Gerrig, 1984: 12; for a similar view see Boulton as quoted in Booth, 1974: 105). By saying *What a lovely day for a picnic* on a stormy day, the ironist assumes the identity of another speaker addressing a gullible audience. The present addressee, however, is supposed to take delight in recognizing both the pretense and the intended attitude of ridicule toward the pretending speaker, the audience, and the utterance. According to Clark (1996: 368), joint pretense is conceived of as a staged communicative act (see also Haiman, 1998; Kotthoff, 1998) where the actual speaker is also an implied speaker performing a sincere communicative act toward an implied addressee who is also the actual addressee. Both actual participants are intended to "mutually appreciate the salient contrasts between the demonstrated and actual situations", so that if asked, the actual speaker would deny meaning for the actual addressee what the implied speaker means for the implied addressee.

According to this view, irony is a two-layered act of communication in which the literal (which is in most cases the salient) meaning is activated and retained by both the speaker and the addressee, who reject it as the intended meaning though they pretend otherwise.

4.1.2.4 The tinge hypothesis

The tinge hypothesis (Dews, Kaplan & Winner, 1995; Dews & Winner, 1995, 1997a, 1999) assumes that irony is used to mute the intended negative criticism. The positive literal meaning of irony (*That was really funny* said on a mean joke) tinges the addressee's perception of the intended meaning. Similarly, the negative literal meaning of ironic compliments (*It's a tough life* said to someone on vacation) mitigates the positively intended meaning. Dews and Winner and their colleagues assume that the contextually incompatible, literal meaning of ironic remarks is processed at some level and interferes with the intended meaning. Following Long and Graesser (1988), they propose a dual-process model "in which comprehension may occur after the recognition of an incongruity or simultaneously" (Dews & Winner, 1997: 405). According to the tinge hypothesis, then, the literal meaning of irony is activated initially, either before or alongside the ironic meaning, and is retained in order to dilute the implicit criticism or compliment. Using a Stroop like interference paradigm, Dews and Winner tested this hypothesis. They asked subjects to judge the intended (rather than the literal) meaning of ironic utterances and recorded their responses. If utterances are shown to take longer to be judged as positive or negative relative to their literal interpretations, this suggests that the contextually incompatible, literal meaning is accessed automatically, interferes with the process, and slows it down. Consistent with the tinge hypothesis, ironies such as *What a lovely day for a picnic* were judged as less aggressive than their literal counterpart *What a lousy day for a picnic* and took longer to be judged as positive or negative relative to their literal interpretations. Dews and Winner concluded that, unlike literal language, irony comprehension involves an obligatory, contextually incompatible, literal phase

(Dews et al., 1995; Dews & Winner, 1997; Winner, 1988)⁴.

4.1.3 THE GRADED SALIENCE HYPOTHESIS

Contra the interactionist view, and partly following Fodor's (1983) modular assumptions, the graded salience hypothesis assumes distinct mechanisms, linguistic and contextual, that run in parallel without interacting initially (Giora, 1997, 1999, in press; Giora & Fein, 1999b; Giora, Fein & Schwartz, 1998; Giora, Peleg, & Fein, 2001; Peleg, Giora & Fein, 2001, in press). The linguistic (lexical) mechanism is a bottom-up, encapsulated machinery that is sensitive only to linguistic information. In contrast, the contextual mechanism involves top-down processes that are sensitive to both linguistic and extra-linguistic knowledge. Unlike the traditional modular assumption (Fodor, 1983), however, the graded salience hypothesis assumes that the encapsulated, lexical access mechanism is ordered: more salient meanings—coded meanings foremost on our mind due to conventionality, frequency, familiarity, or prototypicality—are accessed faster than and reach sufficient levels of activation before less salient ones. According to the graded salience hypothesis, then, coded meanings would be accessed upon encounter, regardless of contextual information or authorial intent. Coded meanings of low salience, however, may not reach sufficient levels of activation to be visible in a context biased toward the more salient meaning of the word (but see Hillert & Swinney, 2001 for a different view).

Contextual, top-down processes may also affect comprehension immediately. A constraining and highly predictive context may avail meanings on its own accord very early on. Nevertheless, it would not penetrate lexical access. Though it has a predictive role that may speed up derivation of the appropriate meaning, it would not obstruct inappropriate, coded meanings upon encounter of the lexical stimulus. Indeed, contextual information may be strong and even faster than lexical processes, so much so that it may avail meanings even before the relevant stimulus is encountered, fostering an impression of direct access. This may be particularly true when the stimulus is placed at the end of a strong sentential context, after most information has been accumulated and integrated, allowing effective guessing and inferential processes. However, these inferential processes do not interact with lexical processes but run in parallel (Giora et al., 2001; Peleg et al., 2001, in press). As shown in Giora et al. and Peleg et al., assuming a simultaneous operation of the encapsulated, linguistic mechanism and the integrative, central system mechanism allows the graded salience hypothesis to predict when contextual information may be faster than, coincidental with, or slower than linguistic processes. Unlike the traditional views, then, the graded salience hypothesis does not always predict slower contextual effects that result in sequential processes. Neither does it assume (as Grice, 1975 does) that activation of a whole linguistic unit should be accomplished before contextual information comes into play. Rather, across the communication path, context and linguistic effects run in parallel, with contextual information availing meanings on its own accord, affecting only the end product of the linguistic process.

As illustration, consider the processes involved initially in trying to make sense of *What a lovely day for a picnic* said on a stormy day. According to this view, upon encounter, the processor would automatically retrieve the salient (literal) meaning of *lovely* ('nice') from the mental lexicon. However, given simultaneous top-down processes, this process will not result in contextual fit. This accidental mismatch between lexical (salient) and contextual (novel) meanings will result in an optimally innovative ironic interpretation. If said on a sunny day, however, the automatically retrieved, lexical meaning will achieve contextual fit, with no novelty involved. The graded salience hypothesis, then, would predict longer reading times for an utterance embedded in an ironically than in a literally biasing context. Note, however, that irony need not build on a salient literal meaning. Its salient meaning can also be figurative as when the conventional metaphor *Children are precious gems* can be used ironically, or when a conventional irony is used innovatively (*Read my lipstick*). In such cases, it is also some nonliteral meaning that should be processed initially

⁴However, though over-polite requests that were perceived as ironic were rated less rude than their (nonironic) under-polite counterparts, they were not rated as less insulting (Kumon-Nakamura, Glucksberg & Brown, 1995). According to some accounts, irony may be a politeness strategy without muting the criticism. In fact Colston (1997) and Toplak & Katz (2000) argue that irony is used to enhance rather than dilute condemnation.

on account of its salience (cf. Pexman et al., 2000). Thus, if ironies and literals differ processing-wise, it is not the literal-nonliteral dichotomy that accounts for their difference, but the salience-nonsalience continuum.

The indirect negation view of irony (Giora, 1995) supplements that graded salience hypothesis with regard to integration processes. According to this view, irony is a form of negation that does not make use of an explicit negation marker. Often, an affirmative (*What a lovely day for a picnic* said on a stormy day) rather than a negative (*What a lousy day for a picnic* said on a sunny day) expression is used to implicate that a specific state of affairs is different or far from the taken for granted, expected (or more desirable) state of affairs made explicit by the expression. Such a view assumes that irony comprehension involves activating the salient, often literal meaning initially. However, it does not assume that the indirectly negated meaning is suppressed and replaced by its opposite, as suggested by the traditional account. Rather, irony entertains both the explicit and derived messages, so that the dissimilarity between them may be computed. By saying *What a lovely day for a picnic* on a stormy day, the ironist points out the extent to which the criticized object (weather) has fallen short of expectations, and is far from being 'lovely'. The indirect negation view thus predicts that the explicit (often) literal meaning of irony activated initially would be retained for purposes of irony interpretation.

As in the case of the standard pragmatic view, longer reading times of nonsalient (ironic) versus more salient (literal or nonliteral) items and longer response times to nonsalient (ironically) related than to salient (literally or nonliterally) related probes support the graded salience hypothesis (Giora, Fein & Schwartz, 1998; Giora & Fein, 1999a; Pexman et al., 2000; Schwoebel et al., 2000). In addition, demonstrated retention rather than suppression of the salient (often) literal meaning of irony supports the direct negation view (see findings in Giora, Fein & Schwartz, 1998; Giora & Fein, 1999b). Empirical evidence, then, supports the graded salience hypothesis and the indirect negation view, and questions the direct access and standard pragmatic models. It shows that, contra the direct access view, salient meanings get activated regardless of context. It further shows that, contra the standard pragmatic model, salient meanings get activated regardless of literality. Diverging from the traditional pragmatic model, it also shows that salient but inappropriate meanings are not suppressed as irrelevant but retained for further processes.

The graded salience hypothesis and indirect negation view and the various versions of the standard pragmatic model, which assume that irony involves processing (aspects of) its literal interpretation, predict that unfamiliar irony will be more pleasurable than its more familiar literal interpretation. Given the optimal innovation hypothesis (7), the graded salience hypothesis further predicts that unfamiliar irony will be more pleasurable than familiar irony (*Read my lips* vs. *Read my lipstick*) and that familiar metaphor (*Children are precious gems*) will be less pleasurable than its nonsalient ironic interpretation (invited by an irony inducing context). These predictions remain to be validated.

5 CONCLUSIONS

What stimuli would incur liking or pleasure? Is it the familiar that is most likable? Is it total novelty that is most pleasing? Is it the figurative that is most enjoyable? The optimal innovation hypothesis predicts that optimal innovation–novel stimuli allowing an insight into the familiar–will be appreciated as most pleasurable, regardless of figurativity.

Findings indeed show that it is optimal innovation that is judged as most pleasing (Giora, in press, Giora, Kronrod et al., 2001; Giora et al., in prep; Kronrod et al., 2000). And though it is costly, because it involves entertaining multiple interpretations simultaneously, optimal innovation is rewarding. On some accounts, its reward is even long-lasting (see Berg & Lippman, 2001; Derks, Gardner, & Agarwal, 1998; Gardner, J. B., & Agarwal Gibbs, 1986; Kreuz, Long & Church, 1991; Lippman & Dunn, 2000; Mcaninch, Austin, & Derks, 1993; Schmidt, 1994; Schmidt & Williams, 2001; Unger, 1996; Vangiffen, & Maher, 1995 on humor and its effect on memory and see also Darling & Civikly, 1987; Kaplan & Pascoe, 1977; McMorris, Urbach, & Connor, 1985; Powell & Andresen, 1985; Ziv, Gorenstein, & Moris, 1986 on the effects of humor on retention of materials).

Appendix A

The theory of joke comprehension proposed here and elsewhere (Giora, 1991) assumes that understanding an optimally innovative discourse such as jokes involves a reinterpretation process in which a salient response to a key word or expression is accessed initially and retained until the disjunctive position. Following this position, it is suppressed to allow for a less salient (but globally compatible) response to emerge. At this point, the initial response has no role in constructing the intended response. Rather, it may stand in the way and interfere with comprehension. According to the retention/suppression hypothesis (Giora, in press), meanings are retained as long as they are conducive to the interpretation process but discarded if they interfere with comprehension (cf. Gernsbacher, 1990). Indeed, comprehenders who are less able at suppressing salient but incompatible meanings have been shown to be poor comprehenders (Gernsbacher, 1990, 1994).

Colston, Giora & Katz (2000) tested the suppression hypothesis, aiming to show that the meaning activated initially is suppressed later on, at offset of the punchline sentence. Results indeed demonstrate that the meaning of the key word, assumed to be salient, was available immediately after offset of the key word (at the end of the first line, see [a] below). Similarly, after offset of the second segment (b), its levels of activation were still marginally significant. However, at the final punchline position (c), the initially activated meaning was no longer available: It was no more activated than the unrelated probe, suggesting that, at this point, it underwent suppression.

- (a) My friend asked me to look at his ear*
 - (b) but it was covered with butter*
 - (c) and salt.*
-
probes (displayed at *)

Vaid et al. (2000) report similar results. They tested one-liner jokes (*I still miss my ex but my aim is getting better*) which were displayed for 4500 ms. They showed that 150 ms after offset of the joke, only the probe related to the contextually appropriate meaning was primed ('hate'), whereas the salient meaning ('love') related to the ambiguous word (*miss*) was not. These results are consistent with the suppression hypothesis, suggesting that after a sufficiently long delay allowing for joke processing to be complete, the salient but contextually inappropriate meaning of the ambiguous word was deactivated.

The suppression hypothesis may indeed account for other findings regarding joke comprehension. Given that suppression comes with a cost (Gernsbacher, 1990), the suppression hypothesis predicts that an interpretation process that does not require suppression would be less costly than one that requires such a process. This prediction is consistent with the findings in Coulson and Kutas (1998b) discussed earlier. Consider, again, example (15) above. While both *husband* and *offspring* were shown to be similarly unpredictable or probable given the preceding context (having cloze probability of 4% and 2% respectively), only *husband* would trigger suppression of the salient meaning of ('proper') *name* so that a less salient concept ('epithets associated with insult') be activated. In contrast, though *offspring* is a low probability ending, suppressing the originally highly salient meaning of *name* is not required. On the contrary, like the probable ending 'child', *offspring* involves retaining the notion of 'proper name' with which it is consistent, since, like 'child', it represents the same salient concept, only by means of an unexpected word. No wonder it took less time to process than *husband*. True, cloze probability is tailored to detect context predictive effects. However, in the specific studies discussed, it also indexes salience (see Giora, in press, Chapter 6 for a full discussion).

Or, consider another example (taken from Coulson & Kutas, 1998b):

- (d) He is so modest he pulls down the shade to change his
 Joke ending: mind
 Nonjoke ending: jacket
 Conventional ending: clothes

According to the suppression hypothesis, for the joke to get through, the salient meaning of *change* ('replace a material object', 'take off cloths'), which is also the one invited by the context (having a cloze probability of 42%), has to be suppressed so that the more novel, less salient meaning of *change* ('change of an immaterial object'), having a cloze probability of 6%, be evoked. Though *changing a jacket* is a less salient, less frequent expression compared to '*changing clothes*' (having a cloze probability of 3%), it need not involve suppressing the salient ('physical replacement') meaning of *change*, because it is an extension of that self-same meaning. The nonjoke ending, then, is not an optimal innovation: no novel sense or concept was involved here. Indeed, subjects took less time to read the nonjoke (*jacket*) than joke (*mind*) ending.

Findings from event-related brain potentials lend further support to the hypothesis that joke comprehension involves suppression. In another study, Coulson and Kutas (1998a and see also Coulson, 2001) compared event-related brain potentials recorded from the scalp as participants read joke and nonjoke sentences. The difficulty of integrating a given word into an established context is correlated with N400 amplitude. N400 amplitude is largest for items with low cloze probability such as semantic anomalies, and smallest for easily integrated items with high cloze probability such as best completions. Accordingly, the suppression hypothesis predicts that joke comprehension would involve N400 amplitude to a greater extent than nonjokes. It also predicts that the effect would be greatest for jokes involving highly salient meanings compared to those involving less salient meanings, since salient meanings are harder to suppress (Giora, in press). Moreover, it predicts that this effect would be more pronounced among good than among poor comprehenders who are deficient at suppression (cf. Gernsbacher & Robertson, 1995).

Findings in Coulson and Kutas (1998a) indeed serve to support the suppression hypothesis. They show that jokes elicited larger N400 components than nonjokes (ending with similarly cloze probability items), and that this effect was greatest where highly salient meanings were involved. Furthermore, they showed that this pattern was most pronounced among participants whose performance on the comprehension questions suggested they understood most of the jokes, suggesting that they suppressed the incompatible, though salient meaning (for a detailed discussion see Giora, in press, Chapter 6).

For Coulson and Kutas (Coulson, 2001; Coulson & Kutas, 1998a,b), however, these findings support the hypothesis that joke comprehension involves a 'frame-shift'. Frame shifting occurs when elements of a given message-level representation are mapped onto a new frame. Jokes, they contend, are deliberately constructed to suggest one frame while evoking elements also consistent with another. The notion of frame shifting is not inconsistent with the assumption of the graded salience hypothesis concerning joke comprehension. It is plausible to assume that a word meaning involves activating its unmarked contextual information or frame. Moreover, jokes do not always hinge on explicit word meanings but rather on frame anticipation. Though the following joke requires the accessing of a less salient reading of *something* as referring to humans, contextual information derived from the 'frame' plays a crucial role in its initial interpretation (as 'a drink'):

- (e) When I asked the bartender for something cold and full of rum, she recommended her husband.

Recall however that optimal innovations require an extra processing effort compared to familiar responses, because they involve processing the salient and more. Findings in Coulson and Kutas, therefore, may not speak to the issue of suppression and may only demonstrate that jokes are optimal innovations. Even the finding that good comprehenders invest more effort than poor comprehenders in processing jokes might be only indicative of poor comprehenders' limited ability to entertain multiple messages on account of limited mental resources (see Miyake, Just & Carpenter, 1994).

REFERENCES

- Attardo, S. (1994). *Linguistic theories of humor*. Berlin: Mouton de Gruyter.
- Attardo, S. (1996a). Humor. In J. Verschueren, J-O Östman, J. Blommaert, & C. Bulcaen (Eds.), *Handbook of Pragmatics* (pp. 1-18). Amsterdam/Philadelphia: John Benjamins.
- Attardo, S. (1996b). Humor theory beyond jokes: The treatment of humorous texts at large. In J. Hulstijn and A. Nijholt (Eds.), *Automatic Interpretation and Generation of Verbal Humor*. IWCH '96. Twente Workshop on Language and Technology 12 (pp. 87-101). Enschede: University of Twente.
- Attardo, S. (2000). Irony as relevant inappropriateness. *Journal of Pragmatics*, 32, 793-826.
- Attardo, S. (2001). *Humorous texts: A semantics and pragmatics analysis*. Berlin: Mouton De Gruyter.
- Attardo, S., Attardo, D. H., Baltés, P., & Petray, M. J. (1994). The linear organization of jokes: analysis of two thousand texts. *Humor*, 7, 27-54.
- Berg, E. M. & Lippman, L. G. (2001). Does humor in radio advertising affect recognition of novel product brand names? *Journal of General Psychology*, 128, 194-202.
- Bornstein, R. F., & D'Agostino, P. R. (1992). Stimulus recognition and the mere exposure effect. *Journal of Personality and Social Psychology*, 63, 545-552.
- Brinker, M. (1988). *Aesthetics as the theory of criticism*. Tel Aviv: Broadcast university/The Ministry of Defense, Israel.
- Brisard, F., Frisson, S., & Sandra, D. (2001). Processing unfamiliar metaphors in a self-paced reading task. *Metaphor and Symbol*, 16, 87-108.
- Colston, H. L. (1997). Salting a wound or sugaring a pill: The pragmatic functions of ironic criticism. *Discourse Processes*, 23, 25-45.
- Colston, H. L., Giora, R., & Katz, A. (2000, July). Joke comprehension: Salience and context effects. Paper presented at The 7th International Pragmatics Conference, Budapest.
- Coulson, S. (2001). *Semantic leaps*. Cambridge and New York: Cambridge University Press.
- Coulson, S., & Kutas, M. (1998a). Frame-shifting: Event-related brain response to jokes. Paper submitted for publication.
- Coulson, S., & Kutas, M. (1998b). Frame-shifting and sentential integration. Paper submitted for publication.
- Curc6, C. (1995). Some observations on the pragmatics of humorous interpretations: a relevance theoretic approach. *UCL Working papers in Linguistics*, 7, 27-47.
- Curc6, C. (1996a). The implicit expression of attitudes, mutual manifestness, and verbal humor. *UCL Working Papers in Linguistics*, 8, 89-99.
- Curc6, C. (1996b). Relevance theory and humorous interpretations. In N. Hulstijn, & A. Nijholt (Eds.), *Automatic interpretation and generation of verbal humor* (pp. 53-68). Twente: University of Twente.
- Curc6, C. (1998). Indirect echoes and verbal humor. In V. Rouchota, & A. H. Jucker (Eds.), *Current issues in relevance theory* (pp. 305-325). Amsterdam and Philadelphia: John Benjamins.
- Curc6, C. (2000). Irony: negation, echo and metarepresentation. *Lingua*, 110, 257-280.

- Darling, A. L., & Civikly, J. M. (1987). The effect of teacher humor on student perceptions of classroom communicative climate. *Journal of Classroom Interaction*, 22, 24-30.
- De Palma, P., & Weiner, J. E. (1992). Accessibility and knowledge representation. *Proceedings of COLING '92*, 1121-1125.
- Derks, P., Gardner, J. B., & Agarwal, R. (1998). Recall of innocent and tendentious humorous material. *Humor-International Journal of Humor Research*, 11, 5-19.
- Dews, S., & Winner, E. (1997). Attributing meaning to deliberately false utterances: The case of irony. In Ch. Mandell, & A. McCabe (Eds.), *The problem of meaning: Behavioral and cognitive perspectives* (pp. 377-414). Amsterdam: Elsevier.
- Dews, S., & Winner, E. (1999). Obligatory processing of the literal and nonliteral meanings of ironic utterances. *Journal of Pragmatics*, 31, 1579-1599.
- Elnatan, I. (2000). Enticement: Aesthetic effects of conventional versus innovative uses of language. Unpublished paper, Tel Aviv University (in Hebrew).
- Elnatan, I. (2002). Degree of innovation and pleasure rating. Unpublished MA thesis, Tel Aviv University (in Hebrew).
- Freud, S. (1905/1960). *Jokes and their relation to the unconscious*. New York: Norton.
- Gibbs, R. W. Jr. (1986). On the psycholinguistics of sarcasm. *Journal of Experimental Psychology: General*, 115, 3-15.
- Gibbs, R. W. Jr. (in press). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics*.
- Giora, R. (1991). On the cognitive aspects of the joke. *Journal of Pragmatics* 16, 465-486.
- Giora, R. (1995). On irony and negation. *Discourse Processes*, 19, 239-264.
- Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, 7, 183-206.
- Giora, R. (1998a). Irony. In: J. Verschueren, J-O Östman, J. Blommaert, & C. Bulcaen (Eds.), *Handbook of pragmatics* (pp. 1-21). Amsterdam: Benjamins.
- Giora, R. (1998b). When is Relevance? On the role of salience in utterance interpretation. *Revista Alicantina de Estudios Ingleses*, 11, 85-94.
- Giora, R. (1999). On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics*, 31, 919-929.
- Giora, R. (2001). Irony and its discontent. In G. Steen, & D. Schram (Eds.), *Psychology of language: In honour of Elrud Ibsch* (pp. 163-182). Amsterdam: John Benjamins.
- Giora, R. (in press). *On our mind: Salience, context and figurative language*. New York: Oxford University Press.
- Giora, R., & Fein, O. (1999a). On understanding familiar and less-familiar figurative language. *Journal of Pragmatics*, 31, 1601-1618. (Special issue on 'Figurative and literal language', ed. by R. Giora).
- Giora, R., & Fein, O. (1999b). Irony: Context and salience. *Metaphor and Symbol*, 14, 241-257.
- Giora, R., & Fein, O. (1999c). Irony comprehension: The graded salience hypothesis. *Humor*, 12, 425-436.

- Giora, R., Fein O., & Schwartz, T. (1998). Irony: Graded salience and indirect negation. *Metaphor and Symbol, 13*, 83-101.
- Giora, R., Kronrod, A., Elnatan, I., & Fein, O. (2001, July). The role of salience in aesthetic creativity. Paper presented at the Eleventh Annual Meeting of The Society for Text and Discourse. The University of California, Santa Barbara.
- Giora, R., Peleg, O., & Fein, O. (2001, May). Resisting contextual information: You can't put a salient meaning down. Paper presented at the international workshop Towards an Experimental Pragmatics. The University of Lyon.
- Giora, R., Zur, A., & Fein, O. (in prep). Degree of innovation and pleasure rating of visual stimuli. Tel Aviv University.
- Haiman, J. (1998). *Talk is cheap: Sarcasm, alienation, and the evolution of language*. Oxford: Oxford University Press.
- Harrison, A. A. (1977). Mere exposure. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 610-646). New York: Academic Press.
- Kaplan, R., & Pascoe, G. C. (1977). Humorous lectures and humorous examples: some effects upon comprehension and retention. *Journal of Educational Psychology, 69*, 61-65.
- Keysar, B., Shen, Y., Glucksberg, S., & Horton, S. W. (2000). Conventional language: How metaphorical is it? *Journal of Memory and Language, 43*, 576-593.
- Kreuz, J. R., Long L. D. & Church B. M. (1991). On being ironic: Pragmatic and mnemonic implications. *Metaphor and Symbolic Activity, 3*(6), 149-162.
- Kronrod, A. (2001). On the role of salience in optimal innovation: Aesthetics as a function of the old/new ratio. Unpublished MA thesis. Tel Aviv University.
- Kronrod, A., Giora, R., & Fein, O. (2000, September). Creative writing: The optimal creative innovation in fixed expressions. The EARLI special interest group writing conference 2000. University degli Studi di Verona.
- Kumon-Nakamura, s. Glucksberg, S., & Brown, M. (1995). How about another piece of the pie: The allusional pretense theory of discourse irony. *Journal of Experimental psychology: General, 124*, 3-21.
- Kunst-Wilson, W. R. & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science, 207*, 1019-1024.
- Lippman, L. G. & Dunn, M. L. (2000). Contextual connections within puns: Effects on perceived humor and memory. *Journal of General Psychology, 127*, 185-197.
- Miyake, A., Just, M. A., & Carpenter, A. P. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language, 33*, 175-202.
- Meaninch, C. B., Austin, J. L., & Derks, P. L. (1993). Effect of caption meaning on memory for nonsense figures. *Current Psychology-Research & Reviews, 11*, 315-323.
- McMorris, R. F., Urbach, S. L., & Connor, M. C. (1985). Effects of incorporating humor in test items. *Journal of Educational Measurement, 22*, 147-155.
- Mukarovský, J. (1932/1964). Standard language and poetic language. In P. L. Garvin (Ed.), *A Prague School reader on esthetics, literary structure, and style* (pp. 17-30). Washington, DC: Georgetown University Press.

- Mukarovsky, J. (1978). *Structure, sign and function*. Yale University Press.
- Peleg, O., Giora, R., & Fein, O. (2001). Salience and context effects: Two are better than one. *Metaphor and Symbol, 16*, 173-192.
- Peleg, O., Giora, R., & Fein, O. (In press). Contextual strength: The whens and hows of context effects. In I. Noveck & D. Sperber (Eds). *Experimental Pragmatics*. Palgrave.
- Pexman P., Ferretti, T., & Katz, A. (2000). Discourse factors that influence irony detection during on-line reading. *Discourse Processes, 29*, 201-222.
- Piven, H. (1999). *Piven, Avodot 1990-1999* (Piven, Works 1990-1999). Tel Aviv: Am Oved.
- Pexman P., Ferretti, T., & Katz, A. (2000). Discourse factors that influence irony detection during on-line reading. *Discourse Processes, 29*, 201-222.
- Powell, J. P., & Andresen L.W. (1985). Humour and teaching in higher education. *Studies in Higher Education, 10*, 79-90.
- Raskin, V. (1985). *Semantic mechanisms of humor*. Dordrecht: Reidel.
- Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111-144). New York: Academic Press.
- Shklovsky, V. (1917/1965). Art as technique. In L. T. Lemon & M. J. Reis (Eds. and Trans.), *Russian formalist criticism: Four essays*, 3-57. Lincoln, NE: University of Nebraska Press.
- Schmidt, S. R. (1994). Effects of humor on sentence memory. *Journal of Experimental Psychology: Learning Memory and Cognition, 20*, 953-967.
- Schmidt, S. R. & Williams, A. R. (2001). Memory for humorous cartoons. *Memory & Cognition, 29*, 305-311.
- Steinhart, E. & Kittay, E. F. (1998). Metaphor. In J. L. Mey (Ed). *Concise encyclopedia of pragmatics* (pp. 576-581). Amsterdam: Elsevier.
- Toplak, M., & Katz, A. N. (2000). On the uses of sarcastic irony. *Journal of Pragmatics, 32*, 1467-1488.
- Unger, L. S. (1996). The potential for using humor in global advertising. *Humor-International Journal of Humor Research, 9*, 143-168.
- Utsumi, A. (2000). Verbal irony as implicit display of ironic environment: Distinguishing ironic utterances from nonirony. *Journal of Pragmatics, 32*, 1777-1807.
- Vaid, J., Hull, R., Gerkens, D., Heredia, & R., Martinez, F. (November, 2000). The time course of script activation in verbal humor processing: Preliminary evidence. Poster presented at the 41st annual meeting of the Psychonomic Society, New Orleans.
- Vangiffen, K., & Maher, K. M. (1995). Memorable humorous incidents: gender, themes and setting effects. *Humor-International Journal of Humor Research, 8*, 39-50.
- Weiner, J. E. (1996). Why is a riddle not like a metaphor? In J. Hulstijn and A. Nijholt (Eds.), *Automatic interpretation and generation of verbal humor* (pp. 111-119). IWCH '96. Twente Workshop on Language and Technology 12. Enschede: University of Twente.
- Yus, F. (1998). A decade of relevance theory. *Journal of Pragmatics, 30*, 305- 345. Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1-27.

- Zajonc, R. B. (1980). Feeling and thinking: Preference need no inferences. *American Psychologist*, *35*, 151-175.
- Zajonc, R. B. (2000). Closing the debate over the independence of affect. In J. P. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition* (pp. 31-58). Cambridge, United Kingdom: Cambridge University Press.
- Ziv, A., Gorenstein, E. & Moris, A. (1986). Adolescents' evaluation of teachers using disparaging humour. *Educational Psychology*, *6*, 37-44.

Frame Blends, Analogies, and Humor

Douglas Hofstadter
Center for Research on Concepts and Cognition
Indiana University
Bloomington, U.S.A
dughof@cs.indiana.edu

Abstract

Analogy is usually seen as a correspondence between the internal structures of two cleanly separated situations or "mental spaces". In fact, however, when humans use analogy, most of the time the two spaces are blurred in subtle manners. Such a blurring of distinct situations is termed by some a "frame blend", by others a "blended space". Many blended spaces, though in some sense incoherent, seem completely normal, and often they provide the most natural way of expressing some complicated or subtle idea. Sometimes, however, blended spaces are shockingly incongruous, and this incongruity is the source of humor. I will show some cartoons taken from a newspaper, in which the humor is totally dependent on the strangeness of the blend made. Then I will show how each cartoon can be "tweaked" in various manners, making its blend of spaces vary in several different dimensions. It is interesting to ask oneself if a very good cartoonist always intuitively hits the optimal blend of spaces. From this I will move into more formal or abstract domains in which analogies and frame blends can be constructed, and I will show how humor, or at least a humor-like effect, can arise in an entirely formal domain when blending is done in a sufficiently incoherent manner. I will also map some standard jokes onto very abstract caricatures of them, in order to show how one and the same humorous phenomenon can arise in two vastly different domains. Finally, I will conclude with the challenge of trying to "translate" jokes from one domain into another, preserving their essence.

Quo Vadis Computational Humor?

Victor Raskin
Purdue University
West Lafayette, Indiana 47907-1356 USA
vraskin@purdue.edu

Abstract

The paper reviews the state of the art in computational humor and proposes a practical problem- and application-oriented approach to it, based on the ontological semantic approach to comprehensive and enriched text meaning representation. This approach minimalizes the gap between the need to base computational humor on top of a fully-developed NLP system and the current state of the art in NLP. The paper also urges the field to expand the range of its application and to go for them and for the funds to implement them competitively and aggressively.

Keywords: computational humor, natural language processing, text meaning representation, business applications

INTRODUCTION

This paper is a follow-up, five and a half years later, of the original Twente International Workshop on Computational Humor paper (Raskin 1996). In fact, the title of this paper is borrowed from a section title of that paper. The earlier paper addressed three issues:

- the need for basing computational humor on a well-defined formal rigorous theory and the script-based semantic theory of humor (SSTH: Raskin 1985) and its extension/revision, the general theory of verbal humor (GTVH: Attardo and Raskin 1991, Attardo 1994), as the prime candidates for this role;
- the feasibility of computational humor before the general problem of natural language processing (NLP), including full meaning representation, has been largely solved;
- the ontological-semantic basis for all work in computational humor.

Significant progress can be reported on all of these “fronts.” In humor theory, not only has the approach gained more currency and supporters (and some very useful detractors) but it has also been successfully extended to longer and more complex texts (Attardo 2001, Ermida 2002). A further extension and revision of GTVH is currently underway (Attardo and Raskin 2003). In NLP, the task of full meaning representation has been demonstrated to be much more feasible, and while problems remain, the disconnect between the needs of computational humor and what NLP can provide has been significantly minimalized. The ontological semantic approach, the theoretical and computational implementation of the early notion of script in SSTH, has been much elaborated upon and now supports a multiplicity of higher-end applications. The situation is much more favorable for a full-steam implementation of real-life, non-toy computational humor systems.

This paper briefly outlines the state of the art in computational humor and focuses on the ontological semantic support for its development and implementation. Section 2, the central part of the paper, draws heavily on Nirenburg and Raskin (2002), the forthcoming major statement on the approach, and I would like to thank my co-author for his kind permission to use some jointly achieved results here.

1 WHY COMPUTATIONAL HUMOR?

1.1 BRIEF HISTORY OF COMPUTATIONAL HUMOR

There are two major motivations for doing something new in research:

- the Everest, “because it’s there,” motivation; and
- the “imperative of usefulness.”

There are more academic ways of referring to these approaches, the method-oriented and problem-oriented approaches, respectively (cf. Nirenburg and Raskin 1996, 2001). According to the first of these, a researcher realizes that his or her favorite method, developed for solving a certain problem, is portable to a different field, and he/she rushes into the field with a “ready-made” solution. There is no particular concern for the necessity of such an application and, at least initially, not much understanding of the actual needs of the field. Both SSTH and the first effort in computational humor (Raskin and Attardo 1994), were clearly method-oriented: the former being an offshoot of a formal semantic theory in linguistics and the latter an extension of that theory and of the computational semantic experience to humor research. LIBJOG, a pretty valueless toy system, or rather family of systems, for generating bad light-bulb jokes, spawned a number of more elaborate toy systems (see, for instance, Binsted and Ritchie 1997), varying only in the authors’ optimism about their results and volume of claims about them. Raskin (1996) was, in fact, an attempt to apologize for introducing something that could only provide entertainment and smugness for the systems’ authors rather than anything useful for either of the two fields, NLP and humor research. Stock (1996), however, smartly and forcefully saved computational humor and established it as a potentially serious field by outlining a number of edutainment and, later (Stock and Strapparava 2002), other useful applications, whose list is still growing, thus putting the field on a firm and respectable problem-oriented foundation.

1.2 STRONG AND WEAK AI HYPOTHESES

The reasons many scholars feel justified in using the method-oriented approach are not always as frivolous as they were made to sound in the previous section. There is a constant built-in justification for that kind of thing in all cognitive areas: one can always claim that this is yet another approach to understand how the human mind works. This is the standard theoretical underpinning of all artificial intelligence (AI) and computational-system work: if we design a system that emulates successfully a human intellectual activity, we may then hope to catch a glimpse into the workings of the mind.

A formal or computational theory may or may not make a claim that it is a model of a natural process. The most well-known claim of this sort is the “strong AI hypothesis” which sees AI “as relevant to psychology, insofar as [it takes] a computational approach to psychological phenomena. The essence of the computational viewpoint is that at least some, and perhaps all, aspects of the mind can be fruitfully described for theoretical purposes by using computational concepts” (Boden 1981: 71-72). Under the “strong AI thesis” (the formulation by the philosopher John Searle 1980: 353; see also Searle 1982a; cf. Searle 1997), “the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states,” a claim that Searle ascribes to Turing (1950) and that forms the basis of the Turing Test. We agree with Moody (1993: 79), that “[i]t is an open question whether strong AI really does represent a commitment of most or many researchers in AI.” So instead of modeling the mind itself, under the “weak AI thesis” “the study of the mind can be advanced by developing and studying computer models of various mental processes” (Moody, 1993: 79-80).

It is very easy to slip from the weak to the strong thesis. Even before it started happening in AI on a mass scale during its 1980s triumphant march to the current oblivion, Chomsky had proceeded in the 1960-70s from a moderate claim that the native speaker must have internalized in his or her mind some such rules as those in transformational generative grammar to the extreme claim that the native speaker must have internalized that particular grammar, as defined by Chomsky at the moment (see Raskin 1976 for a detailed critique of this dangerous slippage and for the pertinent references).

The weak thesis must then be the theoretical justification of computational humor: the systems must help us gain insights into the nature of humor as it is handled by the mind, but do they? The description of LIBJOG in Attardo and Raskin (1994) and some discussion there, as well as Raskin (1996), make it abundantly clear that the system is devoid of any intelligence in the computational sense and is operated on manually compiled lists of targeted groups and their purportedly matching laughable traits, thus providing no further understanding of the mechanisms underlying the jokes than the obvious intuitive hypothesis behind the lists. The same can be repeated of JAPE, the Binsted and Richie toy system for generating cross jokes. The experiment showing that human subjects cannot differentiate between the awful jokes generated by humans and the awful jokes generated by the system proves exactly nothing: bad jokes are bad jokes because they are based on a superficial formula rather than on deep meaning and contextual subtlety. One of the reasons the toy systems are unable to provide insights into the humorous mechanisms inside the mind is that they do not sit on top of a full-meaning NLP system while humor in the mind most definitely does (and some bad jokes do not have to).

1.3 APPLICATIONS OF COMPUTATIONAL HUMOR

Short of theoretical justification, computational humor must then be justified in terms of its applicability for practical purposes. The mature Stock hypothesis is well expressed in the advertisement for this Workshop on its Web site. The following applications are listed there:

- business world applications (such as advertis[ing], e-commerce, etc...);
- general computer-mediated communication and human-computer interaction;
- increase the friendliness of natural language interfaces;
- edutainment and autonomous agents systems.

These are pretty inclusive, and there is no need to get into particulars here, especially since some of those are proprietary and competitive: in fact, all of us in the business should be encouraged to go out and compete against each other, inventing and selling new applications in order to enrich ourselves and the field. I would like to add two more types of applications to the list and comment briefly on them:

- customer acceptance enhancement; and
- humor detection.

The first class of applications encourages customers to accept an unpopular product by rewarding them with humor, in any form of delivery, we are discussing at this Workshop. Information security comes to mind as a prime candidate domain for this endeavor. We have moved NLP forcefully into the business of information assurance and security (IAS) with our work in natural language watermarking and tamper-proofing as well as downgrading (see Raskin 1999, Raskin *et al.* 2001, 2002; Atallah *et al.* 2001). In IAS, the main problem is not the scarcity of systems protecting computer networks but rather the resistance, amounting to deception and sabotage, on the part of the system administrators, to acquiring the necessary knowledge and to making the necessary efforts to install and maintain those IAS systems. The humor interface and/or other humor deliverables may reward the complying sysadmins and cajole them into accepting the IAS protection measures.

The second class of applications is even more interesting, in a way, because it involves analysis rather than generation. Most computational humor applications, like the initial toy systems, involved automatic generation of jokes. In humor detection, an application will search for humor, for instance, with the purpose of cleaning the text up, excluding the possibility of any humorous perversion of the text, for instance, in a presidential address, diplomatic note, or any other deadly serious business. As politicians well know, there is hardly anything that is humor-proof but many speeches open themselves widely to abuse by satirists. On the other hand, the same humor-detection applications can be used to determine the vulnerable spots in a text to be denigrated, e.g., in a political campaign, and then to work in conjunction with humor generation to create appropriate and effective humor.

2 ONTOLOGICAL SEMANTICS AT THE SERVICE OF COMPUTATIONAL HUMOR

2.1 GENERAL OVERVIEW OF ONTOLOGICAL SEMANTICS

Ontological semantic applications include machine translation (MT), information extraction (IE), question answering (QA), general human-computer dialog systems, text summarization and specialized applications combining some or all of the above with additional functionality (e.g., advice giving systems). Of course, such applications are attempted without ontological semantics, or, for that matter, without any treatment of meaning at all. If, however, these applications are based on ontological semantics, then any kind of input to the system (an input text for MT, a query for a question answering system, a text stream for information extraction, etc.) first undergoes several stages of analysis (tokenization, morphological, syntactic, semantic, etc.—see Chapter 8 below for details) that, in the case of success, in the end generate the meaning of a text, “text meaning representation” or TMR. The TMR serves as input to specialized processing relevant for a particular application. For example, in MT, the TMR needs to be translated into a natural language different from the one in which the input was supplied. The program that carries this task out is usually called text generator. In IE, TMRs are used by the special rules as sources of fillers of IE template slots. In question answering, the TMR presents the proximate meaning of the user’s query. The QA processor must first understand exactly what the user wants the system to do, then find the necessary information either in the background world knowledge sources (most often, Fact DB, but sometimes the ontology or the lexicons) and then generate a well-formed answer. Needless to say, all of those applications combine and recombine the same modules, with some variation, and all of those, including whole applications, can be used for various computational humor purposes.

The static knowledge sources include the language-dependent ones—the rules for text tokenization, detecting proper names and acronyms and other preprocessing, “ecological” tasks for morphological, syntactic and ontological semantic analysis. The information for the latter three types of analysis resides largely in the lexicons of the system, though special rules (e.g., syntactic grammars) are separate from lexicons. In the current state of ontological semantics, onomasticons, repositories of proper names, are separated from regular lexicons. The language independent static knowledge sources are the ontology and the fact database (Fact DB). The ontology contains information about how things can be in the world while the Fact DB contains actual facts, that is, events that took place or objects that existed, exist or have been reported to exist. In other words, the ontology contains concept types, whereas the Fact DB contains remembered concept tokens, or instances. Onomasticons contain information about words and phrases in natural language that name remembered concept instances. These concept instance names are also recorded as property fillers in Fact DB frames. Note that the Fact DB also contains other, unnamed, concept instances. More detailed descriptions of all the static knowledge sources are given in Nirenburg and Raskin (2002), Chapter 7.

In most applications of ontological semantics, a side effect of the system's operation is selective augmentation of the Fact DB with the elements of TMRs produced during input analysis stage. This way, this information is remembered and made available for future use. It is in this sense that we can say that ontological semantic applications involve learning: the more they operate, the more world knowledge they record, the better quality results they may expect.

2.2 MEANING REPRESENTATION IN ONTOLOGICAL SEMANTICS

2.2.1 Meaning Proper and the Rest

Consider the following text as input to an ontological-semantic processor.

- (1) Dresser Industries said it expects that major capital expenditure for expansion of U.S. manufacturing capacity will reduce imports from Japan.

In “Computerese,” that is, in the form that we expect that a semantic analyzer would be able to process and represent the above text, the latter will be glossed, for example, as follows:

- (2) A spokesperson for the company called Dresser Industries made this statement: Dresser Industries expects that imports into the US from Japan will decrease through large capital investment for the purpose of expanding the manufacturing potential in the US; the expenditure precedes expansion, which precedes reduction, and all of them take place after the statement.

In a somewhat more formal fashion, the meaning of (1) glossed in (2) can be seen to include the following meaning components:

- (3)
 - (i) that Dresser Industries is a phrase, moreover, a set phrase, a proper name;
 - (ii) that it is the name of a company;
 - (iii) that this name is used in the original text metonymically—the company name, in fact, stands for its unnamed spokesperson(s);
 - (iv) that the spokesperson made a statement (that is, not a question or a command);
 - (v) that the company (once again, metonymically) has a certain belief, namely, an expectation;
 - (vi) that the scope of the expectation is the reduction of imports into US from Japan;
 - (vii) that the reduction of imports is expected to take place through capital investment;
 - (viii) that the purpose of the investment is to increase the capacity for manufacturing in the United States;
 - (ix) that United States refers to a nation, the United States of America, and Japan refers to another nation, Japan;
 - (x) that the object of manufacturing, that is left unnamed in the original text is most likely to refer to goods;
 - (xi) that the decrease occurs in the amount of goods that the United States imports from Japan;
 - (xii) that the time at which reduction of imports occurs follows the time of investment which, in turn, preceded the expansion of manufacturing capacity;
 - (xiii) that the time at which the statement was made precedes the time of investment;
 - (xiv) that what is expanded is not necessarily the actual manufacturing output but the potential for it.

The set of expressions in (3) can be viewed as the meaning of (1). In fact, this is the level at which text meaning is defined in the current implementation of ontological semantics. However, it is important to understand that there may be alternative formulations of what constitutes the meaning of (1) or, for that matter, of any text, and it is clear that humorous interpretations will have their own needs. So, it seems appropriate at this point to discuss the general issue of how exactly to define text meaning. It might come as a surprise that this is not such an easy question! One attempt at making the idea of meaning better defined is the introduction of the notion of literal meaning (cf., e.g., Hausser 1999:20). Thus, we could have declared that what we represent in our approach is the literal meaning of texts. However, this decision meets with difficulties because the notion of literal meaning may not be defined sufficiently precisely. For instance, (3) can be construed as the literal meaning of (1). However, under a different interpretation, deciding to resolve the organization-for-employee metonymy in (3.iii) and (3.v) may be construed as going beyond literal meaning. (3) can be seen as the literal meaning of (1) if one agrees that Dresser Industries, being a company, cannot actually be the agent of saying. If this constraint is lifted, by allowing organizations to be agents of speech acts, then the literal meaning will not require the resolution of metonymy. In other words, this kind of literal meaning will be represented by eliminating (3.iii) and (3.v) from (3). In fact, if this approach is adopted throughout, the concept of metonymy will be summarily dispensed with (Mahesh 1996; Section 8.4.2). As the concept of literal meaning can be understood in a variety of ways, we found it unhelpful for defining which kinds of information belong in text meaning and which remain outside it, while still possibly playing a role (of background knowledge used for inference making in reasoning applications) in text processing in a variety of applications. Computational humor requires even more attention to processing non-literal meaning than the state of the art provides at this time (cf. Nirenburg and Raskin 2002, Chapter 8).

We have just considered a possibility of representing the meaning of (1) using less information than shown in (3). It is equally possible to view an expanded version of (3) as the meaning of (1). One example of such expansion would add statements in (4) to the list (3):

- (4) (i) that the company Dresser Industries exists;
- (ii) that Dresser Industries has an opinion on the subject of reducing imports from Japan;
- (iii) that the most probable source of investment that would lead to the expansion of the US manufacturing capacity is either Dresser Industries itself or a joint venture of which it is a part;
- (iv) that the goal of reducing imports is a desirable one.

(4.i) is known as a(n existential) presupposition for (1). (4.ii) is an entailment of (1). Should they be considered integral parts of the meaning of (1)? Information in (4.iii) and (4.iv) is inferred from (1) on the basis of general knowledge about the world. For example, (4.iii) relies on the belief that if it is not stated otherwise, it is strongly probable that Dresser Industries also plans to participate in the expansion of the US manufacturing capacity. It is noteworthy that, unlike for (4.i) and (4.ii), (4.iii) and (4.iv) are not expected to be always true.

Let us explore a little further what this actually means. One way of approaching the task of determining the exact meaning of a text is by using the negation test, a typical linguistic tool for justifying an element of description by showing that its exclusion leads to some sort of deviance, for instance, a contradiction (see, e.g., Raskin 1985). Indeed, the negation of any element of (3) contradicts some component of the meaning of (1). We may take this as an indication that each element of (3) is a necessary part of the meaning of (1). But is it correct to say that any statement whose negation contradicts (1) is a necessary part of the meaning of (1)? Let us consider a few more cases.

It is easy to see why are (5.1) and (5.2) are contradictory. Each of them consists of (1) and the negation of one of the component clauses of (1). Obviously, the contradiction results from the fact that the negated component is an integral part of the meaning of (1).

- (5) (i) Dresser Industries **said** it expects that major capital expenditure for expansion of US manufacturing capacity will reduce imports from Japan, **and** Dresser Industries **did not say** it expects that major capital expenditure for expansion of US manufacturing capacity will reduce imports from Japan
- (ii) Dresser Industries said it **expects** that major capital expenditure for expansion of US manufacturing capacity will reduce imports from Japan, **and** Dresser Industries said it **does not expect** that major capital expenditure for expansion of US manufacturing capacity will reduce imports from Japan

Similarly, contradictory statements will result from adding the negations of (4.i) and (4.ii) to (1), to yield (6.i) and (6.ii):

- (6) (i) Dresser Industries said it expects that major capital expenditure for expansion of U.S. manufacturing capacity will reduce imports from Japan, and Dresser Industries does not exist;
- (ii) Dresser Industries said it expects that major capital expenditure for expansion of U.S. manufacturing capacity will reduce imports from Japan, and Dresser Industries has no opinion on whether major capital expenditure for expansion of U.S. manufacturing capacity will reduce imports from Japan

The source of contradictions in (6) is different, however, than the source of contradictions in (5). The statements added in (6) do not negate anything directly stated in (1). They negate a presupposition and an entailment of (1), respectively: if it is not presupposed that Dresser Industries exists, (1) makes no sense; if it does not follow from (1) that Dresser Industries has an opinion on the subject of imports from Japan, (1) does not make sense, either. As we can see, the negation tool fails to distinguish between the actual elements of the meaning of (1), on the one hand and the presuppositions and entailments of (1), on the other. This outcome gives us two alternatives—either to include presuppositions and entailments in the meaning of (1) (or, by extension, of any statement) or to ignore the results of the negation test in this case.

This distinction turns out to be problematic for people as well. Thus, delayed recall experiments (Chafe 1977) show something that trial lawyers have always known about witness testimony, namely, that people never recall exactly what was said—only the gist of it—and that they routinely confuse the presuppositions and entailments of a statement with what the statements actually assert. The distinction may, however, be quite important in those NLP applications where it is important to distinguish between what is conveyed by the text directly and what is present only by implication, and computational humor is prominently among those applications. For example, at the text generation step of machine translation, what must be translated is the actually made statements and not what they presuppose or entail, the reason being the assumption that the readers will be able to recreate all the implications that were present but not overtly stated in the original text.

The negation tool does, however, work well for (4.iii) and (4.iv). Adding their negations to (1) yields (7.i) and (7.ii) that are somewhat odd but not contradictory:

- (7) (i) Dresser Industries said it expects that major capital expenditure for expansion of U.S. manufacturing capacity will reduce imports from Japan, and it is not the case that Dresser Industries or a joint venture of which it is a part are the most probable source of investment in the US manufacturing capacity.
- (ii) Dresser Industries said it expects that major capital expenditure for expansion of U.S. manufacturing capacity will reduce imports from Japan, and the goal of reducing imports is not a desirable one.

We conclude that the reason for the absence of contradictions in (7) is that (4.iii) and (4.iv) do not negate any elements of the meaning of (1). In general, we assume that if adding the negation of a statement to another statement is not contradictory, then the former statement does not constitute a part of the meaning of the latter statement. One can also say then that there are no contradictions in (7) because (4.iii) and (4.iv) are possible but not necessary entailments from (1).

Many more such possible statements can be inferred from (1) based on the ontological entry for CORPORATION and related concepts and scripts, such as DO-PUBLICITY, for instance:

- (8) (i) that Dresser Industries has headquarters;
- (ii) that it has employees;
- (iii) that it manufactures particular products and/or particular services;
- (iv) that the addressee of the statement by the spokesperson of Dresser Industries was the general public;
- (v) that the statement has been, most probably, made through the mass media, etc.

Even more inferences can be made from (1) based on the general understanding of goals that organizations and people typically pursue as well as plans that they use to attain those goals:

- (9) (i) that there is a benefit for Dresser Industries in expanding the US manufacturing capacity
- (ii) that capital investment is a plan toward attaining the goal of expanding manufacturing capacity;
- (iii) that this goal can play the role of a step in a plan of attaining the goal of reducing imports; or
- (iv) that Dresser Industries knows about suing mass media as a plan for attaining a variety of goals.

All the inferences in (7 - 9) are not “legal” (cf. Charniak and McDermott 1985:21) deductions but rather abductive, defeasible, negatable inferences. It is for this reason that none of them are included in the specification of the meaning of (1). The distinction between meaning proper, on the one hand, and presuppositions, entailments and inferences, on the other, may not be as important for NLP applications whose results are not intended for direct human consumption, e.g., for text data mining aiming at automatic population of databases. People, however, are capable of generating presuppositions, entailments and inferences on the fly from a brief message. Indeed, brevity is at a premium in human professional and business communication. Text meaning or even condensed text meaning are, thus, the central objects of manipulation in such common applications as machine translation and text summarization, respectively. Clearly, presupposition, entailment, and inferences have a greatly enhanced role to play in humorous texts.

For computers, brevity of the kind to which we are referring has little real physical sense in these days of inexpensive storage devices and fast indexing and search algorithms. What is difficult for computer systems is precisely making reliable and relevant inferences. Therefore, spelling out as many inferences as possible from a text and recording them explicitly in a well-indexed manner for future retrieval is essential for supporting a variety of computational applications, especially for computational humor.

It is important for a computational semantic theory to provide the means of supporting both these precepts—of brevity and of explicitness. A representation of text meaning should be as brief as possible, if it is to be the source for generating a text for human consumption. The knowledge about both the building blocks of the meaning representation and the types of inferences that are possible from a particular text meaning should be stored in an accessible fashion. These kinds of knowledge are interchangeable with the change of inputs—what was a part of text meaning for one source text may end up being a source of inference for another. Any computational semantic application must support this capability of dynamically assigning some of the resident knowledge to direct meaning representations and reserving the rest for possible inferences. In ontological semantics, these goals are achieved through interrelationship among text meaning representations (TMRs), the lexicons and the ontology.

2.2.2 TMR IN ONTOLOGICAL SEMANTICS

Meaning of natural language texts is represented in ontological semantics as a result of a compositional process that relies on the meanings of words, of bound morphemes, of syntactic structures and of word, phrase and clause order in the input text. The meanings of words reside in the lexicon and the onomasticon (the lexicon of names). The bound morphemes (e.g., markers of Plural for nouns) are processed during morphological analysis and get their meanings recorded in special rules, possibly, added to classes of lexical entries. Information about dependency among lexical elements and phrases, derived in syntax, helps to establish relationships of semantic dependency. Word and phrase order in some languages play a similar role.

It is clear then that the knowledge necessary for ontological semantic analysis of text should include not only the lexical material for the language of the text but also the results of the morphological and syntactic analysis of the input text. Let us follow the process of creating an ontological-semantic TMR using the example in (1), repeated here as (10).

- (10) Dresser Industries said it expects that major capital expenditure for expansion of U.S. manufacturing capacity will reduce imports from Japan.

English is a morphologically impoverished language, but morphological analysis of (10) will still yield some non-trivial results:

Root	Part of Speech	Features
Dresser Industries	Phrase Proper	Number: Singular
say	Verb	Tense: Past
it	Pronoun	Number: Singular; Person: Third
expect	Verb	Tense: Present; Number: Singular; Person: Third
that	Binder	
major	Adjective	
capital	Noun	Number: Singular
expenditure	Noun	Number: Singular
for	Preposition	
expansion	Noun	Number: Singular
of	Preposition	
U.S.	Acronym	Number: Singular
manufacturing	Verb	Form: Gerund
capacity	Noun	Number: Singular
reduce	Verb	Tense: Future (will marks this in the text)
import	Noun	Number: Plural
from	Preposition	
Japan	Noun Proper	

Results of syntactic analysis of (10) can be represented in the following structure (which is modeled on the f-structure of LFG (e.g., Levin 1991):

(11)

```
root    say
cat     verb
tense   past
subject

      root dresser industries
      cat  phrase-proper

comp

      root expect
      cat  verb
      tense present
      subject
          root dresser industries
          cat  phrase-proper
      object
          root reduce
          cat  verb
          tense future
          subject
              root expenditure
              cat  noun
              modifier capital
                  cat  noun
                  modifier major
                      cat  adjective
      oblique root for
              cat  preposition
              object root expansion
                  cat  noun
                  oblique root of
                      cat  preposition
                      object root capacity
                          cat  noun
                          modifier root manufacturing
                              cat  verb
                              modifier root u.s.
                                  cat  phrase-proper

      object root imports
              cat  noun
              oblique root from
                  cat  preposition
                  object root japan
                      cat  noun-proper
```

We will now use the results of the morphological and syntactic analysis presented above in building a TMR for (10). TMRs are written in a formal language with its own syntax specified in Nirenburg and Raskin (2002), Section 6.4. For pedagogical reasons, at many points in our presentation here, we will use a somewhat simplified version of that language and will build the TMR for (10) step by step, not necessarily in the order that any actual analyzer will follow.

The first step in ontological semantic analysis is finding meanings for heads of clauses in the syntactic representation of input. In our example, these are say, expect and reduce. As we will see, they all will be treated differently in TMR construction. In addition, the TMR will end up containing more event instances than there are verbs in the original text. This is because ontological semantics is “transcategorical” in that meanings are not conditioned by part of speech tags.

Specifically, in (1), the nouns *expenditure* and *expansion* occupying the syntactic positions corresponding typically to heads of noun phrases, are mapped into instances of event-type concepts in the TMR.

In (12), we present the syntactic-structure (SYN-STRUC) and semantic-structure (SEM-STRUC) components of the entry for *say* in the ontological semantic lexicon of English. The meaning of *say* instantiates the ontological concept INFORM. The representation of this concept, shown in (13), contains a number of properties (“slots”), with a specification of what type of object can be a legal value (“filler”) for each property.

(12)

say-v1

syn-struct

1	root	say		; as in <i>Spencer said a word</i>
	cat	v		
	subj	root	\$var1	
		cat	n	
	obj	root	\$var2	
		cat	n	
2	root	say		; as in <i>Spencer said that it rained</i>
	cat	v		
	subj	root	\$var1	
		cat	n	
	comp	root	\$var2	

sem-struct

1 2	inform			; both syntactic structures have the same
				; semantic structure,
	agent value	^\$var1		; ‘^’ is read as ‘the meaning of,’ and
	theme value	^\$var2		; the variables provide mappings between
				; syntactic and semantic structures

(13)

inform

definition	“the event of asserting something to provide information to another person or set of persons”
is-a	assertive-act
agent	human
theme	event
instrument	communication-device
beneficiary	human

So far, then, the nascent TMR for (1) has the form:

(14)

inform-1

agent	value	_____
theme	value	_____

The arbitrary but unique numbers appended to the names of concepts during ontological semantic processing identify instances of concepts. The numbers themselves are also used for establishing co-reference relations among the same instances. At the next step of semantic analysis, the process seeks to establish whether fillers are available in the input for these properties. If the fillers are not available directly, there are special procedures to try to establish them. If these recovery procedures fail to identify the filler but it is known that some filler must exist in principle, the special filler UNKNOWN is used.

The AGENT slot in (14) cannot be filled directly from the text. The reason for that is as follows. The procedure for determining the filler attempts to use the syntax-to-semantics mapping in the lexicon entry for *say*, to establish the filler for the particular slots. The lexicon entry for *say* essentially states that the meaning, $\wedge \$var1$, of the syntactic subject of *say*, $\$var1$, should be the filler of the AGENT slot of INFORM. Before inserting a filler, the system checks whether it matches the ontological constraint for AGENT of INFORM and discovers that the match occurs on the RELAXABLE-TO facet of the AGENT slot, because *Dresser Industries* is an ORGANIZATION. Note that the ontological status of DRESSER INDUSTRIES is that of a (named) instance of the concept CORPORATION—see Nirenburg and Raskin (2002), Section 4.2.1 for a discussion of instances and remembered instances.

The TMR at this point looks as illustrated in (15).

(15)

inform-1			
	agent	value	Dresser Industries
	theme	value	_____

The theme slot in (14) requires a more complex treatment. The complement of *say* in the syntactic representation (11) is a statement of expectation. According to a general rule, the direct object of the syntactic clause should be considered as the prime candidate for producing the filler for THEME. Expectation, however, is considered in ontological semantics to be a modality and is, therefore, represented in TMR as a property of the proposition that represents the meaning of the clause that modifies it syntactically. Before assigning properties, such as this modality, we will first finish representing the basic meanings that these properties characterize. Therefore, a different candidate for filling the theme property must be found. The next candidate is the clause headed by *reduce*. Consulting the lexicon and the ontology and using the standard rules of matching selectional restrictions yields (16):

(16)

inform-1			
	agent	value	Dresser Industries
	theme	value	decrease-1
decrease-1			
	agent	value	unknown

Continuing along this path, we fill the case roles THEME and INSTRUMENT in (16), as well as their own properties and the properties of their properties, all the way down, as shown in (17):

(17)

inform-1			
	agent	value	Dresser Industries
	theme	value	decrease-1
decrease-1			
	agent	value	unknown
	theme	value	import-1
	instrument	value	expend-1
import-1			
	agent	value	unknown
	theme	value	unknown
	source	value	Japan
	destination	value	USA
expend-1			
	agent	value	unknown
	theme	value	money-1
			amount value > 0.7
	purpose	value	increase-1

```

increase-1
  agent      value  unknown
  theme      value  manufacture-1.theme
manufacture-1
  agent      value  unknown
  theme      value  unknown
  location   value  USA

```

Some elements of (17) are not self-evident and require an explanation. First, the value of the property AMOUNT of the concept MONEY (which is the meaning of *capital* in the input) is rendered as a region on an abstract scale between 0 and 1, with the value corresponding to the meaning of the word major. The same value would be assigned to other words denoting a large quantity, such as *large*, *great*, *much*, *many*, etc. The meanings of words like *enormous*, *huge* or *gigantic* would be assigned a higher value, say, > 0.9. THEME of INCREASE is constrained to SCALAR-OBJECT-ATTRIBUTE and its ontological descendants, of which AMOUNT is one. The filler of the THEME of increase-1 turns out to be the property AMOUNT itself (not a value of this property!) referenced as the THEME of manufacture-1, rendered in the familiar dot notation.

Now that we have finished building the main “who did what to whom” semantic dependency structure, let us add those features that are, in ontological semantics, factored out into specific parameterized properties, such as speech act, modality, time or co-reference. The top proposition in (18) reflects the speech act information that in the text (1) is not expressed explicitly, namely, the speech act of publishing (1) in whatever medium. The speech act introduces an instance of the ontological concept AUTHOR-EVENT (see also Nirenburg and Raskin 2002, Section 6.5).

(18)

```

author-event-1
  agent value unknown
  theme value inform-1
  time
    time-begin > inform-1.time-end
    time-end   unknown
inform-1
  agent value Dresser Industries
  theme value decrease-1
  time
    time-begin unknown
    time-end   (< decrease-1.time-begin) (< import-1.time-begin) (< reduce-1.time-
begin)
                                (< expend-1 .time-begin) (< increase-1.time-begin)
decrease-1
  agent      value  unknown
  theme      value  import-1
  instrument  value  expend-1
  time
    time-begin (> inform-1.time-end) (> expend-1.time-begin) (> import-1.time-begin)
    time-end   < import-1.begin-time
import-1
  agent      value  unknown
  theme      value  unknown
  source     value  Japan
  destination value  USA
  time
    time-begin (> inform.time-end) (< expend-1.begin-time)
    time-end   unknown

```

```

expend-1
  agent      value      unknown
  theme      value      money-1
              amount value > 0.7
  purpose    value      increase-1
  time
    time-begin > inform.time-end
    time-end   < increase-1.begin-time

increase-1
  agent      value      unknown
  theme      value      manufacture-1.theme
  time
    time-begin (> inform.time-end) (< manufacture-1.begin-time)
    time-end   unknown

manufacture-1
  agent      value      unknown
  themevalue unknown
  location   value      USA
  time
    time-begin > inform.time-end
    time-end   unknown

modality-1
  type      potential
  value     1
  scope     decrease-1
  ;this is the meaning of expects in (10)
  ;this is the maximum value of potential

modality-2
  type      potential
  value     1
  scope     manufacture-1
  ;this is the meaning of capacity in (10)

co-reference-1
  increase-1.agent      manufacture-1.agent

co-reference-2
  import-1.theme       manufacture-1.theme

```

The time property values in each proposition, all relative since there is no absolute reference to time in the input sentence, establish a partial temporal order of the various events in (1): for example, that the time of the statement by Dresser Industries precedes the time of reporting. The expected events may only take place after the statement is made. It is not clear, however, how the time of reporting relates to the times of the expected events because some of them may have already taken place between the time of the statement and the time of reporting.

Inserting the value UNKNOWN into appropriate slots in the TMR actually undersells the system's capabilities. In reality, while the exact filler might not be indeed known, the system knows many constraints on this filler. These constraints come from the ontological specification of the concept in which the property that gets the UNKNOWN filler is defined and, if included in the TMR, turn it into what we define as extended TMR (see Nirenburg and Raskin 2002, Section 6.7). Thus, the AGENT of import-1 is constrained to U.S. import companies. The AGENT of expend-1 is constrained to people and organizations that are investors. The AGENT of increase-1 and manufacture-1 is constrained to manufacturing corporations. The THEME of import-1 and manufacture-1 is constrained to GOODS (the idea being that if you manufacture some goods then you do not have to import them). The facts that Dresser Industries is a company while Japan and USA are countries are stored in the onomasticon.

2.3 ONTOLOGICAL SEMANTIC RESOURCES

The resources of ontological semantics, publicly available, provide a much richer and more accurate meaning representation of natural language text than WordNet. The reason for that is simple: WordNet was never intended for this purpose, and it has been used in NLP primarily because it is there and because it can be extended without any direct semantic analysis, something that most NLP groups fear and try their best to avoid. Ontological semantics was created specifically for the purpose of meaning representation in the process of costly semantic analysis performed by 3 tiers of experts with the help of a sophisticated toolbox of semi-automatic acquisition interfaces, optimally combining the native speakers' semantic competence with automatic procedures of its representation. There is no doubt that the amount of sophistication required for humor analysis can only be met by extending ontological semantic resources.

3 CONCLUSION

It is a good time for computational humor. We are, hopefully, past the stage of useless toy systems. We are much close to implementing a full and comprehensive text meaning representation system based on ontological semantics. We have gained a more realistic view of how much this research can actually tell us about the mind. We have gained maturity to see the important role of applications in supporting this important research. We have no more excuse to waste time on any more "bags of tricks" to get a computational system to do something visible but useless. It is time to get to work.

REFERENCES

- Atallah, M., and V. Raskin 2001. Natural language watermarking: Design, analysis, and a proof-of-concept implementation." In: Ira S. Moskowitz (ed.), **Information Hiding**. Berlin-New York: Springer Verlag, 2001, pp. 185-199.
- Attardo, S. 1994. **Linguistic Theories of Humor**. Berlin: Mouton de Gruyter.
- Attardo, S. 2001. **Humorous Texts: A Semantic and Pragmatic Analysis**. Berlin: Mouton de Gruyter.
- Attardo, S., and V. Raskin 1991. Script Theory Revis(it)ed: Joke Similarity and Joke Representation Model. *Humor* 4:3-4, pp. 293-347.
- Attardo, S., and V. Raskin 2003. **General Theory of Verbal Humor**. Berlin: Mouton de Gruyter. In Preparation.
- Binsted, K., and G. Ritchie 1997. Computational rules for punning riddles. *Humor* 10:1, pp. 25-76.
- Boden, M. A. 1981. **Minds and Mechanisms: Philosophical Psychology and Computational Models**. Ithaca, NY: Cornell University Press.
- Charniak, E., and D. McDermott 1985. **Introduction to Artificial Intelligence**. Reading, MA: Addison-Wesley.
- Ermida, I. 2002. Superscripts and the Humor of Narratives. Unpublished Ph.D. Thesis, Department of English, University of Minho, Minho, Portugal.
- Hausser, R. 1999. **Foundations of Computational Linguistics**. Berlin: Springer.
- Levin, L. S. 1991. Syntactic Theory and Processing. In: K. Goodman and S. Nirenburg (eds.), KBMT-89: A Case Study in Knowledge-Based Machine Translation, San Mateo: Morgan Kaufmann, pp. 49-70.
- Mahesh, K. 1996. Ontology Development for Machine Translation: Ideology and Methodology. Memoranda in Computer and Cognitive Science MCCS-96-292, New Mexico State University: Computing Research Laboratory.

- Moody, T. C. 1993. **Philosophy and Artificial Intelligence**. Englewood Cliffs, NJ: Prentice-Hall.
- Nirenburg, S., and V. Raskin 1996. Ten Choices for Lexical Semantics. Memoranda in Computer and Cognitive Science MCCA-96-304, New Mexico State University: Computing Research Laboratory.
- Nirenburg, S., and V. Raskin 2001. Choices in Lexical Semantic Theory. *Computational Intelligence* 17:1, pp. 157-177.
- Nirenburg, S., and V. Raskin 2002. **Ontological Semantics**. Cambridge, MA: MIT Press. Forthcoming. Prefinal Draft at: <http://crl.nmsu.edu/Staff.pages/Technical/sergei/book/index-book.html>.
- Raskin, V. 1976. Generation and performance. *Theoretical Linguistics* 181, pp. 45-61.
- Raskin, V. 1985. **Semantic Mechanisms of Humor**. Dordrecht-Boston-Lancaster: Reidel.
- Raskin, V. 1996. Computer implementation of the general theory of verbal humor. In: J. Hulstijn and A. Nijholt (eds.), **Automatic Interpretation and Generation of Verbal Humor. International Workshop on Computational Humor, IWCH '96. Twente Workshop on Language Technology, TWLT 12**, Enschede, NL: University of Twente, pp. 9-19.
- Raskin, V. 1999. Natural language processing and information assurance and security. Paper delivered at ITC-irst, Trento, Italy.
- Raskin, V., M. J. Atallah, C. J. McDonough, and S. Nirenburg 2001. Natural language processing for information assurance and security: An overview and implementations. In: M. Schaefer (ed.), **Proceedings of NSPW-2000**, New York: ACM Press, pp. 51-65.
- Raskin, V., and S. Attardo 1994. Non-literalness and non-bona-fide in Language: An approach to formal and computational treatments of humor. *Pragmatics and Cognition* 2:1, pp. 31-69.
- Raskin, V., C. F. Hempelmann, K. E. Triezenberg, and S. Nirenburg 2002. Ontology in information security: A useful theoretical foundation and methodological tool." In: V. Raskin and C. F. Hempelmann (eds.), **Proceedings: New Security Paradigms Workshop 2001**, New York: ACM Press, 2002, pp. 67-73.
- Searle, J. R. 1980. Minds, brains, and programs. *Behavioral and Brain Sciences* 3, pp. 417-457. Reprinted in: D. R. Hofstadter and D. C. Dennett (eds.), *The Mind's I*, New York: Bantam, 1981.
- Searle, J. R. 1982. The myth of the computer. *The New York Review of Books*, April 29, 3-6.
- Searle J. R. 1997. **The Mystery of Consciousness. Including exchanges with Daniel C. Dennett and David J. Chalmers**. New York: A New York Review Book. .
- Stock, O. 1996 'Password Swordfish': Verbal humor in the interface. In: J. Hulstijn and A. Nijholt (eds.), **Automatic Interpretation and Generation of Verbal Humor. International Workshop on Computational Humor, IWCH '96. Twente Workshop on Language Technology, TWLT 12**, Enschede, NL: University of Twente, pp. 1-8.
- Stock, O., and C. Strapparava 2002. HAHAcronym: Humorous agents for humorous acronyms. In this volume.
- Turing, A. 1950. Computing machinery and intelligence. *Mind* 59, 434-460. Reprinted in: A. R. Anderson (ed.), **Minds and Machines**, Englewood Cliffs, NJ: Prentice-Hall, 1964, 4-30.

The Structure of Forced Reinterpretation Jokes *

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

Abstract

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

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

1 AIMS

We take “computational humour” to refer to:

- (a) the design, implementation and evaluation of programs which are capable of producing humorous artefacts or behaviours;
- (b) the design, implementation and evaluation of programs which are capable of responding appropriately to humorous artefacts or behaviours;
- (c) the use of computational concepts for studying and theorising about humour.

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

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

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

2 THE CLASS OF JOKES

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

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

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

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

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

The properties and relationships relevant to humour would then be:

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

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

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

3 EXAMINING THE BASIC CONCEPTS

3.1 MEANINGS AND INTERPRETATIONS

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

- (2) ‘Is the doctor at home?’ the patient asked in his bronchial whisper. ‘No,’ the doctor’s young and pretty wife whispered in reply. ‘Come right in.’ (Raskin, 1985)

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

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

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

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

3.2 OBVIOUSNESS

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

3.3 CONFLICT

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

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

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

3.4 COMPATIBILITY

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

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

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

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

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

3.5 ANOTHER VERSION OF CONFLICT AND COMPATIBILITY

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

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

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

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

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

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

4 A FORMALISATION

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

4.1 RELATIVITY

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

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

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

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

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

4.2 INTERPRETING TEXTS

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

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

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

- (a) $S_0 = PB$
- (b) for $1 \leq i \leq n$, there is an element (FC_i^j, CM_i^j) of $M(T_i, S_{i-1})$ such that S_{i-1} can accommodate FC_i^j .
- (c) for $1 \leq i \leq n$, S_i is an element of $I((S_{i-1} \odot FC_i^j) \oplus CM_i^j)$.

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

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

- (a) $S_0 = PB$
- (b) for $1 \leq i \leq n$, S_{i-1} can accommodate FC_i^{max} where (FC_i^{max}, CM_i^{max}) is the maximum element of $M(T_i, S_{i-1})$.
- (c) for $1 \leq i \leq n$, S_i is the maximum element of $I((S_{i-1} \odot FC_i^{max}) \oplus CM_i^{max})$.

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

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

4.3 FORCED REINTERPRETATION JOKES

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

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

- (a) S_{n-1} is the most obvious interpretation of $T_1 \dots T_{n-1}$ based on a sequence S_1, \dots, S_{n-1}
- (b) S_{n-1} cannot accommodate FC_n^{max} where (FC_n^{max}, CM_n^{max}) is the maximum element of $M(T_n, S_{n-1})$.

Definition: A text $T_1 \dots T_n$ has *punchline resolution* with interpretation S'_n if:

- (a) There is a less obvious interpretation S'_{n-1} of $T_1 \dots T_{n-1}$, based on a sequence S'_1, \dots, S'_{n-1} ;
- (b) S'_{n-1} can accommodate FC_n^{max} , where (FC_n^{max}, CM_n^{max}) is the maximum element of $M(T_n, S'_{n-1})$;
- (c) $S'_n = I((S'_{n-1} \odot FC_n^{max}) \oplus CM_n^{max})$.

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

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

To cover jokes such as (3), we need some further definitions. The idea will be to define a situation where information extracted from the punchline (initially interpreted in the context of the set-up) would, if inserted earlier in the text, have led to a different interpretation of the text: for example, if (3) had included the sentence *Charlie owed John two hundred dollars* at an early stage of the narrative. This involves having an interpretation sequence S'_1, \dots, S'_{n-1} which initially is the same as the original most obvious interpretation sequence, but has the punchline information added at some intermediate point, with subsequent stages then making inferences from that enhanced base of propositions.

Definition: Let $T_1 \dots T_n$ be a text. Suppose $T_1 \dots T_{n-1}$ has a most obvious interpretation S_{n-1} based on a sequence S_1, \dots, S_{n-1} . Suppose that the maximum element of $M(T_n, S_{n-1})$ is (FC_n, CM_n) and CM_n (the core meaning part) contains a set of propositions P such that the following holds. There is a sequence of sets of propositions S'_1, \dots, S'_{n-1} such that

- (a) $S'_{n-1} \neq S_{n-1}$
- (b) for some $k, 1 \leq k \leq (n-1)$:
 - (i) for $1 \leq i < k, S_i = S'_i$;
 - (ii) S'_k is the maximum element of $I(((S'_{k-1} \odot FC_k^{max}) \oplus P) \oplus CM_k^{max})$ where (FC_k^{max}, CM_k^{max}) is the maximum element of $M(T_k, S'_{k-1})$;
 - (iii) for $k < i \leq n-1, S'_i$ is the maximum element of $I((S'_{i-1} \odot FC_i^{max}) \oplus CM_i^{max})$, where (FC_i^{max}, CM_i^{max}) is the maximum element of $M(T_i, S'_{i-1})$.

Then $T_1 \dots T_n$ has *punchline revision* with interpretation S'_{n-1} .

4.4 THE CRITICAL INGREDIENT

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

Conjecture: In a text which constitutes a discourse-coherence misunderstanding or has punchline revision, if the resulting interpretation is ABSURD or TABOO, the text constitutes a joke.

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

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

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

5 ANOTHER JOKE CLASS: DRAWING OUT IMPLICATIONS

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

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

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

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

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

- (a) the text $T_1 \dots T_{n-1}$ has a most obvious interpretation S_{n-1} ;
- (b) the maximum element of $I((S_{n-1} \odot FC_n^{max}) \oplus CM_n^{max})$ contains a set A of propositions which
 - are not a subset of CM_n^{max} (i.e. they must be inferred);
 - are not a subset of S_{n-1} (i.e. prior to the punchline, they were not OBVIOUS).

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

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

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

6 DISCUSSION

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

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

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

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

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

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

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

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

REFERENCES

- Charniak, E. and McDermott, D. (1985). *Introduction to Artificial Intelligence*. Addison-Wesley, Reading, Mass.
- Dascal, M. (1985). Language use in jokes and dreams: sociopragmatics vs. psychopragmatics. *Language and Communication*, 5(2):95–106.
- Dolitsky, M. (1992). Aspects of the unsaid in humor. *HUMOR*, 5(1/2):33–43.
- Gärdenfors, P. (1988). *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press, Bradford Books, Cambridge, MA.
- Hobbs, J., Stickel, M., Appelt, D., and Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, 63:69–142.
- Hobbs, J., Stickel, M., Martin, P., and Appelt, D. (1988). Interpretation as abduction. In *Proceedings of 26th Annual Meeting of the ACL*, pages 95–103.
- Jurafsky, D. and Martin, J. (2000). *Speech and Language Processing: an Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice-Hall, New Jersey.
- Norricks, N. R. (2001). On the conversational performance of narrative jokes: toward an account of timing. *HUMOR*, 14(3):255–274.
- Oaks, D. D. (1994). Creating structural ambiguities in humor: getting English grammar to cooperate. *HUMOR*, 7(4):377–401.
- Raskin, V. (1985). *Semantic Mechanisms of Humour*. Reidel, Dordrecht.
- Ritchie, G. (1999). Developing the incongruity-resolution theory. In *Proceedings of the AISB Symposium on Creative Language: Stories and Humour*, pages 78–85, Edinburgh, Scotland.
- Schank, R. and Abelson, R. (1977). *Scripts, plans, goals and understanding*. Lawrence Erlbaum, Hillsdale, N.J.
- van der Sandt, R. (1988). *Context and Presupposition*. Croom Helm, London.

Computers with a Personality? Lessons to be Learned from Studies of the Psychology of Humor.

Willibald Ruch
School of Psychology
Queens University Belfast
Belfast BT7 1NN
N. Ireland, UK
w.ruch@qub.ac.uk

Abstract

The aim of the present paper is to highlight that research on humor remains remarkably incomplete if individual differences are not considered. This also pertains to computational approaches to humor which so far frequently neglected the fact that tastes are different. Irrespective of whether jokes are created by humans or a computer program they will not always find an appreciative audience and it might be fruitful to consider a fit to the targeted recipient. As a step into this direction research from the domain of personality and humor appreciation will be reviewed, and the dimensions associated with whether or not a person will find humor funny will be described. Furthermore, approaches failed to consider the fact that jokes might fail altogether and humans not only respond exclusively positively to humor but the experiential world of the recipient is multidimensional. Therefore, testing the quality of computer-generated humor should go beyond consideration of degree of “funniness”. A few ways will be outlined and some studies exploring the response dimensions will be presented.

Keywords: Humor and Personality, Assessment of Humor, 3WD Humor Test of Personality

1 INTRODUCTION

Imagine the comedian who tells the same type of jokes and funny stories at all performances irrespective of who invites him or her and who the audience is; whether it is on stage in a night club, as a speaker on a companies annual meeting, at his parents wedding anniversary, to his children, neighbors etc. How successful will he be? Won't some audiences find offensive what made others roar? If he does not want to offend anybody—won't the remaining sample of jokes just be plain boring to many? Obviously, such a strategy would not be very successful, and thus a comedian does just what we do, namely to tailor our attempts at humor to our audience. This is done quite intuitively and no systematic study is known on how people actually do so.

If they know the audience well, they might base it on prior experience on what was found funny and what not amusing at all. If not, it is still easy if you have time to “learn” the taste of the audience. One can use quite diverse jokes initially and see which ones they like and which ones they don't. Then one can put this hypothesis to a test and see whether they actually laugh at a joke that one chooses deliberately because it is similar to a prior successful one. Likewise, one can take the prior failure with a bad joke and see whether one is equally unsuccessful with a sister joke. Such a procedure will only work if we have the right taxonomy of jokes; i.e., a rationale that tells which jokes are similar, if not interchangeable (i.e., equal for this given purpose). If the predictions are all correct one can end the learning stage and from now on one has the capacity to optimally serve the audience. But how often do we have a chance for such a trial and error

procedure? An audience may not be that patient and it might be better to start with a clear idea of what will work out and what not.

If one does not have the chance to learn, one still does not need to rely solely on guessing, as we can infer preference from other salient features of the audience. Their current state, for example; are they tired because your routine was set to the end of a tiring working day? Then it might be better not to indulge too much into very sophisticated humor. But not only are the transient emotional states important; more habitual factors, like gender, socio-economic, ethnic background, are crucial. Indeed personality characteristics allow a much better prediction of appreciation of humor that mood does. This has been known for a while; even poets can be quoted on this (e.g., the saying by German philosopher and poet Johann Wolfgang von Goethe "Men show their character in nothing more clearly than by what they think laughable."), and empirical research subsequently provided a scientific foundation for this.

So, what are the chances for the above-mentioned comedian to survive any audience, and even get hired again? How are we perceived in everyday life by the audiences at our attempts to be funny, be it our partner, friend, peers, or strangers? Probably each of us knows somebody who repeatedly tells one type of joke, and we have formed some impression about him or her on this basis. Imagine somebody predominately telling jokes about infidelity, or about homosexuality, more than anything else and on different occasions. What does this tell us about the joke teller? Does he or she presumably have a negative attitude to those topics? That he has a problem with this topic, maybe is in a conflict? Or, that this topic is no problem for the person so that he or she can freely joke about it? Freud (1905) assumed that the super ego does not allow the direct expression of certain impulses and consequently these topics get repressed. However, they may be gratified in a disguised manner as occurs in jokes or dreams. Psychoanalysts therefore considered jokes as a golden path to the conflict areas in the individual and asked them to tell their favorite joke. We "communicate" needs and motivations through what we tell as jokes and through the intensity of our reaction as an audience (Davis & Farina, 1970). Research has shown that Freud was wrong in this respect (see below), but nevertheless, we have to consider that humor induces emotions and audiences may develop ideas about the sender of a humorous communication.

Thus, a more complete view of the humor process should not only involve the analysis of the humorous message but also the states and traits of sender and the receiver, and what they think about each other before and after the message was transmitted. Research has shown that on the side of the sender factors like current mood and motivation, and enduring personality as well as intellectual traits determine whether or not somebody will decide to encode humor, how well he or she is able to do so, what the content or tendency will be, and whether or not considerations about potential effects on the audience exist, etc. (see anthologies by Chapman & Foot, 1977; Goldstein & McGhee, 1972, McGhee & Goldstein, 1983, Ruch, 1998). Likewise, research demonstrated the effects on the side of the receiver. It is evident that degree of amusement depends on a variety of factors, such as his or her traits, the current mood and motivational state of the receiver, physical and social situational factors etc., and it is also evident that the receiver generates a "picture" about features of the sender (i.e., his or her personality, mood or motivations). How much of this reality is considered in research on computational humor?

2 COMPUTATIONAL AND HUMAN HUMOR

When I first got aware of computational approaches to humor at an ISHS conference 10 years ago feelings of awe, curiosity, puzzlement and skepticism alternated. Of course, being able to emulate human behavior is very good proof that one has all the relevant factors under control and understands the processes. So the pursuit of this line of research is a very fruitful one and it will help to test our knowledge and eventually modify theories. It is beyond the scope of this paper to review the approaches and to appraise the state of the art. Rather I would like to address some new frontiers I consider important to consider and attain to in the future. I won't restrict myself to what is feasible right now, so some criteria might seem quite utopistic. And I want to add a few general considerations regarding the narrowness or broadness of the scope of an approach to humor research.

As mentioned above, I found it admirable to have programs generating humor and it is definitely an achievement and a step in the right direction. Obviously, humans undergo an ontogenetic development (McGhee, 1979). The first attempts at humor among children may be moving and highly appreciated by parents, but from a more objective and distant perspective they are not so impressive. Spelling out a taboo word can lead to roaring laughter among youngsters, but children who had already gone through this stage already won't be impressed. Likewise, first attempts at retelling a joke might leave out the essentials in the punch line but it does not matter to peers as sheer incongruity is funny. While children gradually learn what is essential to jokes, there is still room for development as complexity increases and the contents change. Humor of 10 year-olds still does not suffice to amuse older ones, say adolescents or adults. Obviously, programs generating humor develop too but they will have a different development. Earlier versions might occasionally resemble the attempts made by children to generate humor, but not intentionally so. Rather they appear to me as trying to emulate the humor of adults and just sometimes fail to do so. A successful development will then be once the program is capable of producing humor that is indistinguishable from the humor of adults. Before I come back to this point I would like to say that it might be a very difficult task to emulate the developmental stages of children's humor properly.

As a personality psychologist, I am more interested in the "personality" of the program generating humor, the sender, as well as the one of the receiver (if it happens that programs perceiving, understanding, and responding to humor will be created). What personality do they have—or better, what image of a personality do they create? Are they chauvinistic or sexist? Do they move me because of their innocence or naivety? Are they refined in their sense of humor and intellectually entertaining? Do they share any funny idea they have whether or not I might like it, or are they careful not to offend me? Are they cheerful and upbeat, spreading good mood, or are they sarcastic or bitter?

If we expose humans to humor from all existing programs, will they realize that the sources are different? What kind of impressions will they have about the sender (whether or not they are told that the humor is computer-generated). So, does the perceived sense of humor of the personality of the different programs differ? Whether or not they differ—can programs deliberately appear different? Can they emulate different senses of humor? Can they vary their mood and be a grumpy, sarcastic, or cheerful joker on command? Can they deliberately express or suppress libidinal impulses or motivations?

These are studies of cognitions of the receiver about the sender, or require direct judgments about the state and traits or humor style of the sender as done by experts. What about the humorous products? Do they cover the whole span of (joke, pun) humor, or only selected ones? If taxonomies (of jokes, puns etc) exist—does the program emulate all categories and even purposefully so? Would computer and human generated humor end up loading on the same factors (see below)? Or is it the aforementioned comedian who is not able to tailor the jokes to the taste of the audience? Can they please each audience if they are given the prior information that it is either cognitively simple, medium, or complex?

Obviously, we are talking about utopistic goals here. While some goals are not reachable at the moment others might not even be worth pursuing. However, it might be worthwhile to map computer-generated humor in taxonomies of humor created by human beings and see how it stacks off. Also, it might be of interest to study the "personality" of the programs and compare its sense of humor to the ones of humans. And it might also be important to study what cognition a human receiver develops about the sender, when it is exposed to the humor generated. In all cases the criteria for success will be if one deliberately can create the variation that naturally exists in human life.

In the next section I will present one approach to structure the experiential world of the human perceiver of humor. How he or she responds to jokes and cartoons. This will involve the perception of the stimulus properties (i.e., what makes jokes appear similar and others different) as well as the different ways in which people respond to humor. Before doing so I want to briefly address the issue of the different breadth of the scope of the same research field when approached from different disciplines. Although we are all interested in "humor", we have to acknowledge

that for some approaches at least the researcher needs to be explicitly considered and the whole enterprise is best seen as a “humor”-“researcher” dyad. The training and disciplinary background will be different and so some element subjectivity is added. This has to be considered in the evaluation of the outcome of such research. In other approaches the complexity is added by the presence of humans, the sender or receiver of humor, and hence a triad of “humor”-“researcher”-“sender/receiver” needs to be considered. While in the second case the dialog among researchers is important and inter-researcher agreement might rule out subjectivity, in the latter case the subjectivity or individuality is not to be considered “noise” or “error variance” but meaningful variation that is in itself is of interest to study. While one can view computational humor as a dyad of researcher and program/product of program in the early stage it might be reasonable to see it as a three chain enterprise once it is matured. The taste and preference of the researcher, but the layperson needs to be systematically integrated into this process too. While I can see that samples of non-expert research participants are used to evaluate the quality of the humor generated, I would consider this to be a first step only.

3 HUMOR APPRECIATION

Ruch (1981) proposed that a comprehensive assessment of humor should not only cover a taxonomy of humor stimuli but also an investigation of the dimensionality of the responses to humor, as well as a typology of the receiver. The taxonomy of humor stimuli was achieved by a set of factor analytic studies of differing but overlapping sets of jokes and cartoons. In order to get a robust taxonomy samples differing with regard to sex, age, occupation, health status and other variables were used (for an overview see, Ruch, 1992). Most importantly, the first construction samples covered Austrian as well as German subjects. After establishing the taxonomy the items were translated into different languages (e.g., English, French, Hebrew, Russian, Turkish) in order to be able to test its cross-cultural stability (see, Ruch & Hehl, 1998). The validity of the factors was investigated by studying correlations to personality.

Similarly, the dimensions of appreciation were obtained by correlational and factor analytic studies of several rating scales covering different aspects of the responses to humor. Those factors were later validated against responses at other levels, such as facial expression, psycho-physiological responses, or mapping their location in general taxonomies of emotions. The present chapter only highlights some selected outcomes. For the details of studies, and other prior literature and competing findings and models of other researchers the reader is referred to the original studies or earlier reviews by the present of other authors (e.g., Martin, 1998).

3.1 THE EXPERIENTIAL WORLD OF THE RECEIVER

The responses to humor have been conceptualized in quite different ways depending on the research background of the researcher. Jokes are funny, and so the assessment of degree of “funniness” of a joke on a five- or seven-point rating seem to be the most natural criterion to use in experiments. But how does “humorous” or “witty” compare to “funny”? Would not be inclined to reserve those qualifications for different stimuli? Some humor strikes us as droll, clever, or silly? Is “childish” or “silly” just the low end of funniness or is it something different? Obviously many words we use can reflect aspects of amusing or high or poor in quality. But negative responses go beyond “not funny” or “boring”. Jokes can be offensive, embarrassing, or disgusting, to name a few. Thus, one aim would be to map out the dimensions of how we perceive humor.

However, emotion researchers would look less on the perceived stimulus qualities but on the nature and intensity of the emotion induced (to keep matter simple, humor-induced mood changes are neglected, see Ruch & Köhler, 1998). Amusement (hilarity, exhilaration, mirth) would be a natural candidate, but again, humor is capable of inducing more emotions. Amusement is often treated as a facet of happiness, and together with sadness, anger, fear, surprise, contempt and disgust forms the basic emotions distinguished by Ekman (1992). In some of our studies we used Ekman’s basic emotions to cover a broader emotional spectrum and it turned out that subjects made use of the negative emotions quite often. In open ended responses to jokes Ruch and Rath (1993) found that people use other negative emotion words more often. Again, what would be

the dimensions describing the subjective feeling state of individuals exposed to humor? As those different from the dimensions of how we perceive humor? Do more “cognitive” evaluations (clever, stimulated) differ from more affective (funny, amused) ones?

Ruch and Rath (1993) attempted to determine at the experiential level the number and nature of dimensions of response to humor. A sample of 50 male and 50 female adults judged 24 jokes and cartoons on 17 seven-point rating scales. The set of ratings was empirically selected from spontaneous responses of subjects to a set of humor stimuli and represents a variety of aspects of reactions to humor. Positive and negative responses were recorded, as were judgments about perceived stimulus properties and subjects’ own feeling state. Factor analyses of the intercorrelations among the response scales was performed for the joke types separately and combined, and always yielded three factors (see Table 1).

The results displayed in Table 1 suggest that the most distinguishing feature is the hedonic tone of the evaluation; positively and negatively toned ratings were found to load on different factors. Positive responses yielded only one factor (of own positive affect/perceived joke quality), but the negative responses were split up into two correlated dimensions. One of which related more to affective (“offensive”/“indignation”) and the other to cognitive (“simple”/“boredom”) aspects of experience and perception of stimulus properties. All three factors contain evaluations of both stimulus properties and evaluation of the one’s feeling state: thus, this distinction between evaluations of the stimulus and one’s emotional reactions to the stimulus turns out not to be of importance.

While the major dimension blends perceived funniness with the degree of positive emotion induced one should not forget that this is a global dimension only. Variations in attributes (funny, witty, humorous) still exist and might be found of importance for other studies. However, this response dimension also goes along with overt behavior such as smiling and laughter (Ruch, 1995) albeit those converge better intrapsychically than across subjects due to differences in expressivity, and physiological changes (Ruch, 1993). Also the other two-dimensions go along with facial expressions and might reflect different processes (for more details see the study by Ruch & Rath, 1993).

Rating scales ¹	funniness/exhilaration				offensive/indignation				simple/boredom			
	All	INC	NON	SEX	All	INC	NON	SEX	All	INC	NON	SEX
Witty	.98	.80	.66	.92	-.20	-.32	.01	-.03	.01	-.09	-.39	.02
Funny	.94	.85	.66	.87	-.05	-.08	.04	-.01	-.01	-.09	-.45	-.07
Exhilarated	1.00	.92	.72	1.03	-.17	-.23	-.06	-.04	.04	.05	-.31	.16
Amused	.98	.91	.80	.85	-.11	-.22	-.11	.06	.12	.13	-.20	-.08
Original	.70	.75	.83	.77	.25	.16	-.01	.07	-.08	-.22	-.11	-.05
Subtle	.39	.78	.69	.02	.41	.11	.13	.41	-.15	-.08	-.08	-.76
Activated	.35	.79	.91	.03	.53	.28	-.03	.53	-.05	.05	.19	-.71
Puzzled	.31	.50	.96	.30	.70	.63	.03	.81	.07	.04	.32	-.05
Aggressive	.21	.32	.26	.03	.65	.52	.74	.74	.08	.16	-.05	-.14
Tasteless	-.10	-.08	-.23	.05	.83	.81	.89	.70	.01	-.03	-.22	.24
Embarrassing	-.23	-.22	-.08	.10	.94	.92	.85	.81	-.09	-.20	-.08	.09
Indignant	-.21	-.14	.06	-.14	.97	.89	.74	.89	-.06	-.08	.19	-.08
Angered	-.17	-.13	.05	-.25	.87	.80	.49	.84	.08	.13	.42	-.07
Childish	-.15	-.05	.02	.08	-.20	.10	-.11	.22	.46	.54	.81	.55
Simple	-.07	-.10	-.05	.11	-.10	.12	.08	.24	.71	.57	.69	.71
Bored	-.05	-.14	-.10	-.40	-.04	-.08	-.02	.33	.88	.85	.82	.28
Unstimulated	.21	.04	.03	-.05	-.08	-.12	-.04	.35	.95	.97	.90	.54

Table 1: Experiential dimensions of responses to humor as seen in the Promax factor pattern for the 17 rating scales for all humor categories combined and separated in the study by Ruch and Rath (1993)

Note. INC = incongruity-resolution humor. NON = nonsense humor. SEX = sexual humor.

¹The original German terms used were *witzig*, *kindisch*, *aggressiv*, *originell*, *geschmacklos*, *geistreich*, *peinlich*,

3.2 A TAXONOMY OF JOKES AND CARTOONS

What aspects are reflected in individual differences in the perception of humor? Both content and structure have to be distinguished as two different sources of pleasure in humor, and hence one would assume that both are also pivotal in producing individual differences, which, in turn, are relevant in generating a taxonomy of jokes and cartoons. While intuitive and rational taxonomies typically distinguish only between content classes, factor analytic studies show that structural properties of jokes and cartoons are at least as important as their content, with two factors consistently appearing: namely, incongruity-resolution (INC-RES) humor and nonsense (NON) humor. Jokes and cartoons of these factors have different content (e.g., themes, targets) but are similar with respect to the structural properties and the way they are processed. I will briefly describe those factors; for more details on the studies see the original studies or the reviews (Ruch, 1992; Ruch & Hehl, 1998)

Jokes and cartoons of the INC-RES humor category are characterized by punch lines in which the surprising incongruity can be completely resolved. The common element in this type of humor is that the recipient first discovers an incongruity which is then fully resolvable upon consideration of information available elsewhere in the joke or cartoon. There is a certain projective element in these jokes as essential things are not spelled out and have to be supplemented by the recipient; often resolving the incongruity requires attributing motives and traits (e.g., stingy, mean, stupid, absent-minded) to the characters depicted in the jokes. Although individuals might differ with respect to how they perceive and/or resolve the incongruity, they have the sense of having "gotten the point" or understood the joke once resolution information has been identified. At the time this factor was first extracted, it seemed that the two-stage structure in the process of perceiving and understanding humor described by Suls (1972) is a model that fits well to these jokes and cartoons, and hence incongruity-resolution humor was considered to be an appropriate label for that factor.

Nonsense humor also has a surprising or incongruous punch line, however, "... the punch line may 1) provide no resolution at all, 2) provide a partial resolution (leaving an essential part of the incongruity unresolved), or 3) actually create new absurdities or incongruities" (McGhee, Ruch & Hehl 1990; p. 124). In nonsense humor the resolution information gives the appearance of making sense out of incongruities without actually doing so. The recipient's ability to make sense or to solve problems is exploited; after detecting the incongruity he is misled to resolve it, only to later discover that what made sense for a moment is not really making sense. Rothbart and Pien's (1977) impossible incongruities that allow only for partial resolutions are characteristic of the nonsense factor, while their possible incongruities allowing for complete resolutions are more prevalent in INC-RES humor.

While both the incongruity-resolution and the nonsense structure can be the basis for harmless as well as tendentious content, only few contents seem to be salient enough to form independent factors. The pool of jokes and cartoons we analyzed contained different content areas (including aggression), but only sexual humor formed a robust factor overpowering the structure variance. These factors were first extracted in studies of Austrian samples and later replicated in Belgium, England, France, Germany, Israel, Italy, and Turkey (Ruch & Hehl, 1998).

Recently Ruch and Hehl (1998) outlined where the model might need to go both in terms of refining the structural model (e.g., testing for further, perhaps more specific content categories) and the interpretation of the factors. Especially the latter seems necessary, considering that most likely the cognitive processes continue after resolving the incongruity. Unlike after real problem solving, the recipient is aware that the fit of the solution is an "as if"-fit. What makes sense for a moment is subsequently abandoned as not really making sense. At a meta-level we experience that we have been fooled; our ability to make sense, to solve problems, has been misused. Thus, in particular for the impossible incongruities and their partial resolution, it was proposed to move from a the two-step (i.e., step I: detection of incongruity or violation of a build-up expectation; step II: resolution of incongruity) model to a three stage model, where the a third stage refers to detecting that what makes sense is actually nonsense.

lustig, niveaulos, erheitert, gelanqueilt, angeregt, entrüstet, verblüfft, verärgert, amüsiert, and angeödet.

3.3 THE 3 WD HUMOR TEST

The 3 WD ("3 Witz-Dimensionen") humor test (Ruch, 1983) was designed to assess funniness and aversiveness of jokes and cartoons of the three humor categories of incongruity-resolution humor, nonsense humor, and sexual humor. Initially three versions of the test (3 WD-K, 3 WD-A, and 3 WD-B) were constructed. They contain 50 (Form K) or 35 (Forms A and B) jokes and cartoons, which are rated on "funniness" and "aversiveness" using two 7-point scales. The funniness rating ranges from not at all funny = 0 to very funny = 6 and the aversiveness scale ranges between not at all aversive = 0 to very aversive = -6. Forms A and B are parallel tests. They are used together as a long form (with 60 items scored) when reliable measurement is needed or as parallel versions before and after an intervention whose effects have to be evaluated. Form A and B do not overlap, but their purest items form the 3 WD-K, which is a short form. The first five items of each form are used for "warming up" and are not scored. The jokes and cartoons are presented in a test booklet with two or three items on a page. The instructions are typed on the separate answer sheet which also contains the two sets of rating scales.

Six scores can be derived from each Form of the test: three for funniness of incongruity-resolution, nonsense and sexual humor (i.e., INC-RESf, NONf, and SEXf) and three for their aversiveness (i.e., INC-RESa, NONa, and SEXa). These six scores describe an individual's sense of humor at a general level. As mentioned above, sometimes the three subcategories of "pure" sexual humor (PURE SEX), incongruity-resolution based sexual humor (INC-RES SEX) and nonsense based sexual humor (NON SEX) are used in addition to the general sexual humor category. Other indices have been derived as well and were validated in several studies (Forabosco & Ruch, 1994; Ruch & Hehl, 1988; Ruch et al., 1990). For example, a structure preference index was obtained by subtracting INC-RESf from NONf. Similarly, the funniness and aversiveness scores of a humor type could be combined to form a more general appreciation score.

A review of studies of the psychometric properties of the different forms of the 3WD shows that the reliability estimates may be regarded as satisfactory for the scales of all Forms of the 3 WD (Ruch, 1992). The internal consistency varies between .68 and .95, mostly exceeding .80. There is also a sufficiently high degree of equivalence between Form A and B, with parallel test reliability of the six scales ranging from .67 to .93 (with a median of .86) when both forms are filled in on the same day. The samples with a time lag between 2 and 4 weeks yield coefficients between .54 and .85 with a median of .73.

4 HUMOR APPRECIATION AND PERSONALITY

Humor is in the eye of the beholder and thus the identification of those variables that affect the perception of humor is necessary. Why is it that somebody finds a joke absolutely hilarious, the next considers it boring and still another one embarrassing? Many studies have set out to investigate the questions of "what is funny to whom and why" and enriched our understanding of both humor and personality. Conceptually, we need to distinguish among the predilection to certain types of responses and the preference to certain types of stimuli. In other words, is there a generalized tendency to respond positively or negatively to stimuli of all kind (i.e., not only to humor) and what personality traits are predictive of that readiness or typical intensity? Is there an inclination to find certain types of humor funnier than others and what personality traits may account for this tendency?

4.1 INDIVIDUAL DIFFERENCES IN DEGREE OF APPRECIATION

There is increasing evidence that separate factors of positive effect and negative effect can be extracted from different materials (e.g., Watson & Tellegen, 1985). Like funniness and aversiveness these factors are orthogonal. Moreover, extraversion predicts individual differences in positive effect and neuroticism accounts for individual differences in negative affect. Can these relationships can be found in the realm of humor appreciation as well?

There is, indeed, a consistent positive intercorrelation among appreciation of the three humor categories which is low for funniness but relatively high for aversiveness. Thus, there is some room left for the assumption of stable individual differences in the tendencies to find humor generally

more aversive or generally more funny. Since funniness represents the positive responses to humor and aversiveness covers the possible negative ones it could be hypothesized that extraversion correlates positively with funniness of the three humor categories and neuroticism predicts their aversiveness. However, in a review of studies, Ruch (1992) found only spurious effects of extraversion on generalized positive responses to humor. While the zero-order coefficients obtained were overwhelmingly in the expected direction, they generally lack both statistical and practical significance.

However, there seem to be more generalized individual differences in aversiveness and they seem to be correlated with two clusters of predictors (see Table 1). The data collected allow us to confirm the hypothesis that neuroticism is a predictor of negative responses to humor. Typically, scales of neuroticism yield positive correlations, but variables associated with neuroticism were predictive too, like trait-anxiety, depressivity, nervousness, guilt proneness, low ego strength, and even sexual dissatisfaction. The second cluster of variables involved in the general rejection of humor relates to tendermindedness. Humor of all categories is found more aversive by tender than by tough subjects. This could be demonstrated for factors of tendermindedness extracted from personality and attitude inventories, but also related variables were predictive (see Figure 2). Ruch and Hehl (1988) argued that the two groups of predictors might relate to different aspects of aversiveness. The tendermindedness complex might refer to the easiness with which feelings are hurt or subjects feel offended by humor, whereas the neuroticism complex determines the threshold for an negatively toned response and its intensity.

4.2 HUMOR APPRECIATION AND PERSONALITY

The search for predictors of appreciation of humor structure and content was more successful. The rationale for the prediction of personality correlates of appreciation of INC-RES and NON humor was based on the fact that the two humor structures differ with respect to the degree of resolution obtained: in incongruity-resolution humor a complete resolution of the incongruity is possible while there are residual traces of incongruity in nonsense humor. Thus, in INC-RES the resolution of incongruity contributes to appreciation whereas in NON appreciation is based on the existence of residual incongruity. This consideration and evidence from other sources led to the hypotheses that appreciation of the incongruity-resolution structure is a manifestation of a broader need of individuals for contact with structured, stable, unambiguous forms of stimulation, whereas appreciation of the nonsense structure in humor reflects a generalized need for uncertain, unpredictable, and ambiguous stimuli (Ruch, 1992).

Appreciation of incongruity-resolution humor yielded a broad set of predictors (see Figure 1). The single most potent predictor is conservatism, the major dimension underlying social attitudes. According to Wilson's (1973) dynamic theory of conservatism this trait reflects a *generalized fear of both stimulus and response uncertainty*. This should lead more conservative individuals to show greater avoidance and dislike of novel, complex, unfamiliar, incongruous events and to prefer and seek out stimuli which are simpler, more familiar and congruent. This hypothesis was validated for visual art, poetry, and music. Not surprisingly, then, the hypotheses that conservative persons find incongruity-resolution humor more funny than liberals were substantiated in several countries (Ruch, 1992).

The second set of predictors tested the individual's stance towards stimulus uncertainty vs. redundancy more directly using behavioral tests and judgment or creation of art (Ruch & Hehl, 1998). Quite clearly, incongruity-resolution humor is preferred by individuals who generally dislike stimulus uncertainty. For example, appreciation of INC-RES correlated with liking of simple and representational paintings, and with liking simple (such as a triangle, square, or cross) line drawings. Another set of predictors may be circumscribed by inhibitedness, and like stimulus uncertainty it is correlated with conservatism. However, people liking INC-RE humor yield lower scores in depressivity, and they are generally older and tend to fill in questionnaires in a social desirable way.

Nonsense humor is well understood and the correlations are lower in magnitude. Nevertheless, a set of predictors turned up over and over again (Figure 2). While conservatism does not incorporate

the *seeking* of stimulus uncertainty, the trait of sensation seeking, and in particular the component of experience seeking, does. Experience seeking involves the seeking of stimulation through the mind and the senses, through art, travel, even psychedelic drugs, music, and the wish to live in an unconventional style, and there is evidence that it is closely related to the novelty and complexity dimensions of stimuli (Zuckerman 1994). Therefore it was hypothesized and substantiated in several countries that experience seeking is positively related to appreciation of nonsense humor (for details see Ruch 1992). More recently, openness to experience as a basic factor of personality was found to be predictive of nonsense humor (Ruch & Hehl, 1998).

Clearly, liking of nonsense humor is predicted by liking of complexity in a variety of stimuli. The hypothesis that and nonsense humor is appealing to those generally enjoying or searching for uncertainty was also substantiated in the field of aesthetics. For example, appreciation of nonsense correlated positively with liking complex and fantastic paintings (e.g., by Dali), liking of complexity and asymmetry in freehand drawings and polygons, and also with producing complexity in black/white patterns and enjoying and enhancing visual incongruity when wearing prism glasses which distort the visual field (Ruch & Hehl, 1998). A finding yet awaiting replication was that NON goes along with higher sexual libido. They are repeatedly found to be more non-conforming and higher in intelligence. Liking of nonsense peaks between 20 and 35 years of age and declines thereafter.

The prediction of personality correlates of sexual humor can be based on either Freudian theory or salience theory. Freud (1905) hypothesized that repressed needs find relief in jokes and in dreams. Hence, there will be a negative relationship; people repressing their sexual desires will be the ones appreciating sexual content in humor. Salience theory (Goldstein, Suls & Anthony, 1972) predicts a positive relationship; funniness of a particular content in humor will increase with increase in salience of this topic in real life. So, the most promising predictor of appreciation of sexual humor is the individual's sexual experience and attitudes to sex.

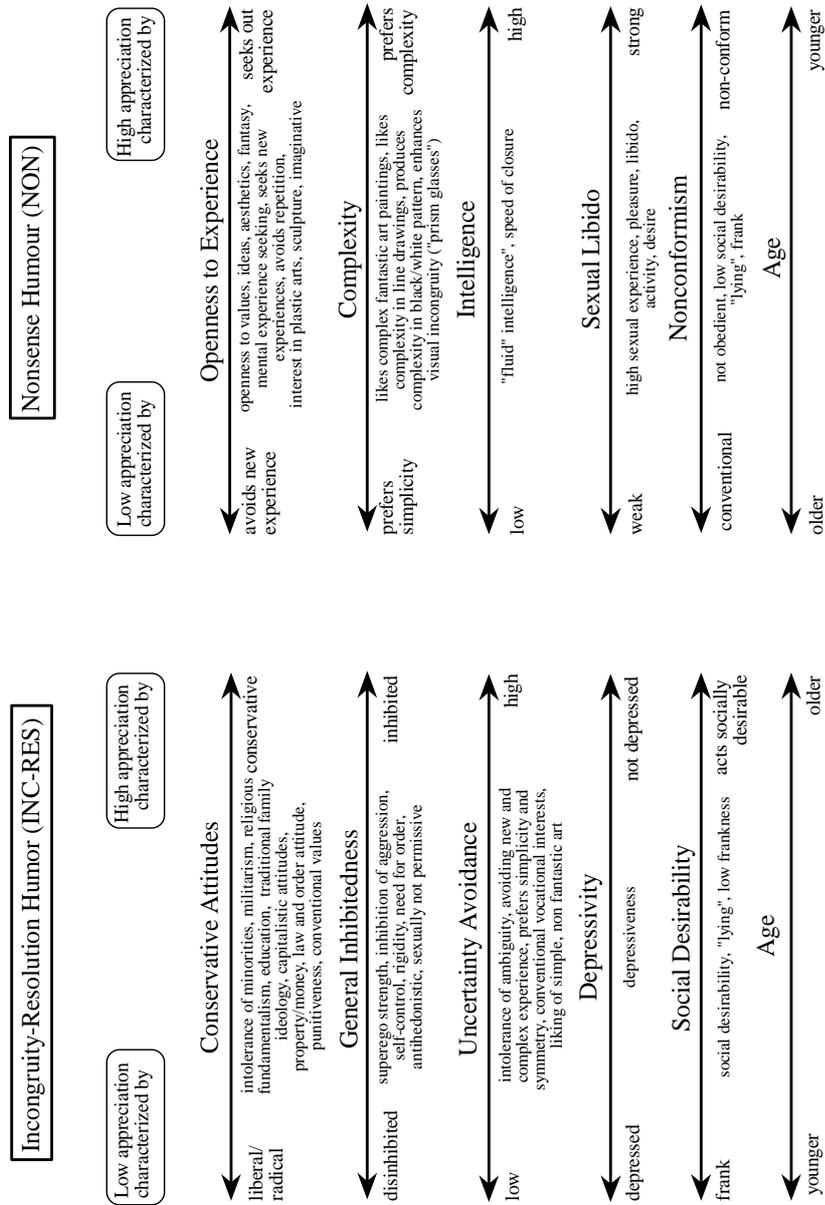


Figure 1: Personality correlates of funniness of incongruity-resolution humor (left) and of nonsense humor (right) in the 3 WD. Figure presents the summary of results across several studies (for details, see Ruch, 1992, Ruch & Hehl, 1998).

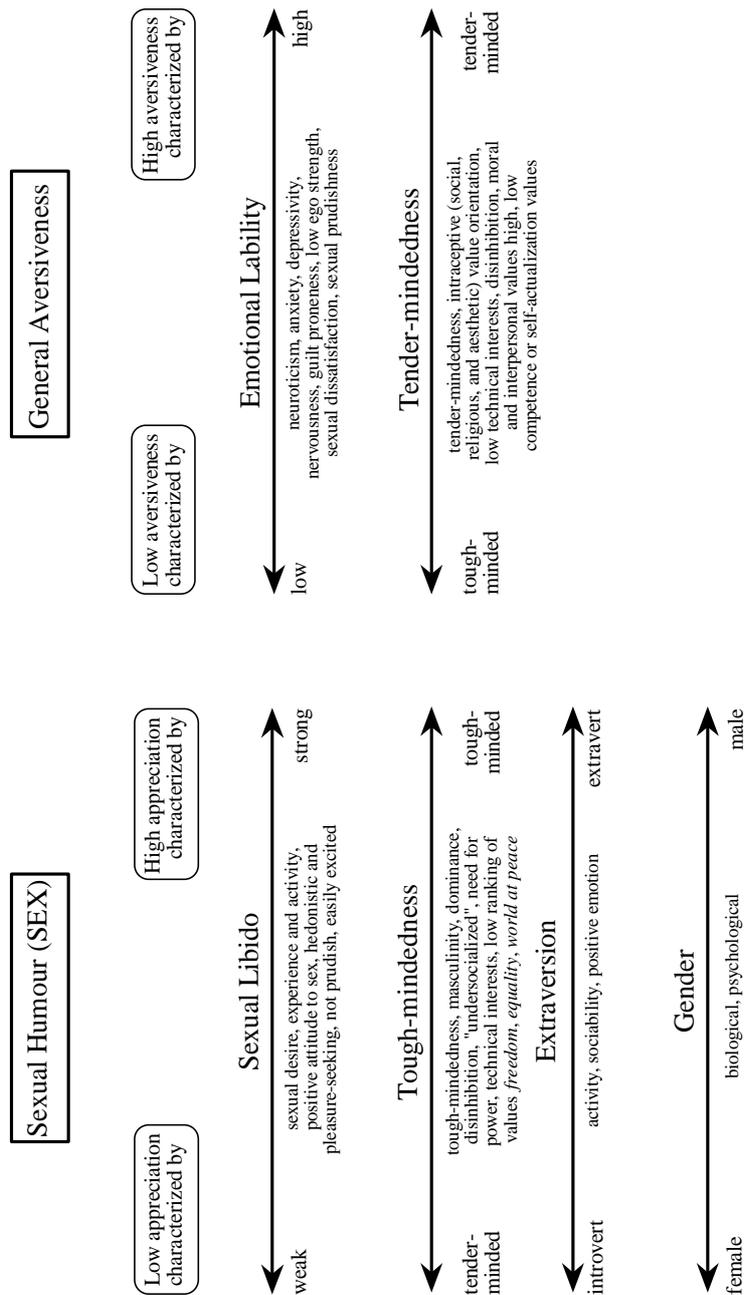


Figure 2: Personality correlates of funniness of sexual humor (left) and of general aversiveness; i.e., the sum of the three aversiveness scores in the 3 WD (right). Figure presents the summary of results across several studies (for details, see Ruch, 1992, Ruch & Hehl, 1998).

There are four predictors of sexual humor, with sexual libido being the most potent (Table 1). The studies clearly show that appreciation of sexual content reflects the degree and valence of salience of this theme; i.e., strength of positive salience contributes to funniness and strength of negative salience contributes to aversiveness of sexual content in humor. The degree of appreciation of the sexual content in humor varies directly with one's attitudes to sex and sexual behavior

rather than inversely, as deducible from Freudian theory. The other cluster of variables predicting sexual humor can be circumscribed by tough-mindedness. Tough- vs. tender-mindedness is the second dimension in the attitude space and orthogonal to conservatism-radicalism. Studies of appreciation of sexual content in humor using the 3 WD converge on the finding that tough-mindedness correlates positively with funniness and negatively with aversiveness of sexual humor (Ruch & Hehl, 1986). The temperamental trait of extraversion is predictive of sexual humor, too. However, this result is not very consistent. Finally, there are gender differences too, which would go along with the other variables (males score higher in tough-mindedness and libido). However, it is not clear whether this is due to the selection of jokes and cartoons used in the 3WD (For a discussion of gender differences see Martin & Ervin-Tripp, 1998).

5 CONCLUSION

The present review was aimed at describing the experiential word of the receiver of humor in terms of a manageable smaller set of descriptive dimensions. Despite the richness of the subjective experience and the tremendous diversity in the humor stimuli, appreciation of humor can be reduced to a two by three matrix. This might not be sufficient for some research approaches but it help as a first orientation in this area. This approach is obviously fuzzy on the side of the encoder or sender, as nothing much is said about the actual process of generating the jokes. It is also not very precise regarding the analysis (or synthesis) of the encoded message, as only *one* global description for a set of jokes is given and this is even done at a more global level.

Nevertheless, the strength of the approach is the focus on the receiver, as also computational humor will want to serve the same client. Is artificially generated humor perceived as artificial by the recipient of a joke? Can computer be programmed in a way that the outcomes is indistinguishable from human humor? For me it seems like that the criterion can be whether or not a random sample of research participants reacts to generated humor in the same way as they do to the 3WD humor test which claims to represent humor appreciation comprehensively (albeit at a very global level). Clearly, the strength of computational approaches is the precision at the stage of the encoder and the control over the variables added or withhold. So far, however, I am missing the consideration of the richness on the side of the receiver as well as the experimental variations on the side of the sender and the message that accounts for the variations in experience on the side of the receiver. It appears to me that a fruitful research endeavor seems to require fusing the strength and expertise of different approaches.

A dialogue across disciplines can not be substituted by writing a book chapter, however, I want to raise a few points that I find worth considering for investigation. What are the necessary variations in the encoding algorithms that generates humor that falls within the INC-RES factor? How to generate humor that jointly loads with the 3WD items on the nonsense factor? Would it makes sense to give the algorithms a personality—a distinct “sense of humor”? There is not only preference for structural types of humor that needs to be considered. Humans make the deliberate choice of being restrained or vulgar in their humor. Humans joke differently when they are in different mood or altered states of motivation. Humans tailor their humor to the audience using information about the audience to optimize (or—in rare cases—minimize) amusement. The audience is sensitive to such variations in their degree of positive and negative responses and makes inferences about the intentions and personality of the sender. Is this a venue computational approaches to humor should consider? Is it feasible at this point in time to implement modules that will make the programs be more “human-like” in their generation of humor? It is the pursuit of questions like this which makes the interdisciplinary field of humor so attractive.

REFERENCES

- Chapman, A.J. and Foot, H.C. (editors) (1977) *It's a funny thing, humour*. Pergamon Press, Oxford.
- Davis, J.M. and Farina A. (1970). Humor appreciation as social communication. *Journal of Personality and Social Psychology*, 15:175-178.

- Ekman, P. (1992). An argument for basic emotions. *Psychological Inquiry*.
- Forabosco, G. and Ruch, W. (1994). Sensation seeking, social attitudes, and humor appreciation in Italy. *Personality and Individual Differences*, 16:515-528.
- Freud, S. (1905). *Der Witz und seine Beziehung zum Unbewußten*. Deuticke, Wien.
- Goldstein, J.H. and McGhee, P.E. (editors) (1977), *The psychology of humor: Theoretical perspectives and empirical issues*. Academic Press, New York.
- Goldstein, J.H., Suls, J.M. and Anthony, S. (1972). Enjoyment of specific types of humor content: Motivation or salience? In Goldstein, J.H. and McGhee, P.E., editors., *The psychology of humor*, pp. 159-171, Academic Press, New York.
- Lampert, M. and Ervin-Tripp, S.M. (1998). Exploring paradigms: The study of gender and sense of humor at the end of the 20th century. In Ruch, W., editor, *The sense of humor: Explorations of a personality characteristic*, pp. 231-270, Mouton de Gruyter, New York.
- Martin, R. A. (1998). Approaches to the sense of humor: A historical review. In Ruch, W., editor, *The sense of humor: Explorations of a personality characteristic*, pp. 15-62, Mouton de Gruyter, New York.
- McGhee, P. E. (1979). *Humor: Its origin and development*. Freeman, San Francisco.
- McGhee, P.E. and Goldstein, J.H. (1983). *Handbook of Humor Research*. New York: Springer.
- McGhee, P.E., Ruch, W. and Hehl, F.-J. (1990). A personality-based model of humor development during adulthood. *Humor. International Journal of Humor Research*, 3:119-146.
- Rothbart, M.K. and Pien, D. (1977). Elephants and marshmallows: A theoretical synthesis of incongruity-resolution and arousal theories of humour. In Chapman, A.J. and Foot, H.C., editors, *It's a Funny Thing, Humour*, pp. 37-40, Pergamon Press, Oxford.
- Ruch, W. (1981). Witzbeurteilung und Persönlichkeit: Eine trimodale Analyse. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 2:253-273.
- Ruch, W. (1983). *Humor-Test 3 WD (Form A, B and K)*. Unpublished manuscript. University of Düsseldorf, Department of Psychology, Düsseldorf, Germany.
- Ruch, W. (1992). Assessment of appreciation of humor: Studies with the 3WD humor test. In Spielberger, C. and Butcher, J. N., editors, *Advances in personality assessment* (Vol. 9), pp. 27-75, Erlbaum, Hillsdale, NJ.
- Ruch, W. (1993). Exhilaration and humor (Chapter 42). In Lewis, M. and Haviland, J.M., editors, *The handbook of emotions*, pp. 605-616, Guilford Publications, New York, NY.
- Ruch, W. (1995). Will the real relationship between facial expression and affective experience please stand up: The case of exhilaration. *Cognition and Emotion*, 9:33-58.
- Ruch, W. and Hehl, F.-J. (1988). Attitudes to sex, sexual behaviour and enjoyment of humour. *Personality and Individual Differences*, 9:983-994.
- Ruch, W. and Hehl, F. J. (1998). A two-mode model of humor appreciation: Its relation to aesthetic appreciation and simplicity-complexity of personality. In Ruch, W., editor, *The sense of humor: Explorations of a personality characteristic*, pp. 109-142, Mouton de Gruyter, New York.
- Ruch, W., and Köhler, G. (1998). A temperament approach to humor. In Ruch, W., editor, *The sense of humor: Explorations of a personality characteristic*, pp. 203-230, Mouton de Gruyter, New York.

- Ruch, W., McGhee, P.E. and Hehl, F.-J. (1990). Age differences in the enjoyment of incongruity-resolution and nonsense humor during adulthood. *Psychology and Aging*, 5:348-355.
- Ruch, W. and Rath, S. (1993). The nature of humor appreciation: Toward an integration of perception of stimulus properties and affective experience. *Humor: International Journal of Humor Research*, 6:363-384.
- Suls, J.M. (1972). A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. In Goldstein, J.H. and McGhee, P.E., editors, *The psychology of humor*, pp. 81-100, Academic Press, New York.
- Watson, D. and Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98:219-235.
- Wilson, G.D. (1973). A dynamic theory of conservatism. In Wilson, G.D., editor, *The psychology of conservatism*, pp. 257-266, Academic Press, London.
- Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. Cambridge University Press, Cambridge.

Humor in Online Romantic Relationships

Aaron Ben-Ze'ev
Department of Philosophy
University of Haifa, Israel
benzeev@research.haifa.ac.il

Abstract

Birds do it, bees do it, and now people with PC's do it. Gloria G. Brame

This paper examines the role of humor in online romantic relationships. I claim that in such relationships, which are based on conversation rather than on external appearance, humor plays a considerably greater role than in offline relationships. This is a positive feature of such relationships as the ability to entertain several different perspectives, which is typical of humor, is a good indicator that such a relationship can endure in the long run. Survival in the long run requires having more than one partial perspective, which is typical of extreme emotional states. In light of the increasing popularity of online relationships, we can expect to see an increase in the role of humor in personal relationships in general.

Keywords: humor, online affairs, emotions, flirting, addiction

1 THE PRESENCE OF HUMOR IN ONLINE RELATIONSHIPS

Your mind is what interests me the most. Bumper Sticker.

The appearance of computer-mediated communication introduces a new type of communication and consequently a new type of personal relationships. There are various kinds of computer-mediated relationships that differ in some significant aspects. My main concern is with one-to-one romantic communication between real people who are not completely anonymous—but whose true identity may still not be fully disclosed.

The complex experience of romantic love involves two basic evaluative patterns referring to (a) attractiveness (or appealingness)—that is, an attraction to external appearance, and (b) praiseworthiness—that is, positively appraising other characteristics. Romantic love requires the presence of both patterns.

Online romantic relationships differ from offline relationships in that they attach less weight to external appearance and more weight to the positive appraisal of the other's characteristics. External appearance is typically not fully known in the early stages of online relationships, and therefore these relationships must rely on other characteristics for creating positive emotional attitudes toward each other. Online relationships prevent people from relying on good looks when evaluating other people and hence they avoid the unjustified advantages that are usually granted to attractive people; these relationships enable people to get to know each other without having to cope with the heavy burden of the attractiveness stereotype (Wallace (1999): p. 138.)

It should be noted that personal traits, including a sense of humor, are also important in offline relationships. In a study of human attraction, six out of ten characteristics considered by both women and men to be most effective in attracting opposite-sex partners had nothing to do with physical attributes and could easily be conveyed in an online communication. Those include a sense of humor, sympathetic toward other people's troubles, good manners, prepared to put effort into spending time together, and willingness to help (Buss (1988) cited in Levine (2000)).

In online relationships, however, the weight of the other person's attraction is considerable smaller, particularly at the beginning of the relationship. Getting to know each other is more

crucial in online relationships. As the information in the first message is quite limited, cases of love at first byte are more rare. Nevertheless, we should not exclude the possibility of such cases—after all, schematic evaluations can also be made on the basis of limited information, which can be revealed in the first conversation. For example, one may detect in the first message a sense of humor and wittiness and instantly fall in love with the sender. As Sandra described her online mate (who has now become her husband): “He was romantic, brilliant, poetic, witty, funny . . . everything I’d dreamed about in a man.”¹

A common phenomenon in offline relationships is the “attractiveness halo,” in which a person who is perceived as beautiful is assumed to be good as well; in online relationships, this is replaced by the “personality halo,” in which a person who is perceived as having a specific, positive personality trait is assumed to have other good characteristics—sometimes even those connected to physical appearance. A sense of humor constitutes a fundamental attribute in the personality halo.

Conversation is of great value in online relationships, and it actually determines the quality of the relationship. The following email message, posted on a message board, exemplifies this:

Subject: The art of the spoken word

Looking for ladies who enjoy being pampered with good conversation and the sensuality of the spoken word. Well-thought conversation is an art, and its appreciation is an extraordinary gift! I love to talk about anything and everything, so if you’re interested in avoiding the immature, crass babble online and having a great talk, please e-mail me or look for me on-line (Cited in Levine (1998)).

It is difficult to imagine the value of conversation being elevated to such heights in an offline relationship. It is refreshing to discover that neither physical contact nor visual content, but rather conversation, is at the heart of true romance and exciting sexual affairs. Conversations are indeed important in forming a good base for a strong relationship. In modern society, we scarcely have the time for genuine and prolonged conversations with those dear to us. In cyberspace, such conversations are common as this is the only means to connect intimately with other people.

Online relationships, which are based on conversations, demand more intense time together—that is, time together without watching television or reading a newspaper. Conversations have a slower pace—they require more time. Although the “slam-bang-thank-you” type of experience also exists in cyberspace, it is less common since a conversation is a kind of ongoing communication. Online conversations force you to interact with your partner in at least some nonsexual aspects. These aspects, as well as the ongoing nature of conversations, encourage the pleasant atmosphere to spill over to the morning after. Most of the time, the impression of what took place the night before is that of a pleasant conversation, like that typical of flirting.

In light of the crucial role of conversation in online romantic relationships, a sense of humor becomes much more important as well. Since online relationships have become more and more prevalent, we can expect an increase in the importance of a sense of humor in human personal relationships.

Online romantic relationships are often a form of flirting. Flirting creates a relaxing, calm, and enjoyable atmosphere. It involves curiosity, humor, imagination, and empathy.

In her discussion on flirting, Miss Etiquette indicates that the meanings of flirting is to be playfully romantic; something of little value or importance. Indeed, her colleague, Miss Manners, considers flirtation “a gentle amusement,” an activity that should be harmless and not lead to anything². Flirting is subtle: it is not an explicit sexual activity, but rather an enjoyable, gentle prelude or substitute for it. Flirting has elements of intellectual teasing flavored by emotional tone. Flirting has seemingly contrary aspects: honesty and a somewhat deceptive attitude (expressed in flattery); caring for others—by listening to and showing interest in them—while not taking the

¹Cyberlove Seminars, <http://www.bluejuice.com/cyberlove/story.html>, “How we met online.”

²Miss Etiquette, Flirting etiquette, http://www.miss-etiquette.com/etiquette_advice.html.

encounter too seriously; being confident and feeling good about yourself while not attaching too much importance to yourself; intelligence mixed with a good sense of humor (Ben-Ze'ev (2002)).

The above contrary aspects are also characteristic of humor. Humor involves honesty, such as when it touches upon the most profound issues in our life; it also involves exaggeration, which may lead to embellishment of the truth and thus to deception. The sensitivity associated with humor indicates that it involves caring for others, but this sensitivity is also associated with not taking others too seriously—sometimes to the extent that other people may be insulted. A sense of humor indicates that we feel good about ourselves, but also that we do not attach too much importance to ourselves. Like flirting, a sense of humor includes a measure of intelligence.

The affinity of online affairs to flirting indicates the affinity of online affairs to games. As in games, in online affairs the process is often more important than the outcome, and as in many games, online affairs are characterized by spontaneity, manifest joy, and a sense of humor (Danet (2001): pp. 7-10).

Online affairs, flirting, games, and humor are essentially intrinsically valuable activities whose value lies in the activity itself, not its results. Although such an activity has results, it is not performed in order to achieve these results. We perform such activities because we enjoy doing so and not because of a certain external goal. The intrinsically valuable nature of online affairs, flirting, and games is another reason for the significant role that humor plays in them.

2 HUMOR AND THE EMOTIONAL INTENSITY OF ONLINE ROMANTIC RELATIONSHIPS

Online sex is a wonderful invention. Now, if only everyone could type faster. Unknown.

The greater role of humor in online romantic relationships raises the issue of whether this role serves to increase or decrease the emotional intensity of these relationships.

Humor is similar to emotions in having a strong element of incongruity or change. Both emotions and humor combine two perspectives—the expected and the unexpected. However, whereas in emotions the simultaneous presence of incongruent perspectives is problematic, and hence requires immediate practical actions, in humor the incongruity is enjoyable and requires no action. The ability to entertain several different perspectives is typical of humor and moderate positions, and is contrary to the partial nature of emotions. A sense of humor is thus often incompatible with an extreme emotional state (Ben-Ze'ev (2000): p. 39). The ability to entertain several alternatives is also a sign of mental health. For example, a person who suffers from paranoia denies that alternatives to his position are possible. Indeed, people often describe their emotionality as a state in which they are unable to think clearly and in particular to appreciate others' points of view (Parrott (1995): pp. 76-77).

Laughing at ourselves serves to distance us from the shaming situation as we join others in taking a fresh perspective at the situation. The new humorous perspective also helps to reduce the significance of the shaming situation. Like humor, confession also involves distancing oneself from the event. By telling others about an event that has shamed us, we join others in a distanced observation of the event and in a way share with them some of the responsibility. Moreover, confession also involves an explicit negative evaluation of the deed—something that also distances ourselves from the event. By negatively evaluating our deed, we consider it as a specific isolated failure, thus keeping our global evaluation of ourselves intact (Ben-Ze'ev (2000): p. 515).

Humor is contrary to extreme emotional states and hence its use in online affairs may decrease the intensity of such affairs. This may be true, but there are other relevant considerations in this regard. First, there are other factors that considerably increase the intensity of online relationships. It is beyond the scope of this paper to describe these factors, and hence I will merely mention some of them: the more pronounced presence of change and available alternatives, the large role of the imagination, and the incomplete nature of online relationships. Second, the use of various perspectives in the initial stages of a relationship is a good indicator that such a relationship can endure in the long run. Survival in the long run requires having more than one partial perspective,

which is typical of extreme emotional states. Moreover, the fact that the relationship has began as an intrinsically valuable activity further contributes to the strength of the relationship.

The reduced importance in online relationships of attractiveness and the increased importance of other personal characteristics is a positive feature in long-term considerations, as the latter are more important for enduring relationships. Getting to know each other's qualities is important in building a strong relationship (McKenna et al. (-)). Humor is an essential factor in this regard.

3 HUMOR AND ONLINE ADDICTION

The Internet is so big, so powerful and pointless that for some people it is a complete substitute for life. Andrew Brown

A major danger inherent in using the Internet is that of becoming addicted to it. This is particularly true concerning those engaged in online romantic and sexual affairs. People, who are careful to avoid offline extramarital affairs, are more easily drawn into online extramarital affairs. One reason for this is that the latter are considered to be less real than the former. When something is considered to be less real, its risks are considered to be less significant. Another reason for the addictive nature of online affairs is the lack of an established and familiar mechanism to warn us against being involved in affairs we do not want to pursue. The permissive nature of cyberspace gives rise to fast online sexual affairs without leaving much time for the participants to hesitate about their willingness to be involved in such relationships.

Online romantic relationships are kind of a new toy with which the human race has not yet learned how to play. People often confuse the toy with reality and ruin their personal as well as professional life. Cybering is similar, in a few significant senses, to taking drugs. Both cases provide easy access to pleasure, which is often based upon virtual realities. In both cases, the tempting results can make people dangerously addicted to the method; people want more and more, but satisfaction is limited and becomes more costly to achieve. An unfulfilled craving for drugs and cybering can cause great distress. Once the first steps are taken in online affairs or drugs, the situation can often run its own course, almost involuntarily. Whereas drugs artificially stimulate pleasure centers in the brain, online conversations artificially stimulate pleasure centers in the mind. Artificial stimulation may appear to be easy and cheap; however, the price can be dear in terms of our overall performance and in particular, in terms of the price that those close to us in our offline circumstances might have to pay. Addiction is indeed common in cyberspace (Cooper et al. (2000)).

One way of reducing the likelihood of becoming addicted to cybering is to use humor in online relationships. Addiction is an extreme behavior that does not encompass other perspectives beside that involved in the addictive behavior. Using humor may be useful in this regard. As indicated, the ability to entertain several different perspectives is typical of humor and of all moderate positions. We should encourage people to entertain several perspectives in their daily life. A useful and enjoyable way to do so is through humor.

Engaging in humor is essentially a social activity. It was found that children who watched humorous films alone laughed less than those watching in pairs, who in turn laughed less than children in groups of four or more. Moreover, children watching humorous material with a friend showed more amusement than those watching with a child they did not know well (Chapman and Wright (1976); Parkinson (1995)). The social aspect of humor may also be a factor in reducing the risk of online affairs. Social relationships give you additional perspectives. Such relationships may also influence us negatively, but in many cases humor can help to prevent the negative social influences from swaying us unduly.

The use of humor in online romantic relationships can be risky as well. The reliance of online relationships on one type of communication may sometimes lead to negative emotions due to misunderstanding. Thus, something that is intended ironically may cause the other person to feel insulted and angry. In offline relationships, other clues clarify and contextualize the irony or humor, so that the intention behind the words becomes obvious; in online communication, such balancing factors are absent. The lean communication of online relationships may also generate intense

positive emotions because the negative aspects of the correspondent tend to remain concealed.

4 CONCLUSION

Let me make this much clear: I don't fish in the desert, I don't sunbathe in the shower, and I do not have cybersex with my husband. Unknown

In this preliminary study, I have suggested that the role of humor in online romantic relationships is greater than that in offline relationships. This is, no doubt, a positive feature of such relationships. Online romantic relationships are not without their own difficulties. It seems that the use of humor can be beneficial for coping with such difficulties as well. Substantiating the above claims and analyzing their implications will require further research.

REFERENCES

- Ben-Ze'ev, A. (2000). *The subtlety of emotions*. Cambridge, Ma.: MIT Press.
- Ben-Ze'ev, A. (2002). Flirting on and offline. *Phenomenology and the Cognitive Sciences*, 1.
- Buss, D. M. (1988). The evolution of human intrasexual competition. *Journal of Personality and Social Psychology*, 54, 616-628.
- Chapman, A. J. & Wright, D. S. (1976). Social enhancement of laughter: An experimental analysis of some companion variables. *Journal of Experimental Child Psychology*, 21, 201-218.
- Cooper, A., Delmonico, D. & Burg, R. (2000). Cybersex users, abusers, and compulsives: New findings and implications. *Sexual Addiction & Compulsivity*, 7, 5-29.
- Danet, B. (2001). *Cyberpl@y: Communicating online*. Oxford: Berg.
- Levine, D. (1998). *The joy of cybersex: A guide for creative lovers*. New York: Ballantine Books.
- Levine, D. (2000). Virtual attraction: What rocks your boat. *CyberPsychology & Behavior*, 3, 565-573.
- McKenna, K. Y. A., Green, A. S. & Gleason, M. E. J. (in press). Relationship formation on the Internet: What's the big attraction? *Journal of Social Issues*.
- Parkinson, B. (1995). *Ideas and realities of emotion*. London: Routledge.
- Parrott, W. G. (1995). The heart and the head. In J. A. Russell, J.M. Fernandez-Dols, A. S. R. Manstead & J. C. Wellenkamp (Eds.), *Everyday conceptions of emotion*. Dordrecht: Kluwer.
- Wallace, P. (1999). *The psychology of the Internet*. Cambridge: Cambridge University Press.

Generation of Idiom-based Witticisms to Aid Second Language Learning

Justin McKay
Binsted-McKay
Tokyo, Japan
justinm@binsted-mckay.com

Abstract

In this paper, we discuss a model of simple idiom-based witticisms, which we have implemented on a computer. The resulting program, WISCRAIC (Witty Idiomatic Sentence Creation Revealing Ambiguity In Context) generates jokes by deriving meaning partially from the normal context in which key words appear and partially from humour-independent lexical entries. WISCRAIC also produces explanations of the jokes it creates, with a view to helping second language learners master the idioms that form the basis of the jokes.

After reviewing some of the humour research that is most relevant to WISCRAIC, we look briefly at the motivation for using humour as a tool for language learning, before examining the design of the system itself.

Finally we present the results of an evaluation of the system's output by a group of joke judges.

Keywords: Computational Humour, Second Language Learning

1. INTRODUCTION

WISCRAIC (Witty Idiomatic Sentence Creation Revealing Ambiguity In Context) is the implementation of a model of a sub-class of puns, namely simple idiom-based witticisms. The program generates jokes by deriving meaning (semantic associations) partially from the normal context of words and partially from humour-independent lexical entries. The system produces a range of jokes, but focuses primarily on witticisms (question-answer and one-liners) that use phonological ambiguity to create jokes based around pseudo-idioms (idioms altered during the course of joke production).

WISCRAIC also produces explanations of the jokes it creates, with a view to helping second language learners master the idioms that form the basis of the jokes. Although the results of our informal evaluation suggest that users were helped by the system, they are not conclusive, and a full-scale formal evaluation is beyond the scope of this research. WISCRAIC is different from previous methods of producing computational humour not only in its capacity to produce explanations of jokes produced, but also in that it tries to derive meaning implicitly from context, rather than using explicit syntactic and semantic categories. For this reason it does not require a structured lexicon which systems such as JAPE [1] depend upon. WISCRAIC's linguistic information is also completely general in nature. It is not tailored in any way for the purposes of joke building. The output of the system was tested on human joke judges with positive results.

2. HUMOUR RESEARCH

2.1 AMBIGUITY AND WIT

Pepicello and Green [12] hold the common view that humour is closely related to ambiguity. Linguistic ambiguity, the type that we are concerned with in WISCRAIC, can take place at a number of different levels:

Phonological - ambiguity concerning the sounds of words

e.g. What bird is lowest in spirits?

A bluebird [12]

This is phonologically ambiguous because 'blue' is a colour, but it also means down or depressed.

Morphological - ambiguity concerning word formation (as inflection, derivation, and compounding)

e.g. The book thief was caught read handed

This is morphologically ambiguous because 'read' is only phonetically identical with 'red' in its past participle form.

Syntactic - ambiguity relating to the rules of syntax or syntactics

e.g. Would you rather have an elephant kill you or a gorilla?

I'd rather have the elephant kill the gorilla [12]

The ambiguity here arises because the question can be parsed in two distinct ways. Under one interpretation, the choice of death is by gorilla or by elephant, whilst under a second, the listener is asked if he would like to be the elephant's victim, or would he prefer that the gorilla were the victim.

The important element which these three jokes share, and which is of particular importance in WISCRAIC, is that the ambiguity, be it phonological, morphological or syntactic creates a point of semantic comparison.

Pepicello and Green claim that this ambiguity must be unsolvable by the listener until the punch line resolves it in some unexpected way. Although this is true of the question-answer format produced by WISCRAIC,

Q. Who broke the woman's hart?

A. The cruel deer-keeper.

there is no such resolution in the one-line witticisms, as there is no punch line.

The cruel deer-keeper broke the woman's hart!

It is probably sufficient therefore, that to the listener of a WISCRAIC witticism, the sentence make sense to a first approximation, and upon further examination (of the text, or simple reflection on the sentence heard), the ambiguity is revealed and a second interpretation is made possible.

Indeed it could be argued that the ambiguity is never resolved, as both interpretations are equally valid.

2.2 HUMOUR COMPUTATION

Attardo and Raskin put together a simple joke generating system, LIBJOG (Light Bulb JOke Generator) [8], mainly to show how poorly simple cut-and-paste methods work. However, the most significant work in the area of computational humour was carried out by Binsted.

Binsted [2] examined the based the linguistic structure of a class of jokes, namely punning riddles, and devised a set of rules that allow for the production of question-answer riddles. These rules were implemented in a system called JAPE.

The final version of JAPE developed by Binsted can generate riddles which:

- use typical subtypes of juxtaposition, substitution or comparison as their punning mechanism
- use the constructed word or phrase in the punch line, rather than the question part of the riddle

A comparison of JAPE and WISCRAIC is given in section 6.

3. HUMOUR AND LEARNING

Minsky's Frame Theory can be applied to linguistic humour, specifically puns, which work by changing the meaning or sense of a word. Minsky believes that a common element to all kinds of humour is "unexpected frame-substitution, in which a scene is first described from one viewpoint and then suddenly -- typically by a single word -- one is made to view all the scene-elements in another, quite different way"[3].

We believe that this frame shifting is of benefit to language learners in that it gives the learner/speaker a way of linking semantic frames, and each connection in the frame network provides another way of remembering a given word, or associated idiom. The good feeling associated with such humorous frame shifting may reinforce and strengthen the connection between words. Even bizarre or tenuous shifts in frame (and possibly logic) may allow a learner to recall a word that cannot be remembered directly.

Research has also been carried out into the use of humour in teaching. This work has shown that humour has benefits in teaching whether it is related to the subject matter or not [5]. There have been no conclusive findings as to whether or not the use of humour aids the learning process. However, research does show that relevant humour actually aids retention [4].

An example of a WISCRAIC witticism and associated explanation is given below:

The friendly gardener had thyme for the woman !

The word time, which is part of the idiom [have,time,for,someone] is a homonym of the word thyme.

A HOMONYM is a word that sounds like another word.

| LINK | between thyme and gardener :

| thyme is a type of plant

| a gardener works with plants

"friendly", which is associated with the idiom [have,time,for,someone] was selected from other adjectives as it has the highest imagability score : 439

By outlining the semantic and phonetic associations underlying the joke, we believe that the system promotes deeper thought about the words being used, introducing new or potentially troublesome (on account of its homonymy) vocabulary in a novel and hopefully memorable way.

4. WHAT MAKES A JOKE?

We will now look briefly at some of the techniques that are used to produce jokes, and how such jokes may be classified. Then we will look in a little more detail at the jokes that WISCRAIC is designed to produce.

4.1 QUESTION-ANSWER RIDDLES

A riddle is a puzzling (and often misleading) question posed as a problem to be solved.

A sub-genre of riddles, which was addressed by Binsted [2] in the development of JAPE, is punning riddles.

A pun consists of using a word in a manner that suggests two or more of its meanings, or the meaning of at least one of its homonyms.

Punning riddles therefore exploit ambiguity in either the question or answer part of a riddle.

An example of the question-answer riddles produced by WISCRAIC is shown below:

Q. Who showed the woman his mussels ?

A. *The vain fisherman*

4.2 WITTICISMS

A witticism is a clever and often ironic remark.
A one-liner can be a very succinct joke or witticism.

Some amusing witticisms (and the people they are attributed to) are given below:

“If other people are going to talk, conversation becomes impossible.” – James McNeill Whistler [13]

“I have had a perfectly wonderful evening. But this wasn’t it.” - Groucho Marx [13]

“Avoid all needle drugs. The only dope worth shooting is Richard Nixon.” – Abbie Hoffman [13]

We can see that this last witticism uses ambiguity of the word ‘dope’ to humorous effect.

An example of a WISCRAIC witticism is :

The obliging dairy farmer met the woman half whey !

WISCRAIC also produces character-based witticisms. If the user enters a name, occupation and adjective describing the character, WISCRAIC will attempt to return a witticism about the named person, searching its existing knowledge bases for the necessary semantic connections.

For instance, if we enter ‘Johnny, deer-keeper, poor’, WISCRAIC will return the following witticism:

Your mate Johnny is a hard up deer-keeper. He really needs doe !

5. SYSTEM DESIGN

A common mechanism in jokes that use phonological ambiguity is substitution.

WISCRAIC uses this mechanism to construct a sentence using an altered idiom that is phonologically identical to the original idiom. The choice of substitution is guided by the availability of a noun phrase profession (e.g. the doctor, the chef) that in some way creates a semantic link, through shared context, to the substitute word.

Strictly, an idiom refers to an expression in a given language whose meaning cannot be derived solely from the meaning of the words comprising the expression. It is important to note that although attention is focused on idioms throughout this thesis, WISCRAIC is designed in such a way that it will handle any verb phrase.

All the witticisms:

- use word-word substitution as their main mechanism
- substitute phonetically identical words
- substitute into an idiom
- **either** use the pseudo-idiom thus constructed in the question part of the joke if it is presented in question-answer format
e.g. Who met the woman half whey ?
The obliging dairy-farmer
- **or** construct a sentence using this pseudo-idiom
e.g. The friendly gardener had a lot of thyme for the woman !

The jokes also use adjectives to ensure that the meaning of the original idiom is suggested, if not obvious. In the examples above, “obliging” suggests “meeting someone half way” while “friendly” hints at “having time for someone”.

The set of preconditions that must be satisfied in order for a joke to be produced are as follows:

- There is an idiom A in the idiom database
- There is a word B in the text of A which the word dictionary shows to be a verb or noun
- There is a word E which is a homophone or alternate meaning of B and is defined in the lexicon

- There is a word C, specified as being suitable for use with idiom A, which the dictionary shows to be an adjective
- There is a profession or character D (in database) which is semantically linked to the substitute E

WISCRAIC attempts to construct a witticism based on a common idiom. It has several distinct knowledge bases and processing modules with which to accomplish this task.

5.1 KNOWLEDGE BASES

- a dictionary of idioms, containing a list of adjectives associated with each idiom.
- a dictionary of professions, containing a list of thematic vocabulary for each that defines a context for the profession.
- a general dictionary of nouns, verbs and adjectives, containing a phonetic transcription and values for four psycholinguistic indicators for each entry.
- a lexicon, containing a list of defining words with each entry.
- a grammar containing information on such things as verb conjugation, valid noun phrases and plural forms of nouns.

A distinction is drawn between "dictionary" and "lexicon" as they are used in WISCRAIC.

WISCRAIC's lexicon defines a relation between each word entered as part of a definition for a lexeme and the lexeme itself. For instance, the definition for the lexeme "bough" asserts that the relation *is_part_of_a* holds between "bough" and "tree". The lexemes in this lexicon are the words that are found as homophones for words comprising the idioms. These relations are not used in the construction of witticisms but are used rather to produce the explanations, as shown above.

WISCRAIC's other knowledge sources however contain no such relations. In the profession database, the profession "chef" for example has "kitchen" as one of the words in its list of associated vocabulary but there is no indication of how the two are related. The reason for this design choice was to stay true to the original idea of manipulating context. The context in which any word appears is not defined through the use of semantic categories, but rather by the frequency with which the given word occurs in writing or speech with other words. Words which are immediately associated with a given word provide a suitable context with which to make jokes based around puns on that word.

Note that all information contained in the lexicon, dictionaries and DCG is completely general in nature. In other words it is humour-independent.

5.2 PROCESSING MODULES

- **Joke Constructor** This module contains information on what elements and relations between elements must be present for a joke to be constructed.
Uses: dictionary of idioms, dictionary of professions, general dictionary, lexicon
- **Surface-form Generator** This module is responsible for taking the elements provided by the Constructor above and converting them into a complete joke form (the form which is printed to screen).
Uses: the grammar
- **Explanation Generator** Takes the elements provided by the Constructor and generates an explanation of the relations between them.
Uses: the grammar

5.3 ADJECTIVES

5.3.1 The Role of the adjective

The sentence "The lumberjack took a bough" is a valid grammatical sentence, but it also semantically valid. It makes perfect sense for a lumberjack, someone who works with trees, to take a branch of a tree for some reason. What we have lost therefore is the meaning of the original idiom "take a bow". This idiom creates the mental image of a performer on stage acknowledging his audience's appreciation.

The delivery of a joke often suggests a joke - the ambiguity or innuendo is stressed through mechanisms such as facial expression or vocal stress. A joke is rarely delivered in the same fashion as a factual statement. Computer generated witticisms do not have access to these mechanisms.

In WISCRAIC, the adjective is used as an indicator of what is to come - the reader expects the regular idiom, and does not spot the ambiguity immediately.

To this end, WISCRAIC selects an adjective from a list of adjectives associated with each idiom to qualify the protagonist in the sentence. For instance, the list for the idiom "take a bow" is: performing, entertaining, flamboyant.

Any of these adjectives may be chosen to qualify lumberjack and will result in a reversion to the original meaning of the idiom "take a bow", and hopefully the realisation on the part of the user that the word bough is replacing bow from that idiom.

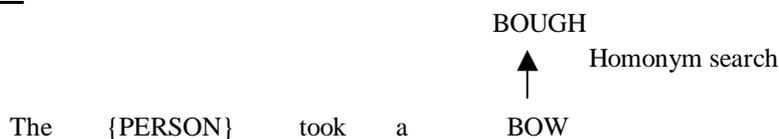
The use of the adjective to suggest the original idiom, and the simultaneous existence of the semantic link between the substituted word and the protagonist of the sentence result in the reader having the simultaneous view that a) the text constitutes a normal usage of the idiom, and b) something is not quite right about that interpretation. In general terms, the reader of a WISCRAIC witticism has in mind the simultaneous view that the situation is normal and also that there is some sort of violation of the way things should be.

5.3.2 Choosing an adjective

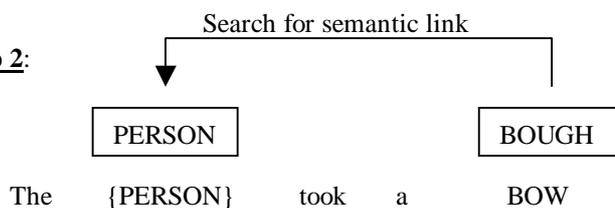
WISCRAIC uses the MRC Psycholinguistic Database to gather information about a number of psycholinguistic indicators for each of the adjectives appearing in the list. The adjective with the highest imageability score is selected first. Upon backtracking, the next highest adjective is selected and so on until the list is depleted. Any adjective appearing in the list that does not have an entry in the MRC Database will never be selected. This ensures that obscure adjectives will not be used in the joke.

5.4 WORKED EXAMPLE

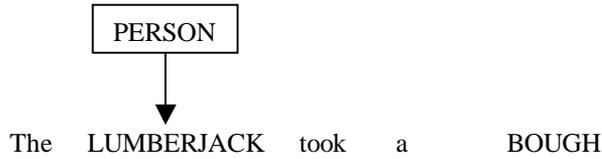
Step 1:



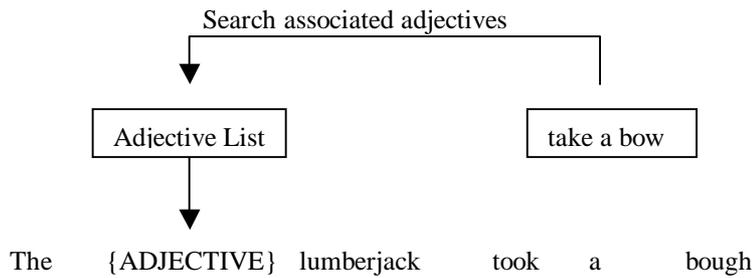
Step 2:



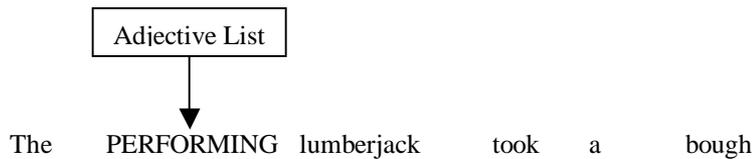
Step 3:



Step 4:



Step 5:



6. DIFFERENCES IN THE DESIGN OF JAPE AND WISCRAIC

There are broad architectural similarities between JAPE and WISCRAIC; both manipulate data from a lexicon, both use DCG grammars to help build surface forms and ensure grammatical correctness and both specify a set of preconditions that must be met in order for a joke to be constructed.

However, there are also important differences in the type of information stored in the respective Knowledge Bases, in the approach to data representation and in the scope and type of jokes produced.

Perhaps the most obvious difference between the two systems is the type of joke each produces. JAPE produces a wide range of question-answer punning riddles built around compound nouns and phrases. WISCRAIC on the other hand focuses principally on witticisms based around idioms and some other verb phrases. WISCRAIC also produces a limited range of more traditional question-answer riddles, like JAPE, though the mechanism used is somewhat different.

One important difference between the two systems is that to a large degree JAPE uses explicit syntactic and semantic information to derive jokes whereas WISCRAIC's knowledge is often implicit. WISCRAIC aims to derive meaning from context rather than word definitions. For instance, JAPE uses categories such as 'inact_verb', 'act_verb', 'used_to', 'used_to_object' and 'specifier'. The information therefore is very structured, and as was the case with JAPE-1, volunteers are quite restricted in how they can define a given word. WISCRAIC uses a more open-ended approach. Only broad syntactic categories such as noun, adjective, and verb are explicitly coded into WISCRAIC. The important element in WISCRAIC is that two words are semantically related in some respect - the nature of the link is not considered in the production of a joke. Of course, WISCRAIC data is not completely free of structure or constraint. When gathering data for the idiom dictionary, volunteers were told to think of adjectives associated with a particular idiom. In the case of the data

on the professions, volunteers were told only that the words they provided should be nouns - the association and reasons behind choosing a particular word were never declared or considered in the design. All that mattered is that there was some type of association in the mind of the volunteer. One of the aims in designing WISCRAIC was to see if a word could be sufficiently defined by context alone to form the basis of jokes that others would find funny and clever, without the basis for this funniness and cleverness being explicitly stated at any stage of production. For this reason it can be said that WISCRAIC's semantic knowledge is largely implicit - it is tacit in the minds of the people who provided the words, yet is sufficiently explicit to allow jokes to be built.

JAPE on the other hand has been designed from a standpoint of having analysed jokes and identified certain semantic and syntactic patterns: this allow rules to be derived which when applied and satisfied will reproduce a particular range of jokes.

7. EVALUATION

Once the development of WISCRAIC's model had been completed and its implementation finished, it was necessary to evaluate its performance.

The purpose of the evaluation was to test whether or not a number of hypotheses hold. The most fundamental question we must answer of course is ``Does WISCRAIC produce jokes ?"

We also wished to determine the factors that contributed to the quality of a joke. In order to evaluate the quality of the jokes (and determine the factors which contributed to or detracted from the quality), questionnaires were developed and distributed to native English-speakers, who judged a selection of WISCRAIC's output.

A further aspect we wished to evaluate is the quality and effect of the explanation generated by WISCRAIC. To this end, another questionnaire was developed and distributed to non-native English speakers. The evaluation of the explanations was not carried out to test any specific hypotheses, but rather to provide a basis for discussion and to give some preliminary indication of the value of the system as a teaching tool.

Through this evaluation we hoped to show that :

- WISCRAIC's output is in fact jokes.
- Puns without the supporting adjective would have lower funniness ratings than the same pun with the adjective.
- Cleverness is correlated with funniness: jokes with higher cleverness ratings than other jokes would also have higher funniness ratings.
- Texts without homonym substitution into the idiom but with semantic links derived from the homonym would sometimes be thought of as weak jokes.
- Texts demonstrating normal use of an idiom would be judged as regular sentences.

7.1 EXPERIMENT DESIGN

The judges' ability to distinguish jokes from non-jokes needed to be checked, so non-jokes were also included in the questionnaires distributed to judges.

Five sets of materials were prepared:

- the WISCRAIC generated jokes.
- the sentences illustrating normal use of the idioms.
e.g. *The mysterious man disappeared into thin air !*
- the WISCRAIC jokes with the adjective removed.
e.g. *The lumberjack took a bough !*
- texts with homonym substitution but no semantic links.
e.g. *The strong policeman met the woman half whey !*
- texts with the original idiom intact but semantic links with a potential, phonetically identical substitute.
e.g. *The strong fisherman showed the woman his muscles !*

Each questionnaire comprised 10 texts to be judged - 6 jokes (set 1 above) and one text each from sets 2,3,4 and 5. It was important that a range of jokes be evaluated and also that the ordering of the jokes, controls and hypotheses-testing texts be varied across the questionnaires. Steps were taken to ensure this were so.

7.2 PARTICIPANTS & PROCEDURE

An initial e-mail asking for volunteer joke-judges was sent to the Masters-level AI and Cognitive Science students of Edinburgh University, as well as a number of outside parties. The mail explained that those willing would be asked to judge a set of jokes as part of project exploring written humour.

The majority of volunteers were British, aged between 25 and 30 and educated to first-degree level or higher.

Each text in the questionnaire has three potential rating slots: Joke Rating, Funniness Rating and Cleverness Rating. The volunteers were asked to indicate whether or not they considered each text to be a joke, by marking a 'J' in the Joke Rating box. For each text marked 'J', two additional values must be provided: a funniness rating and a cleverness rating. The scales for these two ratings are provided in the questionnaire so everyone judges by the same scale.

If a text is not given a 'J' rating, it must be given one of the following ratings (taken from the actual instructions accompanying the questionnaires):

- O - Obscure - I'm not familiar with certain words used.
- S - Not a joke - simply a regular sentence.
- N - Nonsense - doesn't make any sense.
- V - Recognisably an attempt at a joke but doesn't work as a joke for some reason.

7.3 RESULTS & DISCUSSION

We now present a summary of the results and how they relate to our original hypotheses (presented in section 7).

- WISCRAIC produces jokes
Confirmed - 84% considered jokes.
- Lower funniness scores for jokes without adjectives
Unconfirmed - 80% considered jokes - slightly higher funniness score than average.
- Cleverness correlated with funniness
Confirmed - of the jokes rated 5 times or more, 80% had the same funniness and rating scores, either 2 or 3 for both measures.
- Suggested Jokes thought of as weak jokes
Confirmed - 40% judged 'J' - lower than average funniness rating
- Texts showing normal use of idiom judged as regular sentences
Unconfirmed - just over half judged as regular sentences

A factor that affected the results of the evaluation was the number of subjects who actually completed and returned the questionnaires. Of over 50 people who received the request to take part in the evaluation, a small number agreed to do so and only 15 completed questionnaires were eventually returned. Another problem is that because only a small number of questionnaires were returned, the planned distribution of the jokes was essentially lost. Some jokes were evaluated ten times whereas others were only rated once.

Although most of the idioms used by the system should be well known to the volunteers, some, such as "wear a cat suit" might not be instantly recognisable. Also, jokes are not often based around idioms, so people may not be in "joke-mode" when they are presented with texts of this genre.

8. CONCLUSION

We succeeded in developing a model of a sub-class of puns, namely witticisms that use idioms as the basis of the joke. This model was then successfully implemented in a computer program, WISCRAIC.

Three of our original hypotheses have been retained. While the other 2 were not rejected, the evidence was not strong enough to confirm them. The limited evaluation carried out seems to suggest that the explanations did aid understanding and promote learning.

A lack of constraints on some of WISCRAIC's mechanisms means that jokes are prone to tenuous links. This would become more apparent if the system were scaled up to use online rather than hand-built methods.

9. CURRENT RESEARCH

Since the development of WISCRAIC, our focus has been on developing and implementing interactive models of humour. These systems are able to learn new semantic relations through user interaction.

An important part of the interactive process is being able to converse with the system. Consequently, we have concentrated on developing dialogue engines that can introduce jokes (both system and human generated) into dialogue at a contextually appropriate juncture.

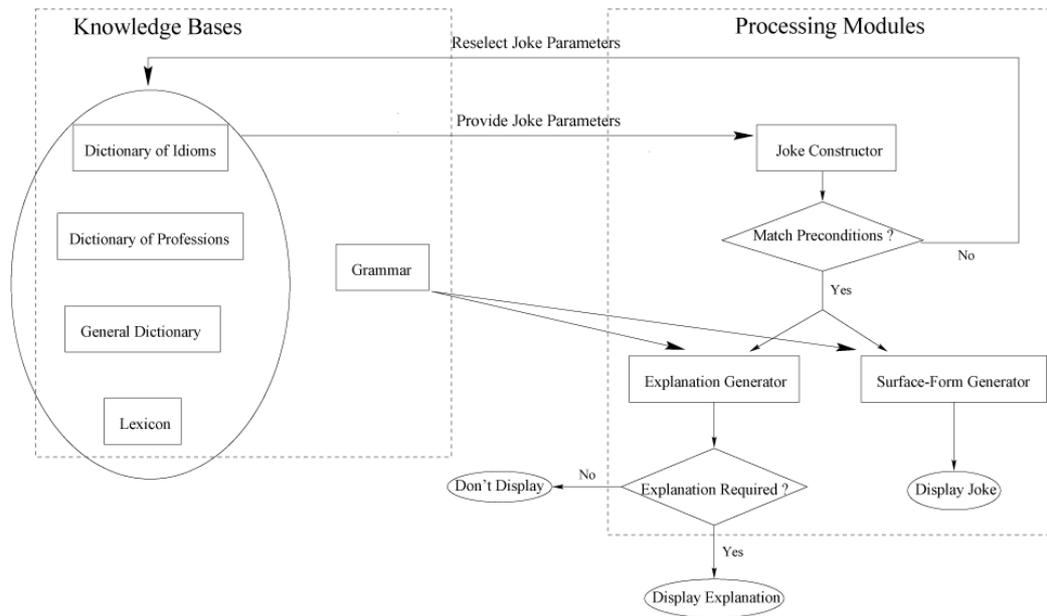
Such systems are still in the early stages of development, but progress so far has been encouraging.

REFERENCES

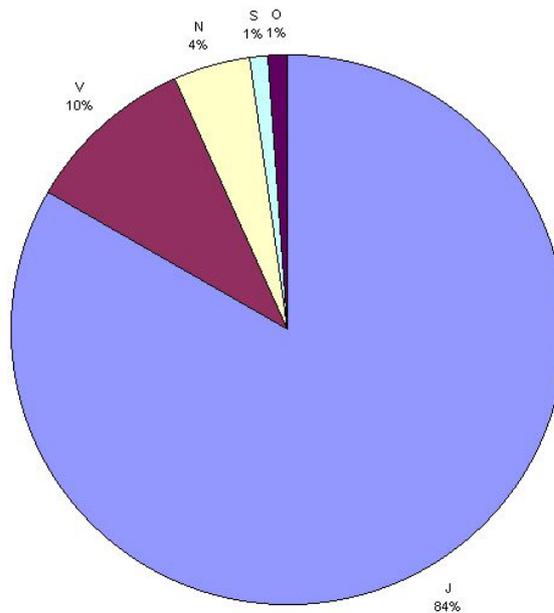
- [1] Binsted, K. (1996). Machine Humour: An implemented model of puns. *PhD Thesis*, Dept Artificial Intelligence, University of Edinburgh, 1996.
- [2] Binsted, K. (1993). Computer Generation of Linguistically Definable Riddles. *Technical Paper No.25*, Dept Artificial Intelligence, University of Edinburgh, 1993.
- [3] Minsky, M. (1981). Jokes and their Relation to the Cognitive Unconscious. *Web Version of Paper*, <http://www.ai.mit.edu/people/minsky/papers/jokes.cognitive.txt>, 1981.
- [4] Kaplan, R.M & Pascoe, G.C. (1977). Humorous Lectures and Humorous Examples, *Journal of Educational Psychology*, Volume 9, pages 61-65, 1977.
- [5] Townsend, A.R, Mahoney, P & Allen, L.G. (1983). Student perceptions of Verbal and Cartoon Humor in the Test Situation, *Educational Research Quarterly*, Volume 7, pages 17-23, 1983.
- [6] Freud, S. (1957). Humour, *Hogarth Press*, Volume 21, pages 161-166, 1957.
- [7] Attardo, S & Raskin V. (1991). Script theory revis(it)ed: joke similarity and joke representation model, *Humor*, Volume 4, No. 3, pages 293-347, 1991.
- [8] Attardo, S & Raskin V. (1993). Non-Literalness and Non-Bona-Fide in Language: Approaches to Formal and Computational Treatments of Humour and Irony, *Unpublished paper*, 1993.
- [9] Redfern, W. (1996). Puns: Second Thoughts, *HUMOR: International Journal of Humor Research*, Volume 9-2, pages 187-198, 1996.
- [10] Veatch, T. C. (1998). A Theory of Humor, *HUMOR: International Journal of Humor Research*, May, <http://www.sprex.com/else/humor/paper/humor.html>, 1998.
- [11] Ritchie, G. (1999). The JAPE riddle system, pages 89-108, 1999.
- [12] Pepicello & Green. (1984). The Language of Riddles, *Ohio State University*, 1984.
- [13] The Urban Legend (2000), WebZine, 29th August 2000, http://homepages.ihug.co.nz/~dbaxo/urban_legend1024.htm

APPENDIX

Information & Processing Flow in WIS CRAIC



How WISCRAIC's output was rated



Real and Fictional Ridicule

A Cognitive Approach for Models of Humour

Francesco Mele
Istituto di Cibernetica “Edoardo Caianiello”
National Research Council (C.N.R.)
Via Campi Flegrei 34 Bld 70
I-80078 Pozzuoli (Naples) Italy
f.mele@cib.na.cnr.it

Abstract

In day-to-day life it can happen that someone makes people laugh because he/she becomes ridiculous, but without any intention of doing so. We refer to this ridicule situation as “real ridicule”. Otherwise people laugh because someone has a ridiculous behaviour intentionally, for example in a theatrical performance, a film, or some fictional events. We refer to this ridicule situation as “fictional ridicule”.

In the paper some formal models about the real ridicule are discussed. The reported models are based on derision theory proposed in Cousuns and Shoham (1994). In these models derision happens when someone’s behaviour results in *inadequacy in pretending or in doing* in relation to a particular action to be performed. It will be shown that the inadequacy is a necessary condition for a ridicule event, but not a sufficient one. It will be shown that the concepts of *wait* and *surprise* can refine the ridicule model in order to *capture* sufficient conditions.

Some fictional ridicule models, reported in the paper, can be considered as a well characterised subclass of the entire field of the humour phenomena.

It will be shown that Mr. Bean and Groucho Marx in Dylan Dog (Dylan Dog a well known Italian comics edited by Bonelli) humour and several kind of humour in some telecasts are examples of fictional ridicule. Ridicule models will be defined using a formalism of intensional logic Mele (1999); Mele and Minei (2000); Mele et al. (1997) based on four basic mental attitudes Cousuns and Shoham (1994); Shoham (1993): belief (bel), goal (goal), capability (can) and intention (int).

1 REAL RIDICULE

The models that we present for real ridicule are derived from the derision theory proposed in Cousuns and Shoham (1994). In this formulation it is stated that:

“the person is doubly inadequate: inadequate in being or in doing, and inadequate in presuming and expecting. There is contrast between that which was expected and foreseen (surprise, unexpected) and that which actually happens.”

For the real ridicule, various basic components will be provided upon which to build the double inadequacy identified in the previous definition. From now on Pr will refer to the principal character of the ridiculous event and A will refer to the onlooker, who is the person that witnesses the ridiculous event, and to whom the beliefs on the observed events are attributed. For the purpose of having a reference example to which variations will be carried out, let’s consider the following ridiculous situation:

A person Pr is walking through a long alley. An onlooker A observes Pr attentively. At the end of this alley there is a deep puddle of water which impedes Pr to proceed.

(e1) Pr stops, (e2) he observes the puddle of water, (e3) he takes a long pole (similar to that used for pole vaulting in athletics). (e4) Pr takes a run-up with the pole. (e5) Instead of sticking the lower end of the pole into the ground, Pr mistakingly sticks it into the puddle. (er) Pr falls, wetting himself completely. (ey) the onlooker A laughs.

(In the following formalism, we will denote with \leftarrow , \wedge and \neg respectively implication, conjunction and negation. In addition we will denote the variables with capital letters and the constants with small letters).

1.1 MODEL COMPONENTS

1. **bel(A, int(Pr, Ex))**

the onlooker A believes that Pr has the intention of reaching a certain objective Ex

In the above example the intention of Pr is relative to the event $Ex = |Pr \text{ jumps over the puddle of water }|$. In this formulation we will not consider how the onlooker A reaches this belief. We will assume that A observes a series of events e1, e2, e3 which make him believe that Pr has the intention of achieving Ex.

2. **bel(A, goal(Pr, Ex))**

the onlooker A believes that Pr wants to reach a certain objective Ex. This belief can be acquired by A either by inference, according to the base axiom proposed in Rao and Georgeff (1991):

$$\mathbf{bel(A, goal(Pr, Ex))} \leftarrow \mathbf{bel(A, int(Pr, Ex))}$$

or because Pr explicitly manifests it (i.e., he tells A, or he does something similar to an explicit communication). In the latter case the onlooker A is supposed to perceive an event $Ez =_{def} \text{send_message(Pr, A, inform, int(Pr, ex))}$; in other words, a person Pr sends A a declarative type message (inform), containing his intention to achieve Ex.

3. **bel(A, bel(Pr, can(Pr, Ex)))**

A believes that Pr believes that he is capable of reaching Ex.

This component can either be inferred by the onlooker by:

$$\mathbf{bel(A, bel(Pr, can(Pr, Ex)))} \leftarrow \mathbf{bel(A, int(Pr, Ex))}$$

A believes that Pr believes that he is capable of reaching Ex if
A believes that Pr has the intention of reaching Ex

or, as in the previous case, because Pr explicitly manifests it through an event $Ez =_{def} \text{send_message(Pr, A, inform, can(Pr, Ex))}$ where he informs that he has the ability to achieve Ex. To communicate that one has particular abilities is an important element for ridicule in that we believe that it strengthens the effect of ridicule in A once Pr demonstrates the lack of his declared abilities.

4. **bel(A, Ex)**

A believes that the objective E_x will be reached.

5. **bel (A, $E_x \leftarrow (E1 \wedge E2 \wedge \dots \wedge E_n)$)**

A believes that the objective E_x will be reached through the plan $E1 \wedge E2 \wedge \dots \wedge E_n$
(We will write ' $\text{bel(A, } E_x \leftarrow \wedge_i E_i)$ ') instead of ' $\text{bel (A, } E_x \leftarrow (E1 \wedge E2 \wedge \dots \wedge E_n)$ ')')

6. **bel (A, $\text{bel(Pr, } E_x \leftarrow \wedge_i E_i)$)**

A believes that Pr believes that the objective of E_x will be reached by $\wedge_i E_i$

7. **bel(A, $\neg \text{can(Pr, Ex)}$)**

the onlooker A believes that Pr is not capable of reaching Ex

This component is essential for inadequacy. It:

- can be acquired by A before Pr fails to reach Ex. (for example if Pr uses a very short pole (the length of an umbrella for example) so that in no way A can believe that Pr will be able to jump over the puddle: this circumstance is already sufficient to create hilarity in A.
- can be acquired by A after Pr fails to reach the objective Ex.
- may not be acquired by A neither before, nor after (for example if Pr sticks the pole correctly into the ground, but the pole breaks, and Pr still falls into the puddle of water).

1.2 FORMULATION OF THE INADEQUACY

The presumptuous act of being or of wanting to reach an objective is an essential component in a model of the real ridicule process.

We will express the inadequacy of doing by the inability of Pr to reach a particular objective Ex.

$$\mathbf{bel}(A, \mathbf{inadequacy_of_doing}(Pr, E_x)) \leftarrow \mathbf{bel}(A, \neg \mathbf{can}(Pr, E_x))$$

A believes that Pr is inadequate of doing if
A believes that Pr is not capable of reaching E_x .

The inadequacy of Pr, in pretending to want Ex, can be defined by starting with the particular belief of the onlooker A who believes that Pr wants to reach E_x . This inadequacy can be defined as follows:

$$\mathbf{bel}(A, \mathbf{inadequacy_in_pretending}(Pr, E_x)) \leftarrow \\ \mathbf{bel}(A, \mathbf{goal}(Pr, E_x)) \wedge \\ \mathbf{bel}(A, \neg \mathbf{can}(Pr, E_x)).$$

A believes that Pr is inadequate in presuming to want Ex if:
A believes that Pr wants Ex and
A believes that Pr is not capable of reaching Ex.

The inadequacy in presuming of the ridiculous person Pr can be defined by the beliefs of the onlooker A, who observes Pr flaunting his confidence in reaching E_x . From these observations A believes that Pr believes that he is able to reach E_x , but at the same time A (for some other reason) does not believe in the abilities of Pr. Formally we propose the following:

$$\mathbf{bel}(A, \mathbf{inadequacy_in_presuming}(Pr, E_x)) \leftarrow \\ \mathbf{bel}(A, \mathbf{bel}(Pr, \mathbf{can}(Pr, E_x))) \wedge \\ \mathbf{bel}(A, \neg \mathbf{can}(Pr, E_x))$$

A believes that Pr is inadequate in presuming E_x if:
A believes that Pr believes that he can achieve E_x and
A believes that Pr is not capable of reaching E_x

A simple derision model can be developed by using the logical disjunction (non exclusive) of the three inadequacies that have been presented.

1.3 WAITING AND SURPRISE

We believe that the formalisation of waiting is a problem of extreme complexity. In the completed study we have taken into account various types of waiting. Here we will only describe two cases:

wait (A, Pr , $\wedge_i E_i$) \leftarrow
bel (A, **bel**(Pr, $E_x \leftarrow \wedge_i E_i$)) \wedge
bel(A, **goal** (Pr, E_x)).

A waits for the conjunction of events $\wedge_i E_i$ if
A believes that Pr believes that from $\wedge_i E_i$ will follow E_x and
A believes that Pr wants E_x

wait (A, Pr , $\wedge_i E_i$) \leftarrow
bel (A, $E_x \leftarrow \wedge_i E_i$) \wedge
bel(A, **goal** (Pr, E_x)).

A waits for the conjunction of events $\wedge_i E_i$ if
A believes that from $\wedge_i E_i$ will follow E_x and
A believes that Pr wants E_x

The waiting in the first definition is stronger in that it is based on a belief of what Pr believes ‘**bel** (A, **bel**(Pr, $E_x \leftarrow \wedge_i E_i$))’. The waiting in the second definition is instead based on what an agent believes to be the plan to reach E_x . This belief has a weak foundation ‘**bel** (A, **bel**(Pr, $E_x \leftarrow \wedge_i E_i$))’ but it is often formed in agents and it is also the cause of major surprises.

Various types of surprises can be taken into consideration, but here only two types of surprises are described: a surprise created in the agent A due to the fact that an element of the plan does not take place, and a surprise due to the objective E_x of the plan not taking place. The following definition includes both cases:

surprise(A, $\neg (E_x \leftarrow \wedge_i E_i)$) \leftarrow
wait(A, $E_x \leftarrow \wedge_i E_i$) \wedge
bel(A, $\neg (E_x \leftarrow \wedge_i E_i)$)

A is surprised that the plan $E_x \leftarrow \wedge_i E_i$ did not take place if
A waits for the plan $E_x \leftarrow \wedge_i E_i$ and
A believes that the plan $E_x \leftarrow \wedge_i E_i$ did not take place

Therefore the above definition contemplates situations in which:

- the objective E_x is reached by Pr even without having activated the plan $\wedge_i E_i$, that is with an alternative *non conventional* plan. Examples: to comb ones eyebrows with a toothbrush, to switch off the light of the room (without getting up) by shooting the light-bulb, to mount a horse using a hoist etc.;
- A expected a failure (considering the inadequacies of Pr in doing and in being capable) at the objective E_x (implementing a particular plan) and instead Pr creates a surprise by reaching the objective via an alternative plan;
- one of the events of the plan $\wedge_i E_i$ does not take place or even when all the elements of the plan take place but E_x is not achieved, that is, something unexpected happens (as an example, when the pole breaks in the case referred to above).

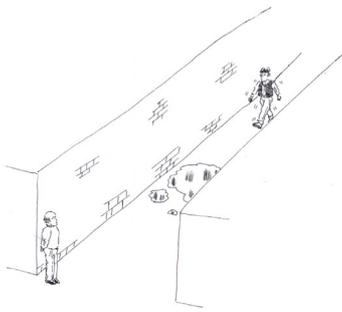
2 RIDICULE MODELS

With the basic components provided, it is possible to construct various components of ridicule. A number of models are presented which lead to ridicule and which are constructed by starting with the components described.

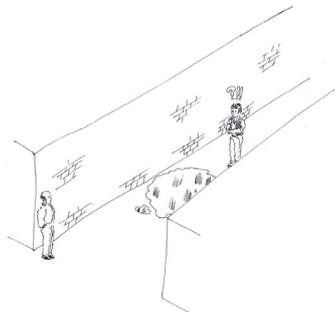
2.1 NO RIDICULOUS EVENT

Before showing some ridicule models, we present a simple case where no ridiculous event happens.

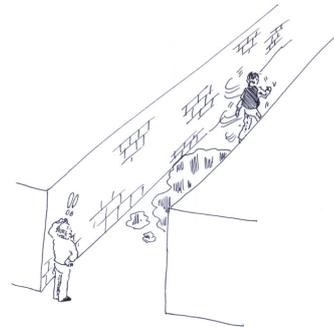
e1	=	Pr stops		A waits
e2	=	Pr observes the puddle of water		Neither surprise nor inadequacy
en	=	Pr makes sure on puddle depth close to		A does not laugh
em	=	Pr decides to come back		



(a) e1



(b) e2

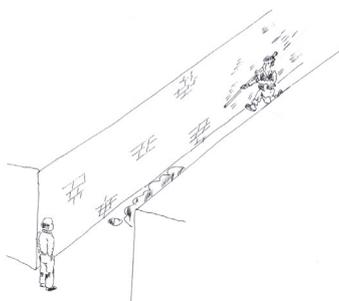


(c) em

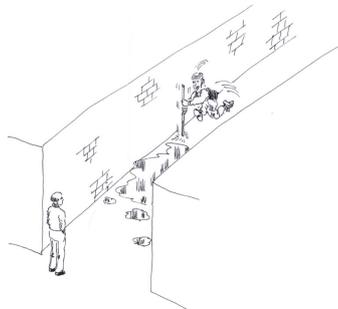
In a first analysis we can say that in such event sequence there is both lack of surprise, and lack of inadequacy, therefore there is the lack of ridicule.

2.2 RIDICULE DUE TO INADEQUACY

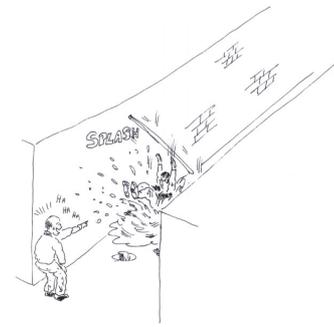
Real ridicule model due to inadequacy is a ridicule frequent case, and it happens in day-to-day life.



(d) e4



(e) e5



(f) e6

e1	=	Pr stops		There is a waiting period: wait (A, Pr, Will Pr jump over the puddle)
e2	=	Pr observes the puddle of water		
e3	=	Pr takes a long pole		Surprise in one of the conditions of the plan Inadequacy (A laughs)
e4	=	Pr takes a run-up with the pole		
e6	=	Pr sticks the pole into the puddle		
er	=	Pr falls, wetting himself completely		

2.3 RIDICULE WITH A DOUBLE WAIT AND A SINGLE SURPRISE

e1	=	Pr stops		First wait (A waits that Pr jumps)
e2	=	Pr observes the puddle of water		
e8	=	Pr takes a short pole		Surprise in one of the conditions of the plan Inadequacy A laughs second wait (A waits that Pr does not jump)
e4	=	Pr takes a run-up with the pole		
e5	=	Pr puts the pool on the dry ground		
ey	=	Pr falls, as the range is short and he wets himself completely		

It is supposed that A has two reasons for laughing: the first reason is due to an *inertia* effect caused by e8 event, the second reason is due to the intrinsic ridicule *ey* event.

2.4 RIDICULE WITH DOUBLE SURPRISE AND DOUBLE WAITING

Now we present a case of ridiculous event that is rare in the day-to-day life although it is a very frequent schema in fictional ridicule.

e1	=	Pr stops		First wait (A waits for Pr to jump)
e2	=	Pr observes the puddle of water		
e9	=	Pr takes an umbrella to use as a pole		Surprise in one of the conditions of the plan Inadequacy , A laughs Second wait, A waits for Pr not to jump A waits for the inadequacy of Pr
e10	=	Pr takes a run-up with the umbrella		
e7	=	Pr hooks onto a bar fixed to the wall, with the handle of the umbrella		
en	=	Pr jumps over the puddle		

In the case of inadequacy another aspect which can be an important component can be added to the model, that is the attempt of Pr to eliminate (reduce) the inadequacy. For example when Pr acts as if he is having fun and he tries to show that he wanted to fall into the water on purpose.

This component is present when Pr knows that he has been ridiculous and with this action he tries (unsuccessfully) to make the onlooker believe that he is in a normal situation.

Finally, we want to point out a last model of ridicule that is present when Pr is not aware of his state of ridicule or ridiculous behaviour. In this model, even if it takes place as frequently as the first, Pr is not aware that he is the object of ridicule. For this reason, it is necessary to add another component of the model that takes place, for example, after an inadequacy, that is:

bel(A , **bel**(Pr , E_x))

A believes that Pr believes that he has reached E_x.

2.5 RIDICULE WITH CONSCIOUSNESS AND HIDDEN INADEQUACY

When inadequacy is present, another aspect, that is an important component of the model, can be considered and it consists in Pr attempting to *hide* the inadequacy. This happens when Pr believes to be ridiculous and he makes the onlooker A believe that he desired and enjoyed falling in the water.

This component is present when Pr is conscious of his ridiculousness and with his action tries (failing) to make someone believe that he is in a normal situation, pleasant too.

2.6 RIDICULE WITHOUT CONSCIOUSNESS

On the contrary of the previous ridicule model, where Pr is conscious of his ridiculous behaviour, in the following model an unconsciousness component is present in the Pr cognitive state. In this model, that is very frequent in day-to-day life, Pr is not aware to be an object of ridicule, and this is the reason for the presence of another ridicule model component. This component is present after that an inadequacy happens, and it is represented by the following conditions:

$$\begin{array}{ll} \text{bel}(A, \text{bel}(\text{Pr}, E_x)) & \text{A believes that Pr believes of having reached } E_x. \\ \text{bel}(A, \neg E_x) & \text{A believes that } E_x \text{ has not been reached} \end{array}$$

2.7 RIDICULE WITHOUT OR WITH WEAK INADEQUACY

Ridicule with weak or without inadequacy needs a wider discussion compared to the previous models.

The considered model includes situations as the following ones:

e1	=	Pr stops		Wait
e2	=	Pr observes the puddle of water		
e3	=	Pr takes a pole		
e4	=	Pr takes a run-up with the pole		
e5	=	Pr puts the pool on the dry ground		
e7	=	The pool breakes		Surprise in the plan
ey	=	Pr falls in the water and he wets himself completely		

It seems that there was not inadequacy on behalf of Pr because the pool does not break due to a Pr incapability. Anyway the onlooker A laughs. In this case A perceives the Pr presumption (in this case we assume presumption). Furthermore, even though Pr falls not voluntary, there is a discrepancy between the goal pursued by Pr and the goal really obtained. If we consider to include in the concept of inadequacy also cases in which Pr is not able to reach E_x , depending or not on his capability, this does not imply a substantial modification of the inadequacy concept. In this last case we simply adopt a concept of effective capability instead of a potential one.

3 FICTIONAL RIDICULE: A SUBSET OF HUMOUR

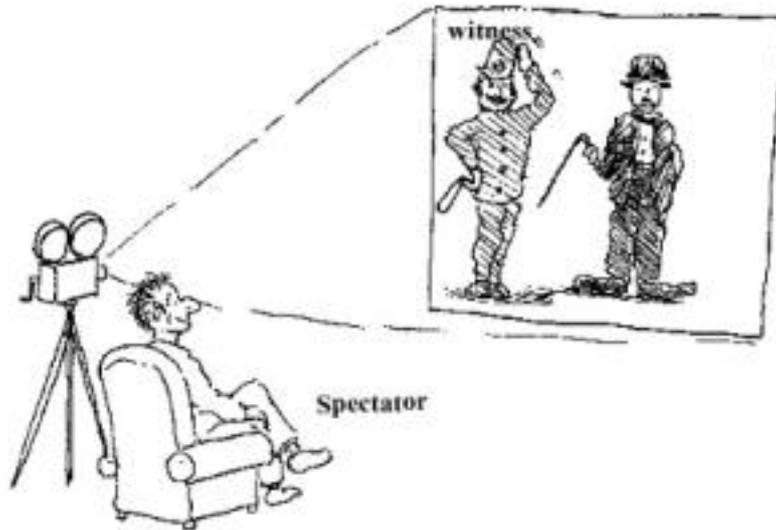
In this section we consider the ridiculous situations that happen in the reality – but represented in a fictional context.

These models, in comparison with more general humour models, such as the two phases ones (inconsistency plus consistency reconstruction) Suls (1976), have the lack of reconstruction although they have the wait and the surprise components. Here we give a characterisation of a subset of phenomena regarding humour that we called fictional ridicule.

3.1 THE FUNCTION OF THE ONLOOKER IN THE REPRESENTATION

Although it is not a necessary condition, the onlooker presence in the representation increases the humour potentialities of the represented ridiculous situations.

The function of a witness, inside a humoristic representation, is to stress the inadequacy of the ridiculous person. The inadequacy is represented by means of a witness embarrassment, or



by means of showing indignation or simply by being silence. In such a way the witness has the function of a *human detector* of inadequacy (visible to the spectator) and the inadequacy itself has the effect to decrease the fruition resistance barrier of the spectator.

So, the witness is a key element for a fictional ridicule model. Till now, the onlooker of the ridicule in real situations takes part into the representation. In the following examples we show that a good ridicule representation can be built when including an onlooker in the representation. The protagonist of the ridicule event pretends with the onlooker to reach a certain goal, also showing some capabilities to be able to do so.

It is worth noting that very rarely the witness laughs.

3.2 MR. BEAN FICTIONAL RIDICULE

Mr. Bean humour is the best example of fictional humour. Mr Bean comic effect has a very large repertory of inadequacies. He often reaches his goals, but he achieves a goal adopting a plane surprising the spectator: he washes his ears with a teeth-brush, he turns off lights with a gun shot to the bulb, and so on. In his humour model Mr. Bean uses very largely a represented witness, that is passive, does not interact with Mr. Bean, is serious and never laughs. (In the figure below, the witness is blind, and the blindness is discovered at the end)



(g) I shot



(h) II shot



(i) III shot



(j) IV shot

In the spectator cognitive state, after the film sequence that ends with the shot I, the following mental attitude arises (according to the component (1)):

(a) $\text{bel}(\text{spectator}, \text{int}(\text{"MisterBean"}, \text{take_off}(\text{trousers})))$.

The (2.1), for values (A::=reader, Pr::= "MisterBean", Ex::= take_off(trousers)), becomes:

(b) $\text{bel}(\text{spectator}, \text{goal}(\text{"MisterBean"}, \text{take_off}(\text{trousers}))) \leftarrow$
 $\text{bel}(\text{spectator}, \text{int}(\text{"MisterBean"}, \text{take_off}(\text{trousers})))$

From (a) and (b) we obtain:

(c) $\text{bel}(\text{spectator}, \text{goal}(\text{"MisterBean"}, \text{take_off}(\text{trousers})))$.

For the considered case we suppose that the plane of (16) $E_x \leftarrow \wedge_i E_i$ ($E_x \leftarrow E_1 \wedge E_2$) believed by the spectator, is:

$E_x ::= \text{take_off}(\text{trousers}) \leftarrow \text{unlace}(\text{belt}) \wedge \text{drop}(\text{trousers})$.

As the following belief holds:

(d) $\text{bel}(\text{spectator}, \text{take_off}(\text{trousers}) \leftarrow \text{unlace}(\text{belt}) \wedge \text{drop}(\text{trousers}))$.

the expression (16) becomes:

(e) $\text{wait}(\text{spectator}, \text{take_off}(\text{trousers}) \leftarrow \text{unlace}(\text{belt}) \wedge \text{drop}(\text{trousers})) \leftarrow$
 $\text{bel}(\text{spectator}, \text{take_off}(\text{trousers}) \leftarrow \text{unlace}(\text{belt}) \wedge \text{drop}(\text{trousers})) \wedge$
 $\text{bel}(\text{spectator}, \text{goal}(\text{"MisterBean"}, \text{take_off}(\text{trousers})))$.

From (c) , (d) and (e) results:

(f) $\text{wait}(\text{spectator}, \text{take_off}(\text{trousers}) \leftarrow \text{unlace}(\text{belt}) \wedge \text{drop}(\text{trousers}))$.

For the sight of shots II and III, the spectator believes that Mr. Bean is not able to pursue the plan (in particular he is not able to drop his trousers) and he makes a revision of his believes that generates a new belief:

(g) $\text{bel}(\text{spectator}, \neg \text{take_off}(\text{trousers}) \leftarrow \text{unlace}(\text{belt}) \wedge \text{drop}(\text{trousers}))$

There is a surprise. In fact, from (17) we obtain:

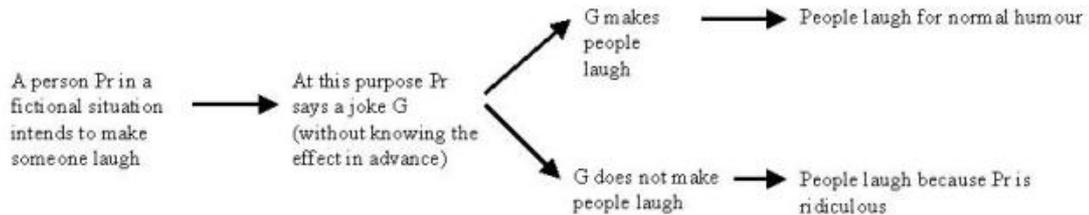
(h) $\text{surprise}(\text{spectator}, \text{take_off}(\text{trousers}) \leftarrow \text{unlace}(\text{belt}) \wedge \text{drop}(\text{trousers})) \leftarrow$
 $\text{wait}(\text{spectator}, \text{take_off}(\text{trousers}) \leftarrow \text{unlace}(\text{belt}) \wedge \text{drop}(\text{trousers})) \wedge$
 $\text{bel}(\text{spectator}, \neg \text{take_off}(\text{trousers}) \leftarrow \text{unlace}(\text{belt}) \wedge \text{drop}(\text{trousers}))$

The surprise is obtained from (h), (g) and (f).

3.3 DOUBLE HUMOUR MODELS

The components of a ridicule model (inadequacy, wait, surprise) are the basis of a more general schema to generate humour.

The model offers two possibilities to make people laugh and for this reason it becomes very *robust*, i.e. it is able to make always people laugh independently from the *quality* of the final joke.



In the humour present in Dylan Dog the main character Groucho assumes to be able to make people laugh.

If his plan is successful, the spectator laughs because the joke makes him laugh. If his plan is unsuccessful, Groucho appears inadequate for some onlooker represented in the scene, therefore the spectator laughs because of the inadequacy of Groucho; in other words the spectator laughs because of the fictional ridicule.

Let's consider the joke (Dylan Dog 27, Pag 64) reported in figure.

The goal of Groucho is to make Dylan Dog laugh: $Ex ::= \text{laugh}(\text{'Dylan Dog'})$



```
e11 = | Groucho says: line 224 |
bel(reader,bel('Groucho',int('Groucho',Ex))
bel(reader,bel('Dylan Dog',bel('Groucho', -can(Groucho',Ex))))
e12 = | Groucho tells the following joke:
      when I was a child my parents
      moved 12 times but I was able to
      find them each time |
e13 = says: Enough please ...
      You're driving me nuts !! |
```

If the joke (e12) makes one laugh we have normal humour. If the joke does not produce laughter we have inadequacy and therefore humour due to the fictional ridicule. The model adopts a witness in the scene (Dylan Dog in e13) to emphasise the inadequacy.

At the end of the fruition process, the cognitive state of the reader is the following:

```
bel(reader,bel('Dylan Dog',-can('Groucho',Ex))
bel(reader,bel('Dylan Dog', inadequacy_in_presuming ('Groucho',Ex))
bel(reader,bel('Dylan Dog', inadequacy_in_pretending ('Groucho',Ex))
```

REFERENCES

- Cousins S. B. and Shoham Y., (1994). Logic of mental Attitudes in AI, *In Lakemeyer G. e Nebel B. (Eds) Foundations of Knowledge Representation and Reasoning*. Springer-Verlag LNAI n.810.
- Castelfranchi C, (1988). *Che figura, Emozioni e immagine sociale*, Il Mulino, Bologna
- Mele F., (1999), *Ridicule and Humour. An intensional logical models for real and represented ridicule*. European Conference of Cognitive Science Pontignano (Siena)
- Mele, F., Minei G. (2000), *Un linguaggio logico temporale per modelli cognitivi*, Atti del convegno nazionale della Mathesis, Teramo 1999
- Mele F., Minei G. Palladino M. (1997), *Magopl: a language for multiagent oriented programming*, Proceedings of the Workshop in Advances in User Modelling, Sixth International Conference on User Modelling (Chia Laguna Sardinia 2-5 June 1997) pp 10-28

Rao A., Georgeff M., (1991). *Modelling Rational Agents within a BDI-Architecture*, KR91, Ed Morgan Kaufman, San, Mateo, CA

Shoham Y. (1993). *Agent Oriented Programming*, Artificial Intelligence, 60:51-92,

Suls J.M.(1976), *L'apprezzamento di barzellette e cartoon in un modello a due fasi: un'analisi del processo informazionale*. In *La psicologia dell'umor* ,Goldstein J.H., McGhee P.E., Franco Angeli Milano

Embodied Agents: A New Impetus to Humor Research

Anton Nijholt
University of Twente
Department of Computer Science
PO Box 217, 7500 AE Enschede
The Netherlands
anijholt@cs.utwente.nl

Abstract

In this paper we survey the role of humor in human-to-human interaction with the aim to see whether it is useful for embodied conversational agents to integrate humor capabilities in their internal model of intelligence, emotions and interaction (verbal and nonverbal) capabilities. For that reason we shortly survey the current state of the art of research in embodied conversational agents, affective computing and verbal and nonverbal interaction. We adhere to the ‘Computers Are Social Actors’ paradigm to assume that human conversational partners of embodied conversational agents assign human properties to these agents, including humor appreciation.

Keywords: Humor, Embodied Conversational Agents, Affective Computing, Nonverbal Communication

1 INTRODUCTION

It is certainly not the case that when we consider research on the role of human characteristics in the user interface of computers no attention has been paid to the role of humor. However, when we compare efforts in this area with efforts and experiments that attempt to demonstrate the positive role of general emotion modeling in the user interface, then we must conclude that the amount of attention is still minimal. As we all know, the computer is sometimes a source of frustration rather than a source of enjoyment. And indeed we see research projects that aim at recognizing a user’s frustration (Klein et al. (1999)), rather than his enjoyment. However, rather than detecting frustration, and maybe reacting to it in a humorous way, we would like to prevent frustration by making interaction with a computer more natural and more enjoyable. For that reason we are working on multimodal interaction and embodied conversational agents. In the interaction with embodied conversational agents, verbal and nonverbal communication are equally important. Multimodal emotion display and detection are among our advanced research issues, and investigations in the role of humor in human-computer interaction is one of them.

Our research is on multimodal interaction with and between embodied conversational agents that inhabit virtual worlds. We envision situations where humans talk to embodied conversational agents in the interface, where maybe it is not at all clear what they represent. Are they completely artificial, autonomous, with built-in intelligence, and put in the environment by their owner or designer? Do they represent a human interactor, including aspects of his personality, beliefs and desires? And is this done ‘off-line’ or even ‘on-line’, where also the interactor’s physical behavior is reflected (not necessarily one-to-one) in the animations of the embodied conversational agent? The user or visitor of an inhabited world may interact with these embodied conversational agents, e.g. to engage in an information service dialogue, a transaction dialogue, to solve a problem cooperatively or to perform a task, or to engage in a virtual meeting. Other obvious applications can be found in the areas of education (including training and simulation), electronic commerce and teleconferencing.

In previous years researchers have discussed the potential role of humor in the interface. However, during these years the potential role of embodied conversational agents was not at all clear, and no attention was paid to their possible role in the interface. Useful observations, also valid when we look at the possible role of embodied conversational agents, were made by Binsted (1995) and Stock (1996). Binsted discusses how humor can make user interfaces friendlier. That is, humans use humor to ease communication problems. In a similar way humor can be used to solve communication problems that arise with human-computer interaction using natural language interfaces. Binsted explains that the kinds of humor to be used do not have to be very sophisticated. Suitable humor that can be used is self-deprecating humor. In some cases deprecating the user or a third party can be appropriate, but this type of humor is very risky. Humor can make a computer more human when it fails and can ease the interaction. Inappropriate humor, however, is irritating and humor should be tailored to the user. When a certain user regularly works with a system, the system can adapt the use of humor to the user's taste. She concludes that humor that is sparingly and carefully used can make natural language interfaces much friendlier. However, there are not that many applications. There are exceptions, for example, Loehr (1999) discusses the use of humor in a natural language robot in the interface, some researchers have been working on adding humor, or rather humorous remarks, to (primitive) embodied agents in the interface or explanation facilities, but these attempts are rather pre-canned and hardly take into account an appraisal of the events that occur or have recently occurred in their world.

About this Paper: In the next section (section 2) we want to look at the role of humor in human-human interaction. We discuss some general issues concerning trust and interpersonal attraction and then, using the CASA paradigm (section 3), start discussing the role of embodied conversational agents in human computer interaction (section 4), including multimodal emotion display, and how we can display appreciation of humor. Section 5 has some observations about feigned and felt emotions when an embodied conversational agent displays them. Section 6 is about the appraisal of events in order to decide that the event is humorous. Section 7 contains the conclusions of this paper.

2 THE ROLE OF HUMOR IN INTERPERSONAL INTERACTION

In interpersonal interactions, either at work or at home, humans use humor, humans smile and humans laugh. Humor can be spontaneous, but it can also serve a social role and be used deliberately. A smile can be the effect of appreciating a humorous event, but it can also be used to regulate the conversation. A laugh can be spontaneous but can also mask disagreement or be cynical. Research has shown that laughs are related to topic shifts in a conversation and phases in negotiations or problem solving tasks. In an educational situation humor can be used by the teacher to catch students' attention but also to foster critical thinking. Humor allows criticism to be smoothed, stress can be relieved and students can become more involved in joint classroom activities by the use of humor. Humor can also be the right answer to frustration. In an (E-)commerce situation we have negotiators that use humor to induce trust.

In this section we want to look at the role of humor in human-human interaction. We survey some results from experimental research. Section 2.1 is devoted to three more general issues, not necessarily connected to a particular domain, but playing a role in human-human interaction: trust, interpersonal attraction and humor support in a conversation. More topics could have been chosen, but some of these arise naturally when in section 2.2 we discuss some domains for which we may expect that in the near future embodied conversational agents can play the roles of one or more of the conversational partners in the current real-life situations. The domains we choose are education, information services and commerce, meetings, and negotiations. The role of humor in these domains is discussed by taking examples from the literature.

2.1 GENERAL ISSUES: SUPPORT, TRUST, AND ATTRACTION

In this paper we are not particularly interested in a possible preconceived aim of a conversational partner to create humor during a conversation or discussion. Rather we look at situations where humor occurs spontaneously during an interaction or where it occurs in a supporting role, for

example to hide embarrassment, to dominate the discussion or to change the topic. Some of these roles will get more attention in section 2.2. Here we have some remarks on humor to induce trust, interpersonal attraction and how to show appreciation of humor during a conversation.

Humor support, or the reaction to humor is an important aspect of personal interaction. We employ a wide range of humor in our conversations and the given support shows the understanding and appreciation of humor. In Hay (2001) it is pointed out that there are many different support strategies. The strategy that can be used in a certain situation is mainly determined by the context of the humorous event. The strategy can include smiles and laughter, the contribution of more humor, echoing the humor, offering sympathy or contradicting self-deprecating humor. There are also situations in which no support is necessary. In order to give full humor support, humor has to be recognized, understood and appreciated. These factors determine our level of agreement on a humorous event and the grade of how we want to support the humor. This paper is in fact a discussion on the possible role of humor support in the context of the design and implementation of embodied conversational agents

Support may show our involvement in the discussion, our motivation to continue and how much we enjoy the conversation or interaction. Similarity in appreciation also supports interpersonal attraction (Cann et al (1997)). This observation is of interest when later we discuss the use of embodied conversational agents in user interfaces. Sense of humor is generally considered a highly valued characteristic of self and others. Nearly everybody claims to have average to above average senses of humor. Perceived similarity in humor appreciation can therefore be an important dimension when designing for interpersonal attraction. In experiments reported by Cann et al (1997) participants had to interact with an unseen stranger. Before the interaction ratings were made of the attitudes of the participants and they were led to believe that the stranger had similar or dissimilar attitudes. The stranger responded either positively or neutrally to a participant's attempt to humor. As a main result it was shown that similarity in humor appreciation was able to negate the negative effects of dissimilarity for other attitudes when looking at interpersonal attraction. In our group we have studied how similarity in attitudes is related to the development of a friendship relationship. The development of a friendship relationship requires time, but especially in the initiation phase the kinds of similarities mentioned above can be exploited. A discussion on friendship in the context of the design and implementation of embodied conversational agents has been presented in Stronks (2002).

Friendship and intimacy are closely related. Hampes (1994) discussed the relation between intimacy and humor and in Hampes (1999) the relation between trust and humor. Trust is an essential aspect of intimacy and the hypothesis that there also exists a correlation between humor and trust was confirmed. There are three key-factors that help us to understand this relationship. The most important factor is the demonstrated relation between humor and extroversion (Ruch (1994)). When we break up extroversion into basic components like warmth, gregariousness, assertiveness and positive emotions it becomes obvious that extroversion involves trust. Another factor, mentioned above, is the fact that humor is closely related to a high self-esteem. People who are proud of who they are, are more likely to trust other persons and to reveal themselves to them. A third factor is that humorous persons are effective in dealing with stress (Fry (1995)). They are well qualified to deal with the stress or anxieties involved in interpersonal relationships and therefore are more willing to enter relationships. A discussion on trust in the context of the design and implementation of embodied conversational agents has been presented in Bickmore et al (2001).

2.2 CONVERSATIONS AND GOAL-DIRECTED DIALOGUES

Humor plays a role in daily conversations. People smile and laugh, certainly not necessarily because someone pursues the goal of being funny or tells a joke, but because the conversational partners recognize the possibility to make a funny remark fully deliberately, fully spontaneously, or something in between, taking into account social (display) rules. We will not go deeply into the role of humor in daily conversations, small talk or in entertainment situations. In daily conversations humor very often plays a social role. The role of humor is hardly discussed in

Deborah Tannen's well known "That's Not What I Meant!" (about "How Conversational Style Makes or Breaks Relationships"), probably simply because the book is not about conversations going well, but about conversations going wrong. Some issues related to the role of humor have been mentioned in the previous section. It is difficult to design experiments intended to find the role played by humor in human-to-human interactions, when no specific goals are defined. Even experiments related to rather straightforward business-to-consumer relationships are difficult to find. Rather we have to deal in these situations with regulations protecting a customer from humor by a salesman (never use sarcasm, don't make jokes on account of the customer, etc.).

We continue this section by discussing the role that humor plays in some domains of interaction. The discussion is certainly not exhaustive and neither is the choice of domains. It is meant to show the importance of the role of humor and therefore the need to consider the role of humor in situations where we try to replace one or more of the interactors in (multi-party) conversations or dialogues with embodied conversational agents. The examples we mention are task-related. For example, the teacher wants to teach a student, someone wants to get information about theatre performances, I want to buy a pair of shoes. As another example, in a meeting we want to exchange information, solve a problem or negotiate with the other participants. In section 4 of this paper we will mention some examples of conversational situations, (not necessarily commercial) information service situations and business-to-consumer situations that have been modeled using embodied conversational agents (among others, a virtual receptionist, a virtual tourist guide, a car saleswoman, a female bartender). However, the (possible) role of humor in these situations is rather restricted and hardly any comparisons have been made with the real-life situations.

When we look at more goal-directed situations, teaching seems to be one field where the use of humor in the teaching process has received reasonable attention. What is the role of humor as a teaching tool? Many benefits have been mentioned and sometimes made explicit in experiments. Among them are motivating effects on attention, promotion of comprehension and retention of information, more pleasurable learning experience, development of affective feelings toward content, fostering of creative thinking, reducing anxiety, etc. The role of humor during instruction, its social and affective functions for teaching and implications for classroom practice has been discussed in several papers (see Price (-) for an example). Despite the many experiments, it seems to be hard to generalize from the experiments that are conducted Ziv (1988).

There are more interesting examples of studies about the role of humor in interaction processes. For example, the role of humor and laughter during negotiation processes is another issue that has received attention. In Adelswaard (1998) several tape recordings made during international negotiations have been analyzed. One of their research questions concerned the interactional position of laughter: When do we laugh during interaction? Different phases during negotiation can be distinguished. Laughing events turned out to be related to the phase boundaries and also to discourse boundaries (topic shifts). Hence, laughter serves interactional goals. The distinction between unilateral and joint laughter is also important. Mutual laughter often reflects consensus, unilateral laughter often serves the same function as intonation. Moreover, this distinction is related to social issues. Who is dominant in the negotiation, who has the advantage, and for what reasons? Attitudes towards topics discussed are also reflected in the laughing events: which topics are important, sensitive or face threatening?

Describing and explaining the occurrence of humor in small task-oriented meetings is the topic of a research study conducted by Consalvo (1989). The study gives good insight in humor's important role within management meetings. A humorous event was defined as the situation in which at least one of the participants laughed. An interesting and unforeseen finding was the patterned occurrence of laughter associated with the different phases of the meeting. The opening phase is characterized by its stiffness and serious tone and the atmosphere of distrust. Humor in this phase is infrequent. This is in contrast with the second, transitional phase that lasts only a couple of minutes and the humorous interactions are frequent and for the first time during the meeting all participants laugh. Their laughter conveys the agreement that the problem can be solved and the commitment of the individual participants. The last phase, the problem-solving phase contains a lot more humorous events than the opening phase, but still less than the transitional phase. Humor echoes the progression of a meeting and can be both constructive or destructive. It can

be an important tool for management and leadership, because appropriate humor can smoothen the task-oriented processes through cultivating an environment where freedom and flexibility will flourish.

Any reader will know about other examples of interactions that involve the role of humor. What about a tourist guide, explaining the history of a cathedral, a bartender that soothes a tense situation, or a car saleswoman who knows how to handle sexual harassment by using humor?

3 COMPUTERS AS SOCIAL ACTORS

In the research on the ‘computers are social actors’ (CASA) paradigm (see e.g. Reeves and Nass (1996)) it has been convincingly demonstrated that people interact with computers as if they were social actors. Due to the way we can let a computer interact, people may find the computer polite, dominant, extrovert, introvert, or whatever attitudes or personality (traits) we can display in a computer. Moreover, they react to these attitudes and traits as if a human being displayed them. As an example, consider the situation where a person interacts with the computer in order to perform a certain task. When, after completing the task, the person is asked by the same computer about its (i.e., the computer’s) behavior, the user is much more positive than when asked this question while sitting behind an other computer. From these CASA experiments we conclude that it is possible, at least in principle, to design systems that are perceived as social actors and that can display characteristics that elicit positive feelings about an interaction, even though the interaction is not considered as perfect from the user’s point of view.

As mentioned above, humor plays an important role in interpersonal interactions. And so do smiles. We will return to the role of smiles later. Will humor in the interface have similar effects as in interpersonal interactions? In Morkes et al. (1998), experiments are reported that have been performed to examine the effects of humor in task-oriented computer-mediated communication and in human-computer interaction. It was shown that humor could have many positive effects. For example, participants who had received jokes during the interaction rated a system as more likable and competent. They smiled and laughed more, they responded in a more sociable manner and reported greater cooperation. The study provides strong evidence that humor should be incorporated in computer mediated communication and human-computer interaction systems.



Figure 1: Examples of 2D and 3D embodied agents

4 EMBODIED CONVERSATIONAL AGENTS

4.1 DEVELOPMENTS IN EMBODIED CONVERSATIONAL AGENTS

In our research on natural interactivity between humans and computers we adhere to the CASA paradigm. Embodied conversational agents (ECAs) have become a well-established research area. Embodied agents are agents that are visible in the interface as animated cartoon characters or animated objects resembling human beings. Sometimes they just consist of an animated talking face, displaying facial expressions and, when using speech synthesis, having lip synchronization. These agents are used to inform and explain or even to demonstrate products or sequences of activities in educational, e-commerce or entertainment settings. Experiments have shown that ECAs can increase the motivation of a student or a user interacting with the system. In Lester et al. (1997) showed that a display of involvement by an embodied conversational agent motivates a student in doing (and continuing) his or her learning task. Some examples of embodied conversational

agents are shown in Figure 1. From left to right we see: Jennifer James, a car saleswoman who attempts to build relationships of affection, trust and loyalty with her customers, Karin, informing about theatre performances and selling tickets, Layla, a virtual receptionist, Carmen, a mother undergoing therapy, and Steve, educating a student.

Current research deals with improving intelligent behavior of these ECAs, but also with adding emotional behavior and personality. Improving intelligent behavior requires using techniques from artificial intelligence, in particular natural language processing. Domain knowledge and reasoning capabilities have to be modeled. Agent models have been developed that allow separation between the beliefs, desires and intentions of an agent. For systems consisting of multiple agents formal communication languages have been developed that allow the transfer of information from one agent to another. Together with dialogue modeling techniques rudimentary natural language interaction with such agents is becoming possible. Speech input remains difficult to realize, speech output can be of acceptable quality.

Despite improvements we cannot expect that in the near future these agents can match human capabilities. In Bates (-) the notion of ‘believability’ was introduced in this field of research. Believability is present in an ECA when despite a lack of realism communication using a multiple of modalities leads to a suspension of disbelief by the human conversational partner. ‘Trust’ in an embodied conversational agent is a related issue. Do we trust an agent that plays the role of a doctor or a salesperson? The role of small talk for, among others, inducing trust in an embodied real estate agent is discussed in Bickmore et al (2001). It is an example of sometimes subtle capabilities that help to improve the believability of an embodied agent. Clearly, humor is very much related to many of the natural interaction issues mentioned above: emotions, personality, attraction, and trust. Until now we have not seen much research going on into embodied agents that interpret or generate humor in the interface. Nevertheless we see attempts to maintain an emotional state and generate facial expressions and other display of emotions from such an emotional state. This may cause an agent to smile at an appropriate moment.

4.2 NONVERBAL AND AFFECTIVE INTERACTION IN EMBODIED AGENTS

An embodied agent has a face. It may have a body, arms, hands and legs. We can give it rudimentary intelligence and capabilities to have verbal and nonverbal interaction. Nonverbal signals come from facial expressions, gaze behavior, eyebrow movements, gestures, body posture, and head and body movements. Nonverbal signals can also be made available in the voice of an ECA. Communicative behavior is dependent on the personality that has been modeled in an ECA.

In previous years we have seen the emergence of affective computing. Although many research results on modeling of affect are available, it is certainly not the case that a comprehensive theory of affect modeling is available. Reasons to include emotion modeling in intelligent systems are, among others, to enable decision-making in situations where it is difficult, if not impossible, to make rational decisions, to afford recognition of a user’s emotions in order to give better and more natural feedback, and to provide display of emotions. Especially when the interface includes an embodied conversational agent, it seems rather obvious that the user expects a display of emotions and some recognition of emotions by the embodied agent. On the other hand, in order to improve the interaction performance of embodied agents they should integrate and use multimodal information obtained from their human conversational partner. Although measurement techniques and technology are becoming available to detect multimodal displayed emotions in human interactants (cameras, microphones, eye and head trackers, expression glasses, face sensors, movement sensors, pressure sensitive devices, haptic devices and physiological sensors) here we will not discuss the display of humor emotions, e.g. enjoyment, by the human interactant (and preferably be perceived by an ECA). Obviously, although useful, we would rather have an ECA understand why the events that take place generate enjoyment by its conversational partner and why the ECA itself should display enjoyment because of its appreciation of the humorous situation. Display of enjoyment should follow from some emotional state that has been computed from sensory inputs of a human interactant, but in the first place from an appraisal of the events that happen or have happened simultaneously or very recently. A usual standpoint is that of appraisal theory, the evaluation of

situations and categorizing arising affective states.

4.3 DISPLAYING HUMOR APPRECIATION IN THE FACE AND IN THE VOICE

In Cowie (2000), when discussing the display of emotions in speech, Roddy Cowie mentioned, “A major topic is raised here for want of a better place. It is humour. Humour appears to have strong links to both control and emotional mixture. It may express anger or bleakness or happiness, and our explorations suggest that it is very often used as the preferred way of signalling these emotions without violating display rules. A useful way of making the point is in terms of artificial agents. If they are going to show emotion, we surely hope that they would show a little humour too.” Facial expressions and speech are the primary sources for obtaining information of the affective state of an interactant. So, we conclude that an embodied conversational agent first of all needs to display emotions and humor appreciation through facial expressions and the voice.

In speech, emotion (or emotion changes) can be detected by looking at deviations from personal, habitual vocal settings of a speaker because of emotional arousal. Cues come from loudness, pitch, vibrato, precision of articulation, etc. See e.g. Kappes et al. (1991) for observations, including a discussion of cues that are related to detecting enjoyment in the voice. The vocalization of laughter is another interesting issue for embodied conversational agents.

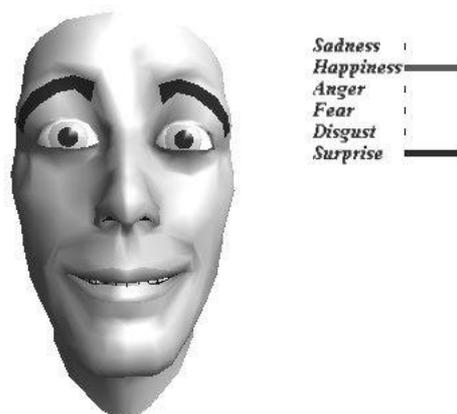


Figure 2: Blending of happiness and surprise

To describe emotions and their visible facial actions, facial (movement) coding systems have been introduced. In these systems facial units have been selected to make up configurations of muscle groups associated with particular emotions. The timing of facial actions has also been described. Using these systems, the relation between emotions and facial movements can be studied. For example, it can be described how emotion representations can be mapped on the contraction levels of facial muscle configurations. Modalities in the face that show affect also include movements of lips, eyebrows, color changes in the face, eye movement and blinking rate. Cues combine into expressions of anger, into smiles, grimaces or frowns, into yawns, jaw-droop, etc. Happiness, for example, may show in increasing blinking rate. Obviously, when using a talking face, a designer can deliberately put emphasis on particular facial actions during interaction. In Figure 2 we display a face from our own research that uses a fuzzy-rule based system to map an emotional state on muscle contraction values (Duy et al. (2001)).

Smiles and laughs are the usual responses to humor. As mentioned in previous sections, laughs and smiles also serve social and (discourse) interactional goals. In Eckman (1985) eighteen different smiles are distinguished. A smile can be a greeting; it can mean incredulity, affection, embarrassment or discomfort, to mention a few. Smiling does not always accompany positive feelings. That makes it important to be able to display the right kinds of smiles at the right time on the face of an embodied conversational agent. Frank and Ekman (1993) discuss the ‘enjoyment’ smile, the particular type of smile that accompanies happiness, pleasure, or enjoyment. The facial

movements that are involved in this smile are involuntary, they originate from other parts in the brain than the voluntary movements and have a different manifestation. Morphological and dynamic markers have been found to distinguish enjoyment smiles from others. The main, best validated marker is known as the Duchenne marker or Duchenne's smile, the presence of orbicular oculi action (the muscle surrounding the eyes) in conjunction with zygomatic major action (muscles on both sides of the face that pull up the mouth corners). The Duchenne smile can reliably be identified visually. Although some people can produce it consciously, it probably is the best facial cue for deciding enjoyment and therefore an embodied conversational agent should show it in the case of sharing humorous events with its human partner. For a survey of hypotheses and empirical findings regarding the involvement of muscles in the laughter facial expression see Ruch and Ekman (2001). Laughter also involves changes in posture and body movements. Again, we need to distinguish between different types of laughter (spontaneous, social, suppressed).

5 FEIGNED OR FELT?

In applications using embodied conversational agents we have to decide which smiles and laughs to use while interacting with a human conversational partner. When a virtual teacher smiles, should it be a Duchenne smile? Is the embodied agent 'really' amused or does it only display a polite smile because it does not really like the joke made by its human conversational partner. Or should it not laugh or smile at all because of this politically incorrect joke? As mentioned by Cowie (2000): "People respond negatively to displays of emotion that are perceived as simulated, and that is a real issue for agents that are intended to convey emotion." Will our attempts to introduce believability not be hampered by the impossibility to convey emotions in a believable way? Maybe we accept poor quality speech synthesis, maybe we accept poor quality facial expression (compared with human speech and human facial expressions), but will we accept the same for emotion display, in particular display related to an appreciation of a humorous event conveyed through these channels? Note, that when we talk about a humorous event, we include events that appear in a story being told by a virtual agent in interaction with a human conversational partner, events that are interpreted from a sequence of utterances in a dialogue, events that are visualized in a virtual environment, or events that need interpretation of integrated virtual and real-life interactions. These are interesting issues, but in our view not different from other observations on believability of embodied agents. In some situations, assuming that quality allows it, a synthesized voice or face may express acted pleasure (or anger), in other situations genuine pleasure (or anger). Whether it sounds or looks sincere depends on being able to suspend disbelief in the human partner of the agent. Interesting in this respect is the work of Marquis and Elliott (1994) who discuss research on embodied poker-playing agents (with a human partner) that can display, based on the OCC model deliberately display false emotions in the face and in the voice.

6 APPRAISAL OF HUMOROUS EVENTS

It is clear from the observations we made in the previous sections that there is a need for models that allow generation, prediction, detection and interpretation of humorous events. There is also a need to be able to generate, display and interpret smiles and laughs in a context that is not necessarily found humorous by an embodied conversational agent, but that nevertheless may lead to smiles and laughs, for example to keep a conversation going. We think that it seems to be quite natural to make a step from event appraisal theories for emotion to appraisal theories for humorous events, in order to obtain embodied conversational agents that smile or laugh at the right moment, making them more believable. Some of these theories have been designed with computation in mind. How can we elicit and display emotions using a computational model? One rather mature theory for calculating cognitive aspects of emotions is the OCC model Ortony et al. (1988), a framework of 22 distinct emotion types. In later years Ortony (2001) it was suggested to collapse this scheme to five distinct positive and five distinct negative affective reactions, under the assumption that this should be sufficient for building believable affective agents ("with the potential for a rich and varied emotional life"). In several (mostly, stripped-

down) versions, the model has been used. E.g., in the OZ-project Reilly and Bates (1992), which is concerned with the development of a theatre world inhabited by emotional agents. In the Carmen project (Marsella et al. (2000)) event appraisal is used to recognize and process feelings of guilt and anger in a setting where an embodied conversational agent talks with a mother of children with leukemia.

We think it is useful to review existing theories and observations concerning the appraisal of (humorous) situations (available as events, in conversation, in verbal descriptions or stories) in terms of possible agent models that include explicit modules for beliefs, desires, intentions and emotions. For example, it would be worthwhile to investigate how such a model can include reasoning mechanisms about situations where there is the feeling that the situation is normal, while at the same time there is a violation of a certain commitment of the agent about how things ought to be. With this view in mind it seems useful to look at the violation theory discussed in Veatch (1998), attempts to define degrees of incongruity Deckers (2001), attempts to define humor in terms of violations of Grice's conversational maxims (Attardo (1993)), proposals to define and explain humor or laughter in terms of perceptual, affective, and response patterns (Russel (2000)), but also 'measures' made possible by a Situational Humor Response Questionnaire (measures the propensity to laugh in a variety of verbally described situations). Annotated verbally described situations could be the basis for (supervised) learning of humorous situations similar as our model of agent emotions learning discussed in Poel et al. (2002).

7 CONCLUSIONS

In this paper we were able to touch upon the state of the art of embodied conversational agents, humorous interfaces and affective computing. We think that from our observations it may become clear that current research on affective computing, research on generating and interpreting facial expressions and research on embodied (and intelligent) agents can and should be combined with humor research. This is for the benefit of humor research and results can help to design new and interesting applications in human-computer interaction using embodied agents.

ACKNOWLEDGEMENTS

This paper was written while we took part in the HAHAcronym project (Humorous Agent for Humorous Acronyms). This is a European IST programme (IST-2000-30039) on Future and Emerging Technologies, done jointly with Istituto Trentino di Cultura (ITC-IRST) in Trento, Italy. Part of the material in this paper was taken from a deliverable we wrote for that project (Mulder and Nijholt (2001)). I am grateful to Matthijs Mulder for his help in finding relevant literature and to Lynn Packwood for proofreading the paper.

REFERENCES

- V. Adelswrd B.-M. Öberg. The function of laughter and joking in negotiation activities. *Humor. J. of Humor Research* (11) 4, 1998, 411-429.
- S. Attardo. Violations of conversational maxims and cooperation: The case of jokes. *Journal of Pragmatics* 19 (1), 1993, 537-558.
- J. Bates. Virtual reality, art, entertainment. *Presence: Teleoperators and Virtual Environments* 1 (1), 133-138.
- T. Bickmore J. Cassell. Relational agents: A model and implementation of building user trust. *Proc. ACM CHI 2001*, Seattle, Washington, 2001.
- N. Bischof. The varieties of smiling and the problem of motivational acclimatization. Abstract. In: *Proceedings 8 European Conference on Facial Expression – Measurement and Meaning*, University of the Saarland, Saarbrücken, Germany, 1999.

- K. Binsted. Using humour to make natural language interfaces more friendly. In: Proceedings of the AI, ALife and Entertainment Workshop, Intern. Joint Conf. on Artificial Intelligence, 1995.
- Bui The Duy, D. Heylen, M. Poel A. Nijholt. Generation of facial expressions from emotion using a fuzzy rule based system. In: Proc. 14th Australian J. Conf. on Artificial Intelligence (AI 2001), Adelaide, Australia, LNAI 2256, M. Stumptner, D. Corbett M. Brooks (eds.), Springer, Berlin, 2001, 83- 94.
- A. Cann, L.G. Calhoun J.S. Banks. On the role of humor appreciation in interpersonal attraction: It's no joking matter. *Humor. J. of Humor Research* (10) 1, 1997, 77-89.
- C.M. Consalvo. Humor in management: no laughing matter. *Humor. J. of Humor Research* (2) 3, 1989, 285-297.
- R. Cowie. Describing the emotional states expressed in speech. Proceedings of the ISCA Workshop on Speech and Emotion. Belfast, September 2000, 11-18.
- R. Cowie et al. (eds.). Speech and Emotion. A conceptual framework for research. Proc. of the ISCA Workshop, Belfast, September 2000.
- L. Deckers. On the validity of a weight-judging paradigm for the study of humor. *Humor. J. of Humor Research* (6) 1, 2001, 43-56.
- P. Ekman. Facial expressions and emotion. *American Psychologist*, Vol. 48, No. 4, 384-392.
- P. Ekman. Telling Lies. W.W. Norton, New York, 1985.
- C. Elliott. Why boys like motorcycles: using emotion theory to find structure in humorous stories. Unpublished paper, School of Computer Science, DePaul University, Chicago, 1999.
- M.G. Frank P. Ekman. Not all smiles are created equal: the differences between enjoyment and nonenjoyment smiles. *Humor. Intern. J. of Humor Research* (6) 1, 1993, 9-26.
- P.S. Fry. Perfection, humor, and optimism as moderators of health outcomes and determinants of coping styles of women executives. In: Genetic, Social, and General Psychology Monographs, Vol.121, 1995, 211-245.
- W.P. Hampes. Relation between intimacy and the Multidimensional Sense of Humor In: Psychological Reports, Vol. 74, 1994, 1360-1362.
- W.P. Hampes. The relationship between humor and trust. *Humor. J. of Humor Research* (12) 3, 1999, 253-259.
- J. Hay. The pragmatics of humor support. *Humor. J. of Humor Research* (14) 1, 2001, 55-82.
- J. Hulstijn A. Nijholt (eds.). Computational Humor: Interpretation and generation of verbal humor. Proc. Twente Workshop on Language Technology 12, Un. of Twente, Netherlands, 1996.
- A. Kappas, U. Hess K.R. Scherer. Voice and emotion. In: Fundamentals of nonverbal behavior. R.S. Feldman et al. (eds.), Cambridge University Press, 1991, 200-237.
- J. Klein, Y. Moon R.W. Picard. This computer responds to user frustration. Theory, design, results, and implications. MIT Media Lab, 1999.
- J.C. Lester, S.A. Converse, S.E. Kahler, S.T. Barlowe, B.A. Stone R. Bhogal. The persona effect: Affective impact of animated pedagogical agents. Proc. CHI '97 Human Factors in Computing Systems, ACM, 359-356.
- D. Loehr. An integration of a pun generator with a natural language robot. In: Proc. Intern. Workshop on Computational Humor, J. Hulstijn A. Nijholt (eds.), University of Twente, Netherlands, 1996.

- S. Marquis C. Elliott. Emotionally responsive poker playing agents. In: Notes for the Twelfth National Conference on Artificial Intelligence (AAAI-94) Workshop on Artificial Intelligence, Artificial Life, and Entertainment, 11-15. American Association for Artificial Intelligence.
- S.C. Marsella, W. Lewis Johnson C. LaBore. Interactive pedagogical drama. In: Proceedings of 4th International Conference on Autonomous Agents 2000, ACM Press, 301-308.
- J. Morkes, H.K. Kernal, and C. Nass. Humor in task-oriented computer-mediated communication and human-computer interaction. In: Proceedings of CHI 98, 1998.
- M.P. Mulder A. Nijholt. Humour research: State of the art. Deliverable D1, IST Programme on Future and Emerging Technologies. Action Line: IST-2000-6.1.1, HAHAcronym: Humorous Agent for Humorous Acronyms, October 2001.
- A. Nijholt. Agents, Believability and Embodiment in Advanced Learning Environments. In: Proc. IEEE International Conf. on Advanced Learning Technologies (ICALT 2001), T. Okamoto, R. Hartley, Kinshuk J.P. Klus (eds.), Madison, Wisconsin, IEEE Computer Society, 457-459.
- A. Ortony. G.L. Clore A. Collins. The Cognitive Structure of Emotions. Cambridge University Press, 1988.
- A. Ortony. On making believable emotional agents believable. In: Emotions in humans and artifacts. R. Trappl P. Petta (eds.), MIT Press, Cambridge, 2001.
- M. Poel, R. op den Akker, A. Nijholt A.J. van Kesteren. Learning emotions in virtual environments. Proc. EMCSR 2002 Symp. Agent Construction and Emotions, Vienna, Austria, April 2002, to appear
- K. Price. Humour and I.T. You must be jokin. WorldWide Web, Manuscript.
- B. Reeves C. Nass. The Media Equation: how people treat computers, televisions and new media like real people and places. Cambridge, Cambridge University Press, 1996.
- W. Reilly, J. Bates. Building emotional agents. Report CMU-CS-92-143, Carnegie Mellon University, 1992.
- W. Ruch. Extroversion, alcohol, and enjoyment. *Personality and Individual Differences* (16), 1994, 89-102.
- W. Ruch P. Ekman. The expressive pattern of laughter. In: Emotion, qualia, and consciousness. A.W. Kaszniak (Ed.), Word Scientific Publisher, Tokyo, 2001, 426-443.
- R.E. Russell. Humor's close relatives. *Humor. J. of Humor Research* (13) 2, 2000, 219-233.
- O. Stock. Password Swordfish: Verbal Humour in the Interface. In: Proc. Intern. Workshop on Computational Humor, J. Hulstijn A. Nijholt (eds.), University of Twente, Netherlands, 1996.
- J.J.S. Stronks. Friendship relations with embodied conversational agents: Integrating psychology in ECA research. M.Sc. Thesis, Parlevink research Group, University of Twente, January 2002.
- T. Veatch. A theory of humor. *Humor. J. of Humor Research* (11) 2, 1998, 161-215.
- A. Ziv. Teaching and learning with humor: Experiment and replication. *J. of Experimental Education* 57 (1), 1988, 5-15.

A Formalism for Misapprehended Identities: Taking a Leaf Out of Pirandello

Ephraim Nissan

School of Computing and Mathematical Sciences, University of Greenwich,
Queen Mary Court, Maritime Greenwich Campus, 30 Park Row
(Old Royal Naval College), Greenwich, London SE10 9LS, England, U.K.
E.Nissan@greenwich.ac.uk <http://www.gre.ac.uk/~E.Nissan> & E.Nissan/CmsNlp

Abstract

Epistemic and behavioural inadequacy do not make for humour by themselves, yet it often involves them. The ‘madman’ theme is prominent in jokes. Non-joke humorous, or even in the main not-humour texts, negotiate a maze of factors, from which humour may emerge. I formalize some beliefs from Pirandello’s *Henry IV* (1922), a tragedy whose protagonist is thought (by pragmatic rationality) to (delusionally) believe he is the medieval emperor of Germany at a time when he has recovered. Different characters are faced with problems rooted in having been schooled or not in medieval history. For example, some women hired to be “Berta of Susa” (Henry IV’s wife), allegedly each myrthfully announced herself to the deluded protagonist as being Berta and as harking back from Susa (a town in Piedmont). Apart from the protagonist epistemic inadequacy, they are laughing because of reasons different from those of his social peers, or of the servants who impersonate his vassals. My formalism (part of a broader project on capturing identity ascriptions) tries to put some order in the quagmire.

1 ELEMENTS FROM A PIRANDELLO PLAY

For all of the difficulty of capturing verbal jokes in AI models (Hulstijn and Nijholt, 1996), taking on non-joke humorous text (or on humour in some non-humour text) is by far a more daunting task (Attardo 1996; Nissan, COLUMBUS, in press). Raskin (1996) and Attardo in their works have shown how, basically, it’s amenability to competing scripts (loosely meant), that undelies humour. In a goal-driven approach to understanding, Dyer emphasized planning-related inadequacies (Dyer *et al.*, 1986, 1988). In this paper, epistemic inadequacy (involving beliefs about one’s identity) motivates behavioral mismatch on a parallel-worlds scale. Yet, epistemic and behavioural inadequacies by themselves wouldn’t make for humour; it’s out of anger, not amusement, that somebody in Italy may snap at you: *Lei non sa chi sono io!* — “If only you knew who I am! [And how well-connected I am!]”.¹ Still, in Pirandello’s *Henry IV*, it’s violent behaviour that a medieval emperor may afford, while a modern private citizen is not supposed to, which time and again is taken by characters to be evidence of an individual (the protagonist) being crazy; the story plot contains three phases in which the protagonist, *A*, impersonates *B*: in turn, *A* is initially acting *B* in a conventional performance, then *A* is delusionally, pathologically *B*, and finally *A* is *B* by imposture, the latter in a special sense that will emerge in the discussion. Here is the précis of

¹Many would poke fun at the very phenomenon, however. For example, Prof. Shmuel Moreh of the Hebrew University in Jerusalem, in a recent commemoration of Prof. Elie Kedourie (1926–1992), a prominent orientalist from the London School of Economics, reminisced how, in Baghdad apparently in the late 1940s, Kedourie and Moreh’s own brother, Jacob (Kedourie’s classmate), used to discuss “Elie’s articles, before and after their publication in Baghdad newspapers. Their discussions were full of humour, sometimes with ironic, absurd and sharp remarks mingled with high bursts of laughter or sardonic smile [...]. One notable example that they would repeat was that of a tribal chief [who was a Member of Parliament, and] who repudiated the censure of the traffic police with the boast of thousands of tribal gunmen at his disposal” (Moreh, 2001, p. 55, col. 3).

Luigi Pirandello's play of 1922, *Enrico IV*; the amount of details included here is intended to do justice to the intricacies of the dynamics of the characters' own beliefs and beliefs they ascribe to some other characters, as the plot unfolds. The play is set in an upper-class milieu, at a villa in Umbria in Pirandello's own days. The protagonist, who is never referred to by his real name, and—at the time in his life when the action as shown takes place—he is about 50. He impersonates Henry IV, the 11th-century emperor of Germany (the one who had to humiliate himself at Canossa), and is surrounded in the villa by servants who impersonate the emperor's young vassals from his secret council. Over twenty years before, acting in the role of Henry IV at a pageant, the protagonist fell off his horse, and the blow on his head caused him to enter a psychologically abnormal state, because of which he believes he actually is Henry IV, who in turn is by some kind of evil magic stopped at the age of 26, when he was begging to have his excommunication revoked at Canossa. His entourage connives in posturing as though this is his real identity indeed. (In one scene, the protagonist recalls a visit, a few months before, by Agnes, the emperor's mother: this was the protagonist's sister in disguise, and she has meanwhile passed away. The protagonist, in that scene, states he has overheard she is dead, but then he reckons aloud that as 'he' is 26, i.e., at the point in time in the life of Henry IV when he went to Canossa, it couldn't be that Agnes is dead, so the sorrow for this event has to be deferred by some more years.)

Unbeknown to his entourage, the deluded man has eventually recovered his old normal self, realized his situation, and become aware that his rival of old, Baron Tito Belcredi, has supplanted him in the good graces of Marquise Matilde Spina, the woman he was courting. Now 45, a widow, and a *fausse blonde*, earlier in life—at the pageant (still a young brunette, like now her daughter Frida)—she had played Marquise Matilde di Toscana (i.e., Matilda of Tuscany), cousin, foe, and host in Canossa of the historical Henry IV. In the villa's "hall of the throne" there are two modern-style realistic portraits, intended to "mirror" the images of how at the pageant 'Henry IV' and 'Matilda of Tuscany' had looked. The mature blonde no longer looks like her old portrait, but Frida, now 19, startlingly does. Frida is betrothed to the young Marquis Carlo Di Nolli.² The protagonist decides to feign that things are as usual, in the sense that he would make his entourage believe that he still thinks of himself as Henry IV, so they are to deal with him accordingly. The action in the play revolves on a visit to the villa by Di Nolli, Matilde, Frida, Belcredi, and a doctor.³ The protagonist had meanwhile revealed to the "vassals" that he has been his real, modern-man self for a while, and has nevertheless feigned and went on with his medieval emperor identity; he encourages the "vassals" to behave likewise. The revelation was leaked to Belcredi and Matilde. The protagonist is talking to the visitors, in particular Matilde and Belcredi, as one who has called the trick, yet persists on purpose, not being eager in the least to re-enter ordinary social life. Even though he no longer harbours amorous feelings for Matilde (Belcredi is rather spiteful as he realizes her still positive attitude to the protagonist), the protagonist reveals things intended to dismantle Belcredi in her esteem (already meager, even though they are lovers); especially, by making Matilde and Belcredi understand he knows it was somebody (Belcredi) who deliberately

²Di Nolli is the one employing the staff and giving orders at the villa, as he is the nephew, guardian and carer of the protagonist since the recent loss of their respective mother and sister. One of the "vassals" procures theatre costumes; we even learn in passing that on occasion—at such times when the protagonist gave order to convoke Henry IV's beleaguered wife, Berta of Susa—different women had been paying visit to 'Henry IV' as occasional, mercenary, **amused** impersonators (who would each claim her name is Berta, and she is from Susa, the town in Piedmont—the historical reference being beyond their grasp).

³By way of a therapeutic attempt devised by the doctor, these eventually masquerade: the doctor and Belcredi as Benedictines, Di Nolli as Henry IV at the pageant of twenty years before, Matilde Spina first as Duchess Adelaide (Berta's mother, the mother-in-law of Henry IV, expected to rebuke the emperor for neglecting her daughter), then both Matilde and Frida as Matilda of Tuscany: supposedly, seeing real persons coming out from behind the respective portraits, the protagonist is to be 'healed' and realize also the time elapsed. When Frida comes out of a niche behind the portrait that so resembles her, and calls out repeatedly "Enrico" (the referent is ambivalent: she is to call out and attract the attention of the protagonist, but Di Nolli is to come out from behind the man's portrait), the protagonist, seeing her in flesh, is quite scared lest he is going mad again. Frida herself is scared and proclaiming this aloud to the other visitors, so in practice the protagonist's own sudden distress is ignored.

caused his horse fall. The protagonist implements a crescendo of steps to put his “friends” in embarrass, partly by blurring the boundaries between his modern and his medieval, imperial identity. The protagonist explicitly mocks the visitors’ masquerade. (He actually regrets his own lost youth. As a deluded person, he had partly dyed his grey hair blond.) They are old, he points out, but as for him time has stopped, he declares in paradoxical derision, it’s Frida he now claims for himself (and here he is posturing as the emperor) as “rightfully” his; he defiantly embraces Frida, laughing scornfully. The visitors rush to separate them, and the protagonist imperiously orders his vassal-impersonating servants to stop them; confused, the “vassals” obey him. Belcredi hurls himself at him; “You are not mad”, he shouts. The protagonist, by reflex, draws the sword of one of the “vassal” impersonators and, repeating interrogatively Belcredi’s statement, wounds him. Matilde shouts that the protagonist is crazy; Belcredi, contradicting her while being taken away, that he is not. Offstage, the hubbub, then a high-pitched shout by Matilde, then silence, signal Belcredi’s death. The protagonist, still onstage with the four young vassal-impersonators, is shocked, and utters the closing line: “Now it’s so indeed, it couldn’t be otherwise” (“Ora sí... per forza...”); then, calling to his vassals: “Here together, here together... and forever!” Social common sense dictates the interpretation of the play’s closing line. The protagonist, having killed a man, could only save himself from the punishment met to sane killers by pretending he is ‘still’ Henry IV.⁴ Perhaps he hopes for detention in the villa.⁵

⁴When, before this summary of Pirandello’s play, I stated that the protagonist’s relation to his pretended personal identity is at first delusional, then as an actor (as indeed, not only he knows he is not the ancient Henry IV, but he eventually makes his servants and then his visitors aware of his acting, and discusses this with them), and finally as an impostor, this imposture of his is not in the sense that *A* pretends he is *B* and that *C* believes that *A* is *B* indeed (where *A* is the play’s unnamed protagonist, *B* is the historical Henry IV, and *C* is whoever in the universe of the play is aware of and has access to the given person), but rather in the sense that *A* has committed a crime, and, starting right afterwards, tries to make the rest of the world believe that his criminal action was committed while he delusionally was *B*, and tries to dissuade them from thinking (the way he did, instead, shortly before the crime) that at the very time he, *A*, was driving the sword into the man who quite shortly afterwards died of his wounds, *A* was aware that he was the modern man *A*, liable under the modern law of the country (whereas an ancient emperor could behave arbitrarily and even kill at will), and that moreover, *A* is thereafter sane and thus punishable by law. To the effect of conveying that false belief about his own liability for the murder, *A*, the protagonist of the play, is going to behave as though delusionally, pathologically *A* believes he is *B* (the ancient emperor) and behaves accordingly.

⁵Not an impossibility, even though typically he should expect to be committed to a hospital for the criminally insane (a motif not unknown to the *noir* cinematic genre). At any rate, this is to go on forever, as long as he is alive. Two factors militate for such an outcome (for him, the lesser evil): firstly, after the wounding, it was Belcredi who realized that the recluse is sane, but Belcredi is dead and the others, notwithstanding his proclamations right before, don’t share this realization, so as long as after the crime the protagonist is taken to be insane, he will not be as liable as a sane offender. Secondly, by behaving insanely after the crime, the protagonist (who realizes the first point) deliberately provides the witnesses with what they take to be evidence that he was insane while committing the crime. The finale makes the play amply satisfy the requirements for qualifying as a tragedy. In fact, the protagonists, as well as Belcredi, and (up to a point) Matilde, all are tragic characters, because not only catastrophe strikes them, but realization dawns on the first two of their respective responsibility in bringing their misfortune about (whereas Matilde, before Belcredi dies, acquires an awareness she didn’t have, and that would have been relevant at turning points in her life history). The protagonist, having become a murderer, no longer has the choice of abandoning his insanity posture and going back to anything resembling a more normal life. As to Belcredi, he dies knowing (a) that this is because of the protagonist’s revenge, a revenge taken impulsively yet while sane; and (b) that the people present, and Matilde in particular, are now aware of Belcredi’s guilt in causing the horse fall. Even though Matilde settles for the insanity explanation of the murder, her final wail signals the loss to death of Belcredi, the convenient lover for whom she was making display of disrespect even before the protagonist demolished him in her eyes. (The benefit that accrued to Belcredi from the horse fall was late to come anyway: Matilde had married somebody else, and later on, once she was widowed, he had become the man in her life but, to her, just a tolerated, belittled *ersatz*. Belcredi during the visit is spited by her being partial to—and, he ironizes, attracted to—the insane recluse, all the more so during the revelations the latter makes.)

In the main, the epistemic states involved are amenable, somewhat laboriously yet with no fundamental inadequacies, a representation of agents' beliefs like the one described in Ballim *et al.*, in press (also see Ballim and Wilks (1991), Maida (1995, 1991), as well as, on logics of knowledge, Fagin *et al.* (1995), and, on uncertainty, uncertain reasoning about agents' beliefs and reasoning, Barnden (2001)). It can be shown however how some devices in the play require, as should be expected, more levels of formalization, in order to account for poetic conventions. Detecting such fine points is inessential for making sense of the plot and for reasoning on the maze of who believes what; it would take an architectural component about literary poetic conventions to model the ability to interpret such cues as well. Only part of those capabilities has to do with analogizing with plot details from a known literary repertoire: this falls outside the scope of this paper. Yet, the interference of a repertoire (intertextual impact on a plot or on wording) features prominently in my COLUMBUS model of a satirical pastiche (Nissan, COLUMBUS, in press).⁶

2 THE FORMALISM FOR IDENTITY IN *Henry IV*

First of all, let us introduce some useful notation. Our interest is practical representation, instead of oriented to abiding to the theoretical neatness of logic-based inferential engines. Nevertheless, by adopting a widespread notation (Lomuscio and Ryan, 1999, §1.3, pp. 120–121) from epistemic logic research (Fagin *et al.*, (1995), let formulae be generated by this grammar: $\varphi ::= p \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \Box_i \varphi \mid C\varphi$ where φ is a formula, $p \in P$ is a propositional atom, $i \in A$

⁶The posture, on the part of the protagonist of Pirandello's *Henry IV*, of replacing Matilde Spina with Frida fits in the same pattern as, in the classic horror film *The Mummy* (a 1932 Universal film directed by Karl Freund and featuring Boris Karloff), the revived mummy of an ancient Egyptian rebel priest, who goes on to pursue the daughter of an archaeologist, in replacement of the ancient woman whose mummy he destroys and who, the story claims, was reincarnated in his modern prey (whom he would like to transform into a fresh mummy). Recognizing this similarity obtains once the recognizing agent detected more elements in the analogy than the mere replacement of a person for another. And indeed, motif indexes and concepts from structural narratology would come handy, to account for that. More remarks are called for: how is the protagonist of *Henry IV* to behave after the murder is suggested allusively at the start of Act One, in a conversation among the young vassal-impersonators in the protagonist's absence from the scene; one of these servants, "Bertoldo" (in real life, Fino Pagliuca) is a newly hired employee, in replacement of one who in real life was called Tito (like Belcredi), and who used to impersonate the bishop Adalbert of Bremen. Tito had died, and Di Nollì did not intend to hire somebody else in his stead, yet had to have that post manned after all, by "Bertoldo", because the protagonist had blamed Adalbert's disappearance on rival courtiers, and had declared he wanted at court Bertoldo. Nobody could tell who, "historically", Bertoldo was, yet they complied. That Act One mentions how "Henry IV" coped with a dead Tito before Tito Belcredi even appears in the play, closes a cycle: it suggests how Henry IV is likely to cope with the aftermath of Belcredi's murder. Yet, it's not like the two Tito characters are identified with each other, other than in the aside detail of the symbolism of there being name-sakes in the play (and which, unlike in real life, can be expected to carry a meaning, because of poetic conventions). One more cycle obtains as Henry IV draws the sword and attacks, at the end of the play. Matilde Spina, recollecting in Act One his horse fall, stated that on coming back to his senses he drew his sword when provoked by revellers in costume, thus with a furor befitting the emperor personality and role, not the actor's; this had signalled his folly. As to Marquise Matilde (Spina) herself playing at the pageant Marquise Matilde (di Toscana), her character in Pirandello's play explains she had chosen that role for the very reason she was a name-sake; she had thus determined the protagonist's gallant response (he would be at her feet like at Canossa), and thus, his choice of his role as Henry IV as being the one accompanying her at the pageant. Still, she admits that at first she couldn't recall the historical connection, other than vaguely. At another level, Pirandello alludes to himself, but only in details, not as a full persona in passing himself off as one of the characters; in fact, in the likeness prescribed for the doctor's, the details of his having a pate and a silvery pointed small beard point to Pirandello (but the doctor having to have a debauched satyr's face presumably cannot be Pirandello's reference to himself! Rather, it perhaps pokes fun at Freudianism's emphasis on the *libido*). Moreover, a biographical reference to Pirandello obtains as Belcredi's idea for the Carnival pageant's theme admittedly came from an illustrated German magazine which Belcredi could only browse and not read, and which depicted the Kaiser at a pageant, clad in the picturesque costume of his former students union, *qua* alumnus of Bonn University, i.e., Pirandello's own *alma mater*.

is an agent, and $\Box_i \varphi$ denotes “Agent i knows φ ”. Moreover, $\Diamond_i \varphi \equiv \neg \Box_i \neg \varphi$ (that is, “Agent i does not know that not φ ”). Here is some shorthand notation:

$E\varphi \equiv \bigwedge_{i \in A} \Box_i \varphi$ “Everyone knows φ ” (each agent came to know φ privately).

$E^2\varphi \equiv E(E\varphi)$ Power notation: “Everyone knows that everyone knows φ ”.

$C\varphi \equiv \bigwedge_{x=1}^{\infty} E^x \varphi$ “ φ is common knowledge”
(the effect if φ was announced publicly).

Let $\boxed{}_A^{-1} B$ stand for: Whatever is known to agent B about A .

I introduce time: $\boxed{}_A^t \varphi$ stands for: At time t , agent A holds the belief φ . $\boxed{}_A^{-1,t} B$ stands for: Whatever, at time t , is known to agent B about A .

t_A^α, t_A^ω respectively stand for: The time at which agent A was born;
The time of the death of agent A .

\mathcal{M}_A stands for: The biographical context of agent A , at time t , $t_A^\alpha \leq t \leq t_A^{\text{now}}$.

$\mathcal{M}_A \equiv \bigcup_{t=t_A^\alpha}^{t_A^{\text{now}}} \mathcal{M}_A^t$ On the left, see one way (a union of sets) to describe the trajectory of biographical-context states in a person’s life from birth to the present. We discard this formula, ridden with repetituousness as it is, also because we need define discrete time increments that suit the trajectory’s evolution. Later on I’ll provide a recursive definition.

Υ stands for: Standard historiography.

\mathcal{H} stands for: Henry IV, Emperor of Germany, as per $\boxed{}_{\mathcal{H}}^{-1} \Upsilon$, which is his standard historiographical image; that is to say, whatever, Φ , is known to standard historiography about \mathcal{H} . Equivalently: Φ such that $\Box_{\Upsilon} \Phi$

$\mathcal{P}, \mathcal{P}_H$ stand for: The protagonist of *Henry IV*; and \mathcal{P} insofar as $\Box_{\mathcal{P}} (\mathcal{P} = \mathcal{H})$

\mathcal{P}_0 stands for: The protagonist, other than when he believes himself to be \mathcal{H} . So it’s: \mathcal{P} , insofar as $\Box_{\mathcal{P}} \neg (\mathcal{P} = \mathcal{H})$

When \mathcal{P} is \mathcal{P}_0 , he holds accurate knowledge about his actual biographical context (including his current milieu): $\left(\boxed{}_{\mathcal{P}_0}^{-1,t} \mathcal{M}_{\mathcal{P}_0}^t \right) = \mathcal{M}_{\mathcal{P}_0}^t$ where: $\begin{cases} \mathcal{M}_A^t \equiv \langle Q_A^t, h_A^t \rangle \\ h_A^t \equiv \langle \hat{e}_A^t, \hat{\mathcal{M}}_A^t \rangle \end{cases}$

and where: $\hat{t} \equiv \max \left(\left\{ \tau \mid \left(t_A^\alpha < \tau \right) \wedge \left(\mathcal{M}_A^\tau \neq \mathcal{M}_A^t \right) \right\} \right)$

The biographical context of agent A at time t is defined here as a doubleton, a tuple composed of the subset of people of A ’s entourage at time t who (by then) are individually known to A , and the history of A at time t .

Of that much, by default, a person is (supposedly) aware: $\boxed{}_A^t \mathcal{M}_A^t$

We'll come back to expectations, but we may assume that (unless we learn otherwise) an agent is aware, at time t , of that same agent's milieu, comprising the people in his (or her) current entourage, as well as that agent's history, which in turn, among the other things, comprises the same agent's past milieu states. The "history" member of the doubleton given above and repeated here: $\mathcal{M}_A^t \equiv \langle Q_A^t, h_A^t \rangle$ can be, in turn, conceived of as a doubleton $h_A^t \& \equiv \&\langle \hat{e}_A^t, \mathcal{M}_A^{\hat{t}} \rangle$ whose own second member is the latest past biographical context of the same agent, so that biographical context is defined recursively:

By our convention: $\hat{e}_A^t \equiv e_A^t - e_A^{\hat{t}}$ just as: $\hat{e}_A^{\hat{t}} \equiv e_A^{\hat{t}} - e_A^{\hat{\hat{t}}}$

That is to say, the symbol \hat{e} (surmounted by the hat) stands for a set (of events in which agent A is knowingly involved) resulting from a difference of two sets, and having occurred in the latest interval of time since the biographical context had last changed: \hat{e} is an increment with respect to the latest earlier point in time when the biographical context was different, whereas without the hat, e stands for a cumulated set: see on the right.

$$e_A^t \equiv \hat{e}_A^t \bigcup_{\tau = t^{\alpha}}^{\hat{t} \text{ (b.d.i.)}} \hat{e}_A^{\tau}$$

b.d.i. \equiv "by discrete increments like in the definition of \hat{t} vis-à-vis t ."

Here is some semantics: e_A^t stands for the set of such current events (here, conventionally, from birth on) that involve the agent A , provided that A is aware of them. As a subset of $\hat{e}_A^{\hat{t}}$, we need include A 's state of relations with Q_A^t . Other elements of $\hat{e}_A^{\hat{t}}$, possibly are events of the Self alone, e.g., some physiological event that stands out, say, a heart attack, or, then, some thoughts that occurred to the given individual.

Likewise, a life-long trajectory of states (as a set of sets, thus as a hypergraph: on hypergraphs, see Berge (1989)) can be reconstituted for \mathcal{M}_A^{τ} doubletons, Q_A^{τ} sets, and h_A^{τ} doubletons.

Now, let us also introduce the following notation for standard expectations:

$\text{exp} \begin{array}{|c|} \hline \square \\ \hline \varphi \end{array}^{-1} u$ stands for: The expected knowledge held by agent u about φ .

$\text{exp} \begin{array}{|c|} \hline \square \\ \hline \end{array}^{-1} u$ stands for: Whatever knowledge is expected to be held by agent u .

Actually, we can rewrite $\begin{array}{|c|} \hline \square \\ \hline \end{array}^t \mathcal{M}_A^t$ this way: $\text{exp} \begin{array}{|c|} \hline \square \\ \hline \end{array}^t \mathcal{M}_A^t$.

In particular, we need to explicitly deal with generally expected knowledge, to represent this:

$\text{exp} \begin{array}{|c|} \hline \square \\ \hline \mathcal{H} \end{array}^{-1} \text{SOC}_{\text{Pir}}^+$ stands for: As much knowledge as could be expected to have been held, concerning \mathcal{H} , by a person from the upper classes of Italian society in Pirandello's own days (in which the play is set).

By default, a person from the lower classes of Italian society of Pirandello's own days is not expected to know anything whatsoever about Henry IV, the Emperor of Germany. (See on the right.)

$$\text{exp} \begin{array}{|c|} \hline \square \\ \hline \mathcal{H} \end{array}^{-1} \text{SOC}_{\text{Pir}}^- = \emptyset$$

In particular, such a person would not be expected to know that there was any time, at which \mathcal{H} existed.

$$\neg \left(\exp \square_{\mathcal{H}}^{-1} \text{soc}_{\text{Pir}}^- \ni \text{“} [\exists t, \text{ at } t : \exists \mathcal{H}] \text{”} \right)$$

Arguably, this is one factor involved in motivating the **amused attitude** (see below on this) allegedly displayed by the female mercenary visitors who (as reported by one of the vassal-impersonators in the play) used to pay visit to \mathcal{P}_H and to each announce herself as being Berta and being from Susa. (To them, these are two separate, though concomitant assertions, as they don't identify the historical character named by the fixed phrase “Berta of Susa”.) An educated reception of that detail from the play will promptly recognize that the vassal-impersonators are servants, thus from the lower classes, and that in turn the social status of the impersonators of \mathcal{H} 's wife was considerably lower than even the servants', according to their basically belonging (respectively) inside or outside the realm of a decent way of life. Yet, out of professional reasons (stable for the servants, but sporadic for the hired women), they have been each made to acquire some subset of as per his standard historiographical image: $\square_{\mathcal{H}}^{-1} \Upsilon$ (i.e., the standard historiographical image of Henry IV, Emperor of Germany), so that they could cater to the needs of \mathcal{P}_H . In fact, the first scene in the play has a new member of staff being taught about \mathcal{H} and about \mathcal{P}_H by his fellow vassal-impersonators V , so that in practice, what both he and the hired women learned about \mathcal{H} , is $\square_{\mathcal{H}}^{-1} V$, or a subset thereof.

Let us turn now to pinpointing and symbolizing points or intervals in time which are relevant for the pageant at which the protagonist had fallen off his horse.

- $t^{y \text{ Canossa}}$ stands for: The time interval (the grainsize being one **year**) in which the Canossa event involving the Henry IV, Emperor of Germany, occurred.
- d_t^{pageant} stands for: Time of the pageant; it's an interval with medium-sized granularity: grainsize is one **day**.
- t^f stands for: Time of \mathcal{P} 's fall off the horse. $t^f \subset d_t^{\text{pageant}}$
- g_t^{s-} stands for: Immediate past time *vis-à-vis* the time t^s of event s , the grainsize value being g . This stands for grainsize also in the following:
- g_t^{s+} stands for: Immediate subsequent time *vis-à-vis* the time t^s of event s .

During an extended interval in one's life, from time $t_q^{\square \mathcal{H}}$, which occurred much later than one's birth, but much earlier than one's death, agents q (i.e., \mathcal{P} or the upper-class members of his milieu) each came to learn the standard historiographical knowledge about \mathcal{H} :

$$\text{At } t, t_q^\alpha \ll t_q^{\square \mathcal{H}} \ll t \leq t_q^\omega, \quad \square_q \left(\square_{\mathcal{H}}^{-1} \Upsilon \right),$$

$$\text{where: } q \in \mathcal{P} \cup \left\{ \mathcal{R} \mid \left(r \in \mathcal{M}_{\mathcal{P}}^t \right) \wedge \left(r \in \text{soc}_{\text{Pir}}^+ \right) \right\}$$

On the day of the pageant, up to his falling off the horse, \mathcal{P} was playing \mathcal{H} , i.e., $\bigcirc_{\mathcal{P}}^{\vartheta} \mathcal{H}$ (where ϑ is a *theatre* mode, versus λ for taking a *legal* role); and moreover, \mathcal{P} (like everybody at the pageant) was conscious of this fact and of this attitude of \mathcal{P} .

Once \mathcal{P} fell off his horse, he became \mathcal{P}_H , but the other people present still believed that he was acting, until, later in the day, a violent reaction of his reveals that he delusionally “is” \mathcal{H} .

$$\begin{aligned}
& \exists j, j \in J, \left(Q_{\mathcal{P}}^{d_t^{\text{pageant}}} \subset J \right) \wedge \left[\text{At } t, t \subseteq d_t^{\text{pageant}} : \right. \\
& \quad \left. \bigcirc_{\mathcal{P}}^{(t)} \mathcal{H} \subset \square_j^{-1,t} \mathcal{P} \right] \wedge \\
& \quad \wedge \left[\text{At } t, \left(t \subseteq d_t^{\text{pageant}} \right) \wedge \left(t \leq t^f \right) : \mathcal{P} \in J \right] \wedge \\
& \quad \wedge \left[\text{At } t, t^f < t < t^{\text{wakeup}} < t^{\text{Play'sPresent}} : \right. \\
& \quad \left. \square_{\mathcal{P}}^t \left(\left(\mathcal{P} = \left(\mathcal{H} \text{ at } t^{\text{Canossa}} \right) \right) \wedge \left(\mathcal{M}_{\mathcal{P}} = \mathcal{M}_{\mathcal{H}} \right) \wedge \right. \right. \\
& \quad \left. \left. \wedge \left[\left(\bigcup_{\tau = t^f}^t Q_{\mathcal{P}}^{\tau} \right) \subset \bigcup_{\tau = t_{\mathcal{H}}^{\omega}}^{\alpha} Q_{\mathcal{H}}^{\tau} \right] \right] \right) \right]
\end{aligned}$$

The latter formula in its entirety states that \mathcal{P} was only party to the knowledge that he was just acting at the pageant, up to his horse fall, whereas the other people of his entourage on that day went on temporarily believing so even after his fall; and that from that fall, up to some time before the play’s present, \mathcal{P} believed that he was (mind this) Henry IV within the time span of a year or so when the event at Canossa occurred (in fact, in the play, \mathcal{P} thinks of “himself” as though he was \mathcal{H} at the age of twenty-six, and he refers to his present time as though his self-humiliation at Canossa was current), yet he also believes that the people in \mathcal{P} ’s life (from the fall, to the present) belong to the set of people in Henry IV’s *entire* life!

$\xi \smile \zeta$ stands for a marital intercourse event between ξ and ζ . An upward crescent stands for marriage, a state whose onset was the wedding (and whose end is divorce or either spouse’s death): $\xi \frown \zeta$. If that state holds, then socially it’s well-regarded and dutiful that $[(\xi \smile \zeta) \text{ at } \{k \mid k \text{ is-a } \text{EventTime} \wedge \{k\} \text{ dots } \text{TimeInterval}(\xi \smile \zeta)\}]$. Royalty (where it applies) is at society’s top. Being a royal consort carries much prestige. Being mad doesn’t. And being promiscuous for a fee is at the lowest end of expected appreciativeness by social norms. Hence the **laughter** of \mathcal{F} , i.e., each impersonator of \mathcal{H} ’s legitimate royal consort, \mathcal{W} . In \mathcal{P} ’s actual life, \mathcal{F} was hired by the “vassal”-in-chief for \mathcal{P} ’s marital purposes:

$$\left(\bigcirc_{\mathcal{F}}^{\vartheta} \mathcal{W} \right) \wedge \left[\square_{\mathcal{F}} \square_{\mathcal{P}} \left(\left(\mathcal{P} = \mathcal{H} \right) \wedge \left(\mathcal{W} \frown \mathcal{H} \right) \wedge \left(\mathcal{F} = \mathcal{W} \right) \right) \right]$$

So when $\mathcal{F} \smile \mathcal{P}$ eventually obtains — \mathcal{F} is thinking — \mathcal{P} would be delusionally associating with the situation a situation diametrically opposed in terms of social prestige. This satisfies “script opposition” as meant in an approach *à la* Attardo & Raskin.⁷ Madness, sex, and posturing, each

⁷Victor Raskin’s “Script-based Semantic Theory of Humor” (SSTH) evolved into Attardo and Raskin’s “General Theory of Verbal Humor” (GTVH). Attardo (1996), discussing non-joke humour, mentioned (p. 95) the existence of an “expansionist approach” (to which neither he, nor I would commit ourselves), according to which, non-joke humour could be treated as complex jokes. Yet, “the broad theory to be developed will [possibly] need to incorporate some technical apparatus which was unnecessary in the SSTH and GTVH” (p. 87). In SSTH, two claims on verbal jokes were combined: “such a text is compatible in part or in full with two different [loosely scripts]” (‘script’ being used loosely), and “these two scripts

are a salient theme in the thematology of humour. Another element in \mathcal{F} 's amusement pertain to the onomastic requirement (which she does not properly understand) for qualifying as the deluded madman's legitimate wife. Complying fulfils the adage: "La ragione si dà ai matti" ("To the mad, say: 'Right'.") There is much more that is worthy to be analyzed in and formalized on Pirandello's *Henry IV*. The tyranny of space prevents me from, e.g., developing here the goal-and-plan kind of analysis of characters' behaviour, authorial strategy and wording tactics, I developed in the COLUMBUS project (Nissan, COLUMBUS, in press).⁸

REFERENCES

- Attardo, S.: Humor Theory Beyond Jokes. In Hulstijn and Nijholt (1996), pp. 87–101.
- Ballim, A. and Wilks, Y.: *Artificial Believers: The Ascription of Belief*. Erlbaum, 1991.
- Ballim, A., By, T., Wilks, Y. and Lieske, C.: Modelling Agent Attitudes in Legal Reasoning. In E. Nissan and A.A. Martino (Eds.), *Software, Formal Models, and Artificial Intelligence for Legal Evidence*, special issue, *Computing and Informatics*, in press.

are opposite in a specially defined sense" (Raskin, 1996, p. 9). Moreover (p. 11), in GTVH "the text of the joke and the script opposition were redefined as two of 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)".

⁸first announced in Nissan (1996). In Nissan ("Collegno", in press), I analyzed beliefs, communication events, and actions from Italy's 1927 amnesia case of the Smemorato di Collegno, contended by two alleged wives: a gripping drama that a 1962 film of comedian Totò turned into a madhouse farce. Amnesia sufferers recovering shreds of identity are of course a world apart from delusions (Rust 1990) and delusional identities, the subject of a classic study by Milton Rokeach (1964) about the psychotic reasoning of three committed patients who each believed he was the same historical character as the other two, without this intruding into their routine coexistence. This is worth mentioning here, as handling paraconsistency in reasoning is important for *Henry IV*. Marvin Minsky (who on jokes wrote Minsky (1980)), at his presentation (Minsky, 2000) at AISB 2000 in Birmingham, announced a forthcoming book, *The Emotion Machine*, on human individual cognition actually several coexisting identities, hence one's contradictory beliefs: "Thus, when someone asks what you 'really' believe [...] that] make[s] sense in the Single-Self realm [which is unrealistic]" (Minsky, 2000). At that same symposium, Keith Oatley (2000) dealt with "three primary kinds of socio-emotional scripts", analyzing in the process Shakespeare's *Henry IV*, *Hamlet*, and *As You Like It*. I would like to end this paper with a remark on a kind of multimedial humour, namely, a musical joke in an operatic context. The opportunity for this is provided by a theme shared by the following example, and Pirandello's *Henry IV*: a character's disillusionment with the possibility to escape into an alternative world. James Parakilas (1993/94) pointed out: "There are no examples of exoticism in the Western musical tradition more famous than the 'Gypsy' numbers in *Carmen*: [...]. A number of scholars have examined the techniques of Bizet's exoticism, [...]. But there has been no comparable consideration of why this musical exoticism [...] lasts for only half the opera. Once Don José stops Carmen in the middle of her castanet dance, there is no Gypsy or Spanish sound to her music for the rest of the opera, [...]. Musical exoticism in *Carmen* is a matter of dramatic structure, not simply of local color. Its function is not to characterize Carmen and her fellow Gypsies as much as to map a change in the relationship between Carmen and Don José. The silencing of the exotic music in the middle of the opera marks the moment when Don José follows Carmen into her world and discovers that he can neither force her to be the Carmen he had dreamed of nor escape from his own world, from himself" (p. 33). "The dramatic function of the Gypsy music is clarified only after Don José, responding to the bugle call, makes Carmen stop her dance and put down her castanets. In so doing he drives Gypsy music out of the opera. Once he has taken up life among the Gypsies, exchanging the lure for the experience of their life, they take on the style of the culture he thought he was escaping. In fact, the music the Gypsy smugglers come onstage singing at the beginning of act 3 ("Ecoute, compagnon, écoute!") is no Gypsy dance; it is all business, all discipline, practically a march [...]. This music can be heard as a grim musical joke on Don José, showing that for him there is no escape from the soldier's life, no escape from his culture and the rule of force that it has imposed on other cultures" (p. 46, my underlining). "Ironically, it is Don José's perspective that is represented by the use of Gypsy music in the opera; in effect, the Gypsy music is Bizet's operatic substitute for Mérimée's literary device of letting Don José narrate his own part of the story" (p. 47). Actually, the irony arises indeed from there being—as required by Sperber and Wilson (1981), Wilson and Sperber (1992)—both echoic mention, and negation.

- Barnden, J.A.: Uncertain Reasoning About Agents' Beliefs and Reasoning. In: A.A. Martino and E. Nissan (Eds.), *Formal Approaches to Legal Evidence*, special issue, *Artificial Intelligence and Law*, 9(2/3), 2001, pp. 115–152.
- Berge, C.: *Hypergraphs: Combinatorics of Finite Sets*. North-Holland, Amsterdam, 1989.
- Dyer, M.G., Flowers, M. and Reeves, J.F.: A Computer Model of Irony Recognition in Narrative Understanding. Tech. Rep., Computer Science Dept., UCLA, 1986.
- Dyer, M.G., Flowers, M. and Reeves, J.F.: Recognizing Situational Ironies: A Computer Model of Irony Recognition and Narrative Understanding. Tech. Rep., UCLA, 1988.
- Fagin, R., Halpern, J., Moses, Y., Vardi, M.: *Reasoning about Knowledge*. MIT Press, 1995.
- Hulstijn, J. and Nijholt, A. (Eds.): *Automatic Interpretation and Generation of Verbal Humor: Proceedings of the 12th Twente Workshop on Language Technology*, 1996.
- Lomuscio, A. and Ryan, M.: An Algorithmic Approach to Knowledge Evolution. In B. Knight and E. Nissan (Eds.), *Temporal Logic in Engineering*, special issue, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing (AIEDAM)*, 13(2), 1999, pp. 119–132.
- Maida, A.S.: Maintaining Mental Models of Agents Who Have Existential Misconceptions. *Artificial Intelligence*, 50, 1991, pp. 331–383.
- Maida, A.S.: Review of Ballim/Wilks (1991). *Minds and Machines*, 5(2), 1995, pp. 277–280.
- Minsky, M.: Jokes and the Logic of the Cognitive Unconscious. MIT AI Memo 603, 1980.
- Minsky, M.: Future Models for Mind-Machines (draft). In Sloman (2000), pp. 124–129.
- Moreh, Sh.: A Tribute to Elie Kedourie. (Reviewing: *Elie Kedourie [...]*, ed. by S. Kedourie; Frank Cass, London, 1998; = originally: supplement to *Middle Eastern Studies*, 33(4), 1998.) *The Scribe* (The Exilarch Foundation, London), no. 74, 2001, pp. 55–56.
- Nissan, E.: From ALIBI to COLUMBUS. In Hulstijn and Nijholt (1996), pp. 69–85.
- Nissan, E.: The COLUMBUS Model (2 parts). *International Journal of Computing Anticipatory Systems*, in press.
- Nissan, E.: A Formalism for Competing Claims of Identification: The “Smemorato di Collegno” Amnesia Case. *Informatics and Computing*, in press.
- Oatley, K.: Shakespeare's Invention of Theatre as Simulation That Runs on Minds. In Sloman (2000), pp. 102–110.
- Parakilas, J.: The Soldier and the Exotic, Part I. *The Opera Quarterly*, 10(2), 1993/94, 33–56.
- Pirandello, L.: *Enrico IV*. Bemporad, Florence, 1922; also: B.M.M., Mondadori, Milan, 1948; Oscar Mondadori 233, 1966, 1974. English translation: *Henry IV*, trans. J. Mitchell, Eyre Methuen, London, 1979; trans. F. May, The University of Leeds, 1956. Also in *Three Plays: Six Characters in Search of an Author, Henry IV, Right You Are! (If You Think So)*, trans. A. Livingston, Dent, London, 1922, 1923, 1931.
- Raskin, V.: Computer Implementation of the General Theory of Verbal Humor. In Hulstijn and Nijholt (1996), pp. 9–19.
- Rokeach, M.: *The Three Christs of Ypsilanti: A Psychological Study*. Knopf, New York, 1964; Arthur Barker Ltd., London, 1964; Columbia University Press, New York, 1981.
- Rust, J.: Delusions, Irrationality and Cognitive Science. *Philosophical Psychology*, 3(1), 1990, pp. 123–138.

- Sloman, A. (Ed.): *Proceedings of the AISB'00 Symposium on How to Design a Functioning Mind*, Birmingham, 17–20 April 2000.
- Sperber, D. and Wilson, D.: Irony and the Use-Mention Distinction. In: P. Cole (Ed.), *Radical Pragmatics*. Academic Press, New York, 1981, pp. 295–318.
- Wilson, D. and Sperber, D.: On Verbal Irony. *Lingua*, 87, 1992, pp. 53–76.

HAHAcronym: Humorous Agents for Humorous Acronyms

Oliviero Stock and Carlo Strapparava
ITC-irst, Istituto per la Ricerca Scientifica e Tecnologica
I-38050 Trento, ITALY
{stock|strappa}@itc.it

Abstract

HAHAcronym¹ is the first project concerned with computational humor sponsored by the European Commission. The project is meant to convince about the potential of computational humor, through the demonstration of a working prototype and an assessment of the state of the art and of scenarios where humor can add something to existing information technologies. The main goal of HAHAcronym is the realization of an acronym *ironic* re-analyzer and generator as a proof of concept in a focalized but non restricted context. In order to implement this system some general tools have been adapted, or developed for the humorous context. A fundamental tool is an incongruity detector/generator: in practice There is a need to detect semantic mismatches between expected sentence meaning and other readings, along some specific dimension (i.e. in our case the acronym and its context). For all tools, particular attention has been put on reusability.

1 INTRODUCTION

It is time that Computational Humor is perceived as a field that can deliver something useful. The current dynamics of research support, for good or for bad, is based on the prospect that the research themes are conducive to practical results in more or less short terms. In the field of information technology, generally there is an especially limited amount of patience for the first visible outcome of research activity. Another element for pushing toward some concrete results is, on the other side, that by means of those constructive experience we, ourselves, can become more aware of problems and necessities. Besides anything else we believe there are real opportunities out there, that not always require a complete modeling of humor processes. We are convinced that deep modeling of humor in all of its facets is not something for the near future. The phenomena are too complex; humor is one of the most sophisticated forms of human intelligence. It is *AI-complete*: the problem of modeling it is as difficult to solve as the most difficult Artificial Intelligence problems. But some steps can be followed to achieve results. And when something is realized we can note that humor has the methodological advantage (unlike, say, computer art) of leading to more directly falsifiable theories: the resulting humorous artifacts can be tested on human subjects in a rather straightforward manner.

Basically, in order to be successfully humorous, a computational system should be able to: recognize situations appropriate for humor; choose a suitable kind of humor for the situation; generate an appropriately humorous output; and, if there is some form of interaction or control, evaluate the feedback.

And indeed society needs humor, not just for entertainment. In the current business world, humor is considered to be so important that companies may hire ‘humor consultants’. Humor can be used “to criticize without alienating, to defuse tension or anxiety, to introduce new ideas, to bond teams, ease relationships and elicit cooperation”.

¹HAHAcronym is a Future Emerging Technologies (FET) EC project (contract number IST-2000-30039) in the context of Information Society Technologies (IST) programme. The consortium is composed by ITC-irst (coordinator partner) and by the Centre for Telematics and Information Technology/Parlevink of University of Twente.

So, looking at computational humor from an application-oriented point of view, one assumption is that in future human-machine interaction, humans will demand a naturalness and effectiveness that requires the incorporation of models of possibly all human cognitive capabilities, including the handling of humor. There are many *practical* settings where computational humor will add value. Among them there are: business world applications (such as advertisement, e-commerce, etc. . .), general computer-mediated communication and human-computer interaction, increase in the friendliness of natural language interfaces, educational and edutainment systems.

Not necessarily applications need to emphasize interactivity. For instance there are important prospects for humor in automatic information presentation. In the Web age presentations will become more and more flexible and personalized and will require humor contributions for electronic commerce developments (e.g. product promotion, getting selective attention, help in memorizing names etc.) more or less as it happened in the world of advertisement within the old broadcast communication.

We are concerned with systems that automatically produce humorous output (rather than systems that appreciate humor). Some of the fundamental competencies are within the range of the state of the art of natural language processing.

In one form or in another humor is most often based on some form of incongruity. In verbal humor this means that at some level different interpretations of material must be possible (and some not detected before the culmination of the humorous process) or various pieces of material must cause perception of specific forms of opposition. Natural language processing research has often dealt with ambiguity in language. A common view is that ambiguity is an obstacle for deep comprehension. Most current text processing systems, for example, attempt to reduce the number of possible interpretations of the sentences, and a failure to do so is seen as a weakness of the system. The potential for ambiguity, however, can be seen as a positive feature of natural language. Metaphors, idioms, poetic language and humor use the multiple senses of texts to suggest connections between concepts that cannot, or should not, be stated explicitly. Fluent users of natural language are able to both use and interpret ambiguities inherent in the language and verbal humor is one of the most regular uses of linguistic ambiguity.

The work presented here is based on various resource for natural language processing, adapted for humor. It is a small step, but aiming at an appreciable concrete result. It has been developed within the first European Project devoted to computational humor. A visible and evaluable result were at the basis of the deal. We have proposed a situation that is of practical interest, where many components are present but simplified with respect to more complex scenarios. The goal is a system that makes fun (Victor Raskin says *profane*) existing acronyms, or, starting from concepts provided by the user, produces a new acronym, constrained to be a word of the given language. And, of course, it has to be funny.

HAHAcronym is the first project concerned with computational humor sponsored by the European Commission. The project is meant to convince about the potential of computational humor, through the demonstration of a working prototype and an assessment of the state of the art and of scenarios where humor can add something to existing information technologies.

The main goal of HAHAcronym is the realization of an acronym *ironic* re-analyzer and generator as a proof of concept in a focalized but non restricted context. In order to implement this system some general tools have been adapted, or developed for the humorous context. A fundamental tool is an incongruity detector/generator: in practice There is a need to detect semantic mismatches between expected sentence meaning and other readings, along some specific dimension (i.e. in our case the acronym and its context). For all tools, particular attention has been put on reusability.

2 STATE OF THE ART ON COMPUTATIONAL HUMOR

So far only very limited effort has been put on building computational humor prototypes.

Indeed very few working prototypes that process humorous text and/or simulate humor mechanisms exist. Mostly they are concerned with rather simple tasks.

There has been a considerable amount of research on linguistics of Humor and on theories

of semantics or pragmatics of humor (Attardo, 1994; Attardo and Raskin, 1991; Giora and Fein, 1999); however, most of the work has not been formal enough to be used directly for computational modelling of humor. In this sense Pepicello and Green's work on puns (Pepicello and Green, 1984) is an exception.

Within the artificial intelligence community, most writing on humor has been speculative (Minsky, 1980; Hofstadter et al., 1989). Minsky made some preliminary remarks about how humor could be viewed from the artificial intelligence/cognitive science perspective, refining Freud's notion that humor is driven by our mental "censor" which control inappropriate thoughts and feelings. Utsumi (1996) outlines a logical analysis of irony, but this work has not been implemented in any way. Among other works: De Palma and Weiner (1992), who have worked on knowledge representation of riddles, Katz (1993), who attempted to develop a neural model of humor. Ephratt (1990) has constructed a program that parses a small range of ambiguous sentences and detects alternative humorous readings.

Probably the most important attempt to create a computational humor prototype is the work of Kim Binsted and Graeme Ritchie (1997).

They have devised a formal model of the semantic and syntactic regularities underlying some of the simplest types of punning riddles. A punning riddle is a question-answer riddle that uses phonological ambiguity. The three main strategies used to create phonological ambiguity are syllable substitution, word substitution and metathesis.

The last mention is for Jester (Goldberg et al., 2001), an on-line joke recommending system using collaborative filtering in order to recommend jokes adapted to the user's sense of humor'. It does not implement any linguistic technique for humor, rather it uses statistical techniques to recommend jokes based on the user's ratings of a set of sample jokes.

Almost all the approaches try to deal with the incongruity theory at various level of refinement (Köstler, 1964; Raskin, 1985). The incongruity theory focuses on the element of surprise. It states that humor is created out of a conflict between what is expected and what actually occurs in the joke. This accounts for the most obvious features of much humor: ambiguity or double meaning. Moreover, what we have tried to exploit in the HAHACronym project is incongruity planned as terms that are 'inconsistent', not fitting well together, disjointed, unsuitable. In the prototype primarily we have exploited an "off-the-shelf" resource (WORDNET DOMAINS) to find and employ incongruous terms in re-analyzing and generating funny acronyms.

3 THE HAHACRONYM PROJECT

As said HAHACronym is the first European project about computational humor. It is part of the Future Emerging Technologies Program within the Information Society Technologies section of The Fifth FP.

The main goal of the project is the realization of an acronym re-analyzer and generator as a proof of concept in a focalized but non restricted context. We had to propose something that we would be able to develop in a short period of time (few months) that would be meaningful, well demonstrable, that could be evaluated along some pre-decided criteria, and that is conducive to a subsequent development in a direction of potential industrial interest. So for us it was essential that a) the work could have many components of a larger systems, simplified for the current setting; b) we could reuse and adapt existing relevant linguistic resources; c) some simple strategies for humor effects could be experimented.

One of the purpose of the project was to show that using "standard" resources (with some extensions and modifications) and suitable linguistic theories of humor (i.e. developing specific algorithms that implement or elaborate theories), it is possible to implement a working prototype.

For that, we have taken advantage of specialized thesauri and repositories and in particular of WORDNET DOMAINS, an extension developed at ITC-irst of the well-known English WORDNET. In WORDNET DOMAINS, synsets are annotated with subject field codes (e.g. MEDICINE, ARCHITECTURE, LITERATURE. . .) providing cross-categorical information. WORDNET DOMAINS is organized for multilinguality and an Italian extension is already available. Other important computational tools are: a rule database of semantic field oppositions with humorous potential;

a parser for analyzing input syntactically and a syntactic generator of acronyms; general lexical resources, e.g. acronym grammars, morphological analyzers, rhyming dictionaries, proper nouns databases.

The initial results have been submitted to an evaluation by human subjects that eventually have given indications for re-use of the semantic tools. Even if the scenario of this assessment project is simplified, the tools developed will be reusable for other applications and, in principle, portable to other languages.

Beside the prototype, the project produces an assessment of applied opportunities in the market; an assessment of the state of the art in the field and of what can be done in a short to medium term. The working demonstrator is meant to convince about the feasibility and the potential impact of a larger applied project.

4 RESOURCES AND GENERAL ARCHITECTURE

In order to implement the HAHAcronym prototype, we have refined existing resources and we are developing general tools useful for humorous systems. As we will see, a fundamental tool is an incongruity detector/generator: concretely we need to be able to detect semantic mismatches between word meaning and sentence meaning (i.e. in our case the acronym and its context). For all tools, particular attention is put on reusability.

Analyzing or generating verbal humor, calls for exploitation and adaptation and elaboration of natural language processing resources such as lexicons, part-of-speech taggers, parsers, annotation tools, knowledge representation formalisms.

The starting point for us consisted in making use of some “off-the-shelf” resources, such as WORDNET DOMAINS (an extension of the well-known English WORDNET) and standard parsing techniques. The tools resulting from the adaptation will be reusable for other applications, and are portable straightforwardly to other languages (e.g. WORDNET DOMAINS is multilingual).

4.1 WORDNET AND WORDNET DOMAINS

WORDNET is a thesaurus for the English language inspired by psycholinguistics principles and developed at the Princeton University by George Miller (Miller, 1990; Fellbaum, 1998). It has been conceived as a computational resource, therefore improving some of the drawbacks of traditional dictionaries, such as circularity of definitions and ambiguity of sense references. Lemmata (about 130,000 for version 1.6) are organized in synonym classes (about 100,000 *synsets*). WORDNET can be described as a “lexical matrix” with two dimensions: a dimension for *lexical relations*, that is relations holding among words, and therefore language specific, and a dimension for *conceptual relations*, which hold among senses (the synsets) and that, at least in part, we consider independent from a particular language. A synset contains all the words by means of which it is possible to express the synset meaning: for example the Italian synset {calcium, calcio, Ca} describes the sense of “calcio” as a chemical substance, while the synset {calcio, pedata} describes the sense of “calcio” as a leg movement. A list of the main relations present in WORDNET follows.

4.1.1 AUGMENTING WORDNET WITH DOMAIN INFORMATION

Domains have been used both in linguistics (i.e. Semantic Fields) and in lexicography (i.e. Subject Field Codes) to mark technical usages of words. Although this is useful information for sense discrimination, in dictionaries it is typically used for a small portion of the lexicon. WORDNET DOMAINS is an attempt to extend the coverage of domain labels within an already existing lexical database, WORDNET (version 1.6). The synsets have been annotated with at least one domain label, selected from a set of about two hundred labels hierarchically organized. (Figure 1 shows a portion of the domain hierarchy.)

We have organized about 250 domain labels in a hierarchy (exploiting Dewey Decimal Classification), where each level is made up of codes of the same degree of specificity: for example, the second level includes domain labels such as BOTANY, LINGUISTICS, HISTORY, SPORT and RELI-

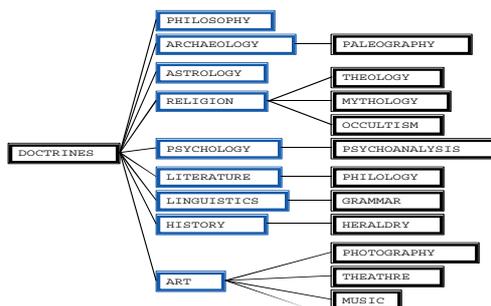


Figure 1: A sketch of the domain hierarchy

GION, while at the third level we can find specialization such as AMERICAN_HISTORY, GRAMMAR, PHONETICS and TENNIS.

Information brought by domains is complementary to what is already present in WORDNET. First of all a domain may include synsets of different syntactic categories: for instance MEDICINE groups together senses from Nouns, such as `doctor#1` and `hospital#1`, and from Verbs such as `operate#7`. Second, a domain may include senses from different WORDNET sub-hierarchies (i.e. deriving from different "unique beginners" or from different "lexicographer files"). For example, SPORT contains senses such as `athlete#1`, deriving from `life_form#1`, `game_equipment#1`, from `physical_object#1`, `sport#1` from `act#2`, and `playing_field#1`, from `location#1`.

4.1.2 OPPOSITION OF SEMANTIC FIELDS

On the basis of well recognized properties of humor accounted for in many theories (e.g. incongruity, semantic field opposition, apparent contradiction, absurdity) we have modelled an independent structure of domain opposition, such as RELIGION vs. TECHNOLOGY, SEX vs. RELIGION, etc. . . We exploit these kind of opposition as a basic resource for the incongruity generator.

4.1.3 ADJECTIVES AND ANTONYMY RELATIONS

Adjectives play an important role in modifying and generating funny acronyms. So we gave them a thorough analysis. WORDNET divides adjectives into two categories. *Descriptive adjectives* (e.g. `big`, `beautiful`, `interesting`, `possible`, `married`) constitute by far the largest category. The second category is called simply *relational adjectives* because they are related by derivation to nouns (i.e. `electrical` in `electrical engineering` is related to noun `electricity`). To relational adjectives, strictly dependent on noun meanings, it is often possible to apply similar strategies as those exploited for nouns. Their semantic organization, though, is entirely different from the one of the other major categories. In fact it is not clear what it would mean to say that one adjective "is a kind of" (ISA) some other adjective. The basic semantic relation among descriptive adjectives is antonymy. WORDNET proposes also that this kind of adjectives are organized in clusters of synsets associated by semantic similarity to a focal adjective. Figure 2 shows clusters of adjectives around the direct antonyms *fast/slow*.

4.1.4 EXPLOITING THE HIERARCHY (E.G. GEOGRAPHIC NAMES)

It is possible to exploit the network of lexical and semantic relations built in WORDNET to make simple ontological reasoning. For example, if a noun or an adjective has a geographic location meaning, the pertaining country and continent can be inferred.

4.2 RHYMES

The HAHAcronym prototype takes into account word rhymes and the rhythm of the acronym expansion. To cope with this aspect we got and reorganized the CMU pronouncing dictionary

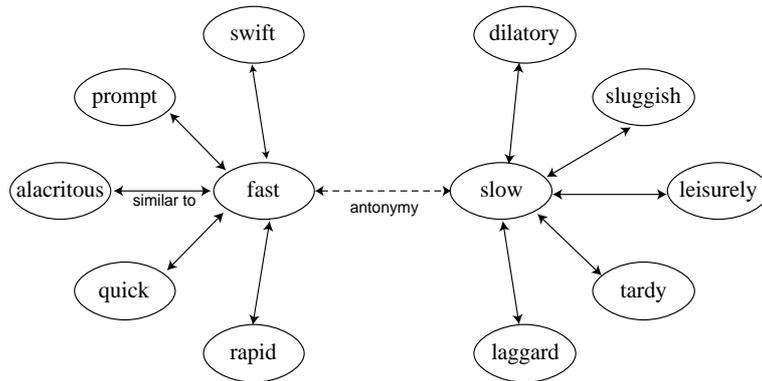


Figure 2: An example of adjective clusters linked by antonymy relation

(<http://www.speech.cs.cmu.edu/cgi-bin/cmudict>) with a suitable indexing. The CMU Pronouncing Dictionary is a machine-readable pronunciation dictionary for North American English that contains over 125,000 words and their transcriptions.

Its format is particularly useful for speech recognition and synthesis, as it has mappings from words to their pronunciations in the given phoneme set. The current phoneme set contains 39 phonemes; vowels may carry lexical stress. (e.g. 0 No stress, 1 Primary stress, 2 Secondary stress).

The current phoneme set has 39 phonemes, not counting variations for lexical stress.

4.3 PARSER AND GRAMMAR

Word sequences that are at the basis of acronyms are subject to a well-defined grammar, simpler than a complete noun phrase grammar, but complex enough to require a nontrivial analyzer. We have decided to use a well established nondeterministic parsing technique (ATN-based parsing). Ordinarily, an ATN parser has three components: the ATN itself, that represent the grammar in the form of a network, an interpreter for traversing it, and a dictionary (possibly integrated with a morphological analyzer). As obvious at this point for the third component we use WORDNET; integrated with an ad-hoc morphological analyzer. As far as the interpreter is concerned, we developed an ATN compiler that translate ATN's directly into Lisp code (i.e. Lisp augmented with nondeterministic constructs). Figure 3 sketches a portion of the acronym grammar.

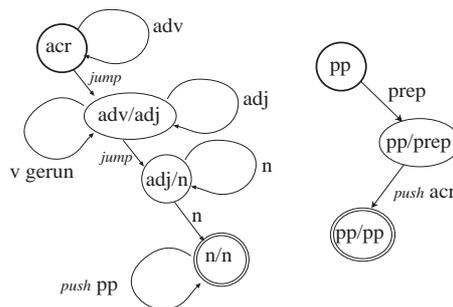


Figure 3: A simplified grammar

Even if for the generation part we do not traverse the grammar, we exploit it as the source for syntactic constraints also there.

4.4 A-SEMANTIC DICTIONARY

The A-semantic dictionary is a collection of hyperbolic/epistemic/deontic adjective/adverbs. This resource is particularly useful when it is necessary to complete a forming acronym (especially dur-

ing the generation of new acronyms). Some examples of this kind of words are: abnormally, abstrusely, adorably, exceptionally, exorbitantly, exponentially, extraordinarily, voraciously, weirdly, wonderfully. This resource is hand-made using various dictionaries as information sources.

Other important lexical resources are: an euphemism dictionary, a proper noun dictionary, lists of typical foreign words commonly used in the language with some strong connotation.

4.5 IMPLEMENTATION

To get an ironic or “profaning” re-analysis of a given acronym, the system follows various steps and strategies. The main elements of the algorithm can be schematized as follows:

- acronym parsing and construction of a logical form
- choice of what to keep unchanged (typically the head of the highest ranking NP) and what to modify (e.g. the adjectives)
 - using semantic field oppositions,
 - keeping the initial letter, rhyme and rhythm (the modified acronym should sound similar to the original as much as possible)
 - for adjectives, basing reasoning mainly on WORDNET antonymy clustering
 - using the a-semantic dictionary
- look for possible substitutions
 - using semantic field oppositions,
 - keeping the initial letter, rhyme and rhythm (the modified acronym should sound similar to the original as much as possible)
 - for adjectives, basing reasoning mainly on WORDNET antonymy clustering
 - using the a-semantic dictionary

Figure 4 and 5 show respectively a sketch of the architecture and an example of the system strategies.

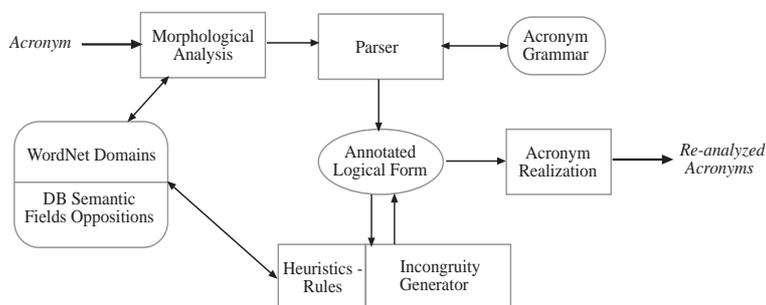


Figure 4: Acronyms Reanalysis: a sketch of the demonstrator architecture

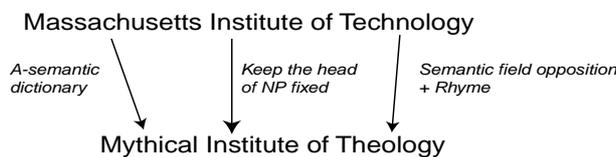


Figure 5: An example of acronym reanalysis

Making fun of existing acronyms is basically using irony on them, desecrating them with some unexpectedly contrasting but otherwise consistently sounding expansion.

In practice, as mostly you do not come across acronyms and their expansion an ironizing system, that for instance is part of a funny mail replier, will first look up for the *good* expansion in a large repository (The Acronym Finder <http://www.acronymfinder.com/>).

Then the process would start leading to a reformulation of the original sequence of words.

As far as acronym generation is concerned, the problem is more complex. We constrain resulting acronyms to be words of the dictionary. The system takes in input some concepts (actually synsets, so that even input to this system can result from some other processing, for instance sentence interpretation) and some minimal structural indication, such as the semantic head. The primary strategy of the system is to consider as potential acronyms words that are in ironic relation with the input concepts. By definition the acronyms will have to satisfy constraints, such as to include the initial letters of some lexical realization of the inputs words synsets, with some elaboration that will grant the sequence of initials on the basis of the constraints of the overall acronym syntax. In this primary strategy, the ironic reasoning comes mainly at the level of acronym choice in the lexicon and in the definition of the fillers of the *open slots* in the acronym.

For example, giving as input “fast” and “CPU”, we get **static, torpid, dormant**. (Note that the complete synset for “CPU” is {processor#3, CPU#1, central_processing_unit#1, mainframe#2}. So we can use in the acronym expansion a synonym of “CPU”. The same happens for “fast”.) Once we have an acronym proposal, a syntactic skeleton has to be filled to get a correct noun phrase.

For example given in input “fast” and “CPU”, the system selects **TORPID** with the possible syntactic skeleton:

< adv >_T < adj >_O Rapid Processor < prep > < adj >_I < noun >_D

or

< adj >_T < adj >_O Rapid Processor < prep > < noun >_I < noun >_D

where “rapid” and “processor” are synonyms respectively of “fast” and “CPU” and the notation *< Part_of_Speech >_Letter* means a word of that particular part_of_speech with *Letter* as initial.

Then the system fills this syntactic skeleton with strategies similar to those described for re-analysis.

5 EXAMPLES AND PRELIMINARY EVALUATION

Here below some examples of acronym re-analysis are reported. As far as semantic field opposition is concerned we have slightly tuned the system towards the domains **FOOD**, **RELIGION** and **SEX**. For each example we report the original acronym, the re-analysis and some comments about the strategies followed by the system.

ACM - Association for Computing Machinery
→ **Association for Confusing Machinery**

The acronym expansion is simple. The system keeps all the heads (the head of the whole NP and the head of the PP) and works on the adjective, considering mainly the rhyme and using the a-semantic dictionary.

CCTT - Close Combat Tactical Trainer (Army second generation virtual trainer.)
→ **Cold Combat Theological Trainer**

This is an example of two changes: antonym strategy for the first adjective and semantic opposition found in the **RELIGION** domain that modifies ‘Tactical’ into ‘Theological’.

CHI - Computer Human Interface
→ **Computer Harry_Truman Interface**

An unexpected result, mainly achieved exploiting rhyme.

DMSO - Defense Modeling and Simulation Office.
→ **Defense Meat_eating and Salivation Office**

The two modifications are coherent into the **FOOD** semantic fields. In general the system can choose either to keep coherence among the modification or to exploit contrast picking them from different ‘opposite’ semantic fields, as in the following example:

IST - Institute for Simulation and Training.

→ Institute for Sexual_abstention and Thanksgiving

FCCSET - Federal Coordinating Committee for Science, Engineering and Technology

→ Femoral Coordinating Committee for Sword_dance Earring and Theology

This example shows how too many modifications can produce a re-analysis that does not resemble the original acronym at all, with poor irony effect.

FLOPS - Floating Point Operations Per Second (A measure of computing power.)

→ Fucking Point Operations Per Second

Typical use of a taboo world.

Testing the humorous quality of texts or other verbal results is not an easy task. There are some relevant studies though, such as (Ruch, 1996; Köhler and Ruch, 1996). HAHAcronym for the moment has only some initial results for the re-analysis part. A preliminary evaluation with a group of about 30 university students has shown good results. The students had to choose among five levels of amusement (from `very_funny` to `not_funny`). The result are very encouraging. None of the re-analyses proposed to the students were been rejected as completely non-humorous. More than 50% were considered good enough.

At the writing time we have not started the evaluation of the complete prototype. The evaluation experiments will be organized as described in the schemas that follow.

Acronym re-analysis evaluation: Acronym humorous re-analysis, given an existing acronym

- Number of panel members: 20
- Method: for every re-analysis, each panelist gives from 1 to 5 (max) points. (100 pt is the maximum score given 20 testers)
- Number of experiments: 40
- Success threshold: in 60% of the given acronyms a re-analysis will have a score that exceeds 55/100
- The results will also be differentially compared with the score obtained by randomly generated (but syntactically correct) data obtained by the basic machinery, disengaging the specific resources and heuristics. 65% of acronyms will have a more humorous evaluation with the full machinery.

Acronym creation evaluation. Acronym humorous creation from scratch (i.e. from a set of concepts).

- Number of panel members: 20
- Method: for every acronym generation, each panelist gives from 1 to 5 (max) points. (100 pt is the maximum score given 20 testers)
- Number of experiments: 30
- Success threshold : in 45% of the sets of input concepts, provided there is a solution, a generated acronym will have a score that exceeds 55/100

6 PROBLEMS AND CONCLUSIONS

At the time of this writing the acronym re-analyzer is completed in its main structure and one of the main question is the ranking of possible re-analyses so that the ones that are supposed to be funnier appear at the top. The system is flexible and novel strategies can be added. Ranking a

priori is easy: it is based on a combination of values attributed by hand to the various strategies. Ranking a posteriori is much more difficult and involves some initial steps in modeling humor appreciation.

The generator is not completed yet, but the whole prototype will be completed within two months.

The implementation is in Allegro CommonLisp and the system could be easily integrated within a number of hosting applications.

We consider this as a first step of a set of possible developments of practical impact.

Some possible applied scenarios:

1. educational software for children based on the idea of machine assisted joke generation, thus helping children explore word-meanings;
2. a system that uses humor as a means to promote products and to get the user's attention in an electronic commerce environment;
3. an appealing names generator for products and merchandise (e.g. goods, commodities, pharmaceutical products, etc. . .). A final version of the prototype could be put on the Web to be tested: e.g. as a service for finding humorous acronyms for European Projects.

If there doubts about the possible practical utility of a system like this, let us just consider the case of European project proposals. The first field of the proposal is for the name, that must be an acronym. When a proposal is prepared, the energy of the most senior scientists involved in the consortium goes in finding a a good acronym. Being funny is a plus. A very conservative estimate is that on average to prepare a proposal may require a couple of person-months. For the name we think it is fair to say that it may take two person-days fo the best senior scientists. Every year proposals are submitted to the European Commission, numbering well over 10.000. So at minimum 20.000 days are spent on (if possible funny) acronyms, i.e. some 100 person-year of the most senior personnel involved: the cost can be estimated at over 10 million euro. A computational humor system resulting from this prototype would cost much less. . .

REFERENCES

- Attardo, S. (1994). *Linguistic Theory of Humor*. Mouton de Gruyter, Berlin.
- Attardo, S. and Raskin, V. (1991). Script theory revis(it)ed: joke similarity and joke representation model. *Humour*, 4(3):293–347.
- Barsoux, J. L. (1994). *Funny Business*. Cassel.
- Bergson, H. (1899). *Le rire. Essai sur la signification du comique*. Paris.
- Binsted, K. and Ritchie, G. (1994). An implemented model of punning riddles. In *Proc. of the 12th National Conference on Artificial Intelligence (AAAI-94)*, Seattle.
- Binsted, K. and Ritchie, G. (1997). Computational rules for punning riddles. *Humor*, 10(1).
- Ephratt, M. (1990). What's in a joke. In Golumbic, M., editor, *Advances in AI: Natural Language and Knowledge Based Systems*, pages 406–450. Springer Verlag.
- Fellbaum, C. (1998). *WordNet. An Electronic Lexical Database*. The MIT Press.
- Freud, S. (1905). *Der Witz und Seine Beziehung sum Unbewussten*. Deutike, Leipzig and Vienna.
- Gibson, D. (1994). Humor consulting laughs for power and profits in organizations. *Humor*, 7(4):403–428.
- Giora, R. and Fein, O. (1999). Irony: Context and salience. *Metaphor and Symbol*, 14:241–257.

- Goldberg, K., Roeder, T., Gupta, D., and Perkins, C. (2001). Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 4(2):133–151.
- Gupta, D., Digiovanni, M., Narita, H., and Goldberg, K. (1999). ester: Efficient rating prediction for statistical joke retrieval. Technical report, UC Berkeley. <http://shadow.ieor.berkeley.edu/humor>.
- Hofstadter, D., Gabora, L., Raskin, V., and Attardo, S. (1989). Synopsis of the workshop on humor and cognition. *Humor*, 2(4):293–347.
- Katz, B. F. (1993). A neural resolution of the incongruity and incongruity-resolution theories of humor. *Connection Science*, 5(1):59–75.
- Köhler, G. and Ruch, W. (1996). Sources of variance in current sense of humor inventories: How much substance, how much method variance? *Humor: International Journal of Humor Research*, 9:363–397.
- Köstler, A. (1964). *The act of creation*. Hutchinson, London.
- McGhee, P. (1989). *Humor and Children's Development*. Haworth Press.
- Miller, G. (1990). An on-line lexical database. *International Journal of Lexicography*, 13(4):235–312.
- Minsky, M. (1980). Jokes and the logic of the cognitive unconscious. Technical report, MIT Artificial Intelligence Laboratory. AI memo 603.
- Morkes, J., Kerna, H., and Nass, C. (1998). Humor in task-oriented computer-mediated communication and human-computer interaction. In *Proc. of CHI 98*.
- Morkes, J., Kernal, H., and Nass, C. (1999). Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of SRCT theory. *Human-Computer Interaction*, 14:395–435.
- Palma, P. D. and Weiner, E. J. (1992). Riddles: accessibility and knowledge representation. In *Proc. of the 15th International Conference on Computational Linguistics (COLING-92)*, volume 4, pages 1121–1125, Nantes.
- Pepicello, W. and Green, T. (1984). *The language of the riddles*. Ohio State University.
- Raskin, V. (1985). *Semantic Mechanisms of Humor*. Dordrecht/Boston/Lancaster.
- Ruch, W. (1996). Special issue: Measurement approaches to the sense of humor. *Humor: International Journal of Humor Research*, 9.
- Stock, O. (1996). "Password Swordfish": Verbal humor in the interface. In Hulstijn, J. and Nijholt, A., editors, *Proc. of International Workshop on Computational Humour (TWLT 12)*, University of Twente, Enschede, Netherlands.
- Utsumi, A. (1996). Implicit display theory of verbal irony: Towards a computational theory of irony. In Hulstijn, J. and Nijholt, A., editors, *Proc. of International Workshop on Computational Humour (TWLT 12)*, University of Twente, Enschede, Netherlands.

given lexicon, while verbal humour of the second kind would represent an open set relying on serendipity, the ability to creatively establish links based on world knowledge, or the ability to construct a humorous scenario.

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

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

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

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

2 TWO SETS OF EXAMPLES

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

I hate chemistry, said Tom *acidly*.

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

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

I think I've caught one, said Tom with <i>bated</i> breath.	(cf. bated breath)
I have to cut the grass again, Tom <i>moaned</i> .	(cf. mow-ned)

but all share the same essential characteristics.

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

categories, including the following:

Number of bridges

m = multiple semantic bridges

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

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

syn = a semantic bridge based on synonym

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

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

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

hyp = a semantic bridge based on hyponymy

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

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

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

seq = sequential or cyclic relations

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

Types of complex lexical or semantic relations

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

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

enc = a relation in based on world knowledge

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

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

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

One upon a time there was a beautiful princess

Tom began *grimly*. (Grimms' fairy tales)

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

Look above the stable, Tom said loftily.
I like the spacious new apartment, Tom said loftily.
I'm a hardworking stonemason, said Tom callously.
And they lived happily ever after, said Tom grimly.
Once upon a time there were two brothers, said Tom grimly.

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

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

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

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

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

John is violent. He is razing cattle.
John ate a dollar coin. He is passing the buck.

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

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

- a) As a first step, seventy adjectives with noun homophones and seventy nouns with noun homophones were taken from a list on the web (Cooper, 1999). Since these contained only homophones spelled differently, they were supplemented from another source (Franklyn, 1966). Sixteen words were obscure enough that the persons judging the puns would be unlikely to recognize them. These were removed.
- b) The second step involved picking collocations of these words from the Oxford English dictionary. Each of the homophones was looked up, and the first common phrase agreeing with our syntax requirements was chosen. If a common phrase with the proper syntax could not be found for a word or its homophone, both words were deleted. The resulting phrases were added to the lexicon as units, as were the individual words of the phrases.
- c) For each of the words, volunteers were given a questionnaire and asked to provide words related to the various nouns, adjectives and verbs in the lexicon. It is interesting to note the diversity which this produced. Thus, apart from the usual basic lexical relations like synonymy, antonymy, meronymy, etc., we find encyclopedic relationships such as *sailor* – *pier*, *diver* – *coral*, *grave* – *bier*. In the first two cases, the relationship between the two words is based on typical locations associated with the first term, while in the third example, the relationship is based on shared membership in a common semantic domain (cemetery and burial rituals and objects).
- d) In a last step, other words were added to the lexicon, including determiners, conjunctions, prepositions, etc. called for by the syntax specifications.

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

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

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

1-2

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

2-3

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

3-4

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

4-5

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

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

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

<i>Generated pun</i>	<i>Lexical relations</i>	<i>Avg score</i>	<i>Standard deviation</i>
The housewife captures a murderer. A cereal killer.	Encyclopedic/synonymy	1.43	0.5
Joan looks at a musical sign on the station. A bass clef.	Hyponymy/encyclopedic	1.43	0.76
John gives a peck in the home. A buss shelter.	Synonymy/hyponymy	1.71	1.46
The chimney sweep digs a burrow. A grate hole.	Encyclopedic/synonymy	1.86	1
Joan visits a relation on the mound. An aunt hill.	Hyponymy/synonymy	1.88	0.5
Joan visits a grave in the basement. A bier cellar.	Encyclopedic/synonymy	1.88	0.82
The butcher commits a carelessness. A gross negligence.	Encyclopedic/synonymy	2	1.46
The social worker hates a foe. A hostel enemy.	Encyclopedic/synonymy	2	1.07
The cleaner loves an individual. A pail person.	Encyclopedic/synonymy	2	1.34
The vocalist arrives at a settlement. A bass camp.	Hyponymy/synonymy	2.13	0.69
Joan bears stress on the boardwalk. Pier pressure.	Hyponymy/synonymy	4	1.15
The textile worker fulfils a requirement. A dyer need.	Encyclopedic/synonymy	4	0.75
The general performs an operation. A major surgery.	Sequence/synonymy	4	0.5
Joan kisses a hero at the disco. A knight club.	Encyclopedic/synonymy	4.25	0.82
The sailor bears a stress. Pier pressure.	Encyclopedic/synonymy	4.38	0.47

3 DISCUSSION

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

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

It's freezing in here, said Tom coldly

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

I'm from a Humberside port, said Tom ghoulishly

relies on the reader or hearer knowing that there is a city called Goole in this region, knowledge

that is far from universal. If this knowledge is not available to the hearer or reader of the pun, the pun will fail.

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

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

As an initial working hypothesis, we could suggest that if the type of lexical relation is held constant, then the number of active links in a pun will be correlated with its cleverness. In a second test, if the number of links is held constant, we can ask whether cleverness is correlated with different types of links. For example, it appears, intuitively, that synonym-based Tom Swifties are less clever than those based on more complex lexical relations.

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

Tony Blair MP – I'm Tory Plan B

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

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

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

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

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

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

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

REFERENCES

- Binsted, K. (1996). Machine humour: An implemented model of puns. Dissertation, University of Edinburgh, Scotland.
- Binsted, K. and G. Ritchie (1994). A symbolic description of punning riddles and its computer implementation. Research Paper 688, University of Edinburgh, Scotland.
- Cooper, A. (1999) Alan Cooper's Homonyms. Available at <http://www.cooper.com/alan/homonym.html>.
- Cruse, D.A. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.
- Franklyn, J.H. (1966) (Ed.) *Which Witch? Being a grouping of phonetically compatible words*. Hamish Hamilton, London, 1966.
- Lenat, D., R.V. Gupta (1989). *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project*. Reading, MA: Addison-Wesley.
- Lessard, G. and M. Levison (1992). Computational Modelling of Linguistic Humour. *ALLC/ACH92*, Oxford, England.
- Lessard, G. and M. Levison (1993). Computational Modelling of Riddle Strategies. *ACH/ALLC93*, Georgetown, USA.
- Lessard, G. and M. Levison (1995). Linguistic and cognitive underpinnings of verbal humour. *International Cognitive Linguistics Association conference*, Albuquerque, NM.
- Levison, M. and G. Lessard (1992). A System for Natural Language Generation. *Computers and the Humanities*, 26:43-58.
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.

Raskin, V. (1985). *The Semantic Mechanisms of Humour*. Dordrecht, Reidel.

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

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.
R. Leermakers (*Philips Research Labs, Eindhoven.*) Recursive Ascent Parsing.
H. Harkema & M. Tomita (*University of Twente, Enschede & Carnegie Mellon University, Pittsburgh*)
A Parsing Algorithm for Non-Deterministic Context-Sensitive Languages.
G.J. van der Steen (*Vleermuis Software Research, Utrecht*) Unrestricted On-Line Parsing and Transduction with Graph Structured Stacks.
J. Rekers & W. Koorn (*CWI, Amsterdam & University of Amsterdam, Amsterdam*) Substring Parsing for Arbitrary Context-Free Grammars.
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 Parallellization 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.
Eds. H.J. op den Akker, A. Nijholt & W. ter Stal, 115 pages.

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.
G.J. van der Steen & A.J. Dijenborgh (*Vleermuis Software Research, Utrecht*) Lingware: The Translation Tools of the Future.
T. Vosse (*NICI, Nijmegen*) Detecting and Correcting Morpho-syntactic Errors in Real Texts.
C. Barkey (*TNO/ITI, Delft*) Indexing Large Quantities of Documents Using Computational Linguistics.
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 Deltra in Natural Language Front-ends.
J. Odijk (*Philips Research Labs, Eindhoven*) The Automatic Translation System ROSETTA3.
D. van den Akker (*IBM Research, Amsterdam*) Language Technology at IBM Nederland.
M.-J. Nederhof, C.H.A. Koster, C. Dekkers & A. van Zwol (*University of Nijmegen, Nijmegen*) The Grammar Workbench: A First Step Toward Lingware Engineering.
-

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.
- A.J.M.M. Weijters & J. Thole** (*University of Limburg, Maastricht*) Speech Synthesis with Artificial Neural Networks.
- 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. Scholtes** (*University of Amsterdam, Amsterdam*) Neural Data Oriented Parsing.
- E.F. Tjong Kim Sang** (*University of Groningen, Groningen*) A connectionist Representation for Phrase Structures.
- M.F.J. Drossaers** (*University of Twente, Enschede*) Hopfield Models as Neural-Network Acceptors.
- P. Wyard** (*British Telecom, Ipswich*) A Single Layer Higher Order Neural Net and its Application to Grammar Recognition.
- N.E. Sharkey & A.J.C. Sharkey** (*University of Exeter, Exeter*) A Modular Design for Connectionist Parsing.
- 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. Miikkulainen** (*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

Eds. D. Nauta, A. Nijholt & J. Schaake, 114 pages.

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. Ginzburg (*University of Edinburgh*) Enriching Answerhood and Truth: Questions within Situation Semantics.

J. Schaake (*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.

R. Scha (*University of Amsterdam*) Understanding Media: Language vs. Graphics.

L. Boves (*University of Nijmegen*) Spoken Language Interfaces.

J. Nerbonne (*University of Groningen*) NL Interfaces and the Turing Test.

K. Simons (*Digimaster, Amstelveen*) "Natural Language": A Working System.

P. Horsman (*Dutch National Archives, The Hague*) Accessibility of Archival Documents.

W. Sijtsma & O. Zweekhorst (*ITK, Tilburg*) Comparison and Review of Commercial Natural Language Interfaces.

J. Schaake (*University of Twente*) The Reactive Dialogue Model: Integration of Syntax, Semantics, and Pragmatics in a Functional Design.

- D. Speelman** (*University of Leuven*) A Natural Language Interface that Uses Generalised Quantifiers.
R.-J. Beun (*IPO, Eindhoven*) The DENK Program: Modeling Pragmatics in Natural Language Interfaces.
W. Menzel (*University of Hamburg*) ASL: Architectures for Speech and Language Processing
C. Huls & E. Bos (*NICI, Nijmegen*) EDWARD: A Multimodal Interface.
G. Neumann (*University of Saarbrücken*) Design Principles of the DISCO system.
O. Stock & C. Strapparava (*IRST, Trento*) NL-Based Interaction in a Multimodal Environment.
-

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.
R. Bod (*University of Amsterdam*) Data Oriented Parsing as a General Framework for Stochastic Language Processing.
M. Stefanova & W. ter Stal (*University of Sofia / University of Twente*) A Comparison of ALE and PATR: Practical Experiences.
J.P.M. de Vreught (*University of Delft*) A Practical Comparison between Parallel Tabular Recognizers.
M. Verlinden (*University of Twente*) Head-Corner Parsing of Unification Grammars: A Case Study.
M.-J. Nederhof (*University of Nijmegen*) A Multi-Disciplinary Approach to a Parsing Algorithm.
Th. Stürmer (*University of Saarbrücken*) Semantic-Oriented Chart Parsing with Defaults.
G. Satta (*University of Venice*) The Parsing Problem for Tree-Adjoining Grammars.
F. Barthélemy (*University of Lisbon*) A Single Formalism for a Wide Range of Parsers for DCGs.
E. Csuhaaj-Varjú & R. Abo-Alez (*Hungarian Academy of Sciences, Budapest*) Multi-Agent Systems in Natural Language Processing.
C. Cremers (*University of Leiden*) Coordination as a Parsing Problem.
M. Wirén (*University of Saarbrücken*) Bounded Incremental Parsing.
V. Kubon & M. Platek (*Charles University, Prague*) Robust Parsing and Grammar Checking of Free Word Order Languages.
V. Srinivasan (*University of Mainz*) Punctuation and Parsing of Real-World Texts.
T.G. Vosse (*University of Leiden*) Robust GLR Parsing for Grammar-Based Spelling Correction.
-

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.

Proceedings Twente Workshop on Language Technology 7 (TWLT 7)

Computer-Assisted Language Learning

Eds. L. Appelo, F.M.G. de Jong, 133 pages.

Preface and Contents

L.Appelo, F.M.G. de Jong (*IPO / University of Twente*) Computer-Assisted Language Learning: Prolegomena

M. van Bodegom (*Eurolinguist Language House, Nijmegen, The Netherlands*) Eurolinguist test: An adaptive testing system.

B. Cartigny (*Escape, Tilburg, The Netherlands*) Discatex CD-ROM XA.

H.Altay Guvenir, K. Oflazer (*Bilkent University, Ankara*) Using a Corpus for Teaching Turkish Morphology.

H. Hamburger (*GMU, Washington, USA*) Viewpoint Abstraction: a Key to Conversational Learning.

J. Jaspers, G. Kanselaar, W. Kok (*University of Utrecht, The Netherlands*) Learning English with It's English.

G. Kempen, A. Dijkstra (*University of Leiden, The Netherlands*) Towards an integrated system for spelling, grammar and writing instruction.

F. Kronenberg, A. Krueger, P. Ludewig (*University of Osnabruek, Germany*) Contextual vocabulary learning with CAVOL.

S. Lobbe (*Rotterdam Polytechnic Informatica Centrum, The Netherlands*) Teachers, Students and IT: how to get teachers to integrate IT into the (language) curriculum.

J. Rous, L. Appelo (*Institute for Perception Research, Eindhoven, The Netherlands*) APPEAL: Interactive language learning in a multimedia environment.

B. Salverda (*SLO, Enschede, The Netherlands*) Developing a Multimedia Course for Learning Dutch as a Second Language.

C. Schwind (*Universite de Marseille, France*) Error analysis and explanation in knowledge based language tutoring.

J. Thompson (*CTI, Hull, United Kingdom/EUROCALL*) TELL into the mainstream curriculum.

M. Zock (*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
- P. van Alphen, C. in't Veld & W. Schelvis** (*PTT Research, Leidschendam, The Netherlands*) Analysis of the Dutch Polyphone Corpus.
- H.J.M. Steenken & D.A. van Leeuwen** (*TNO Human factors Research, Soesterberg, The Netherlands*) Assessment of Speech Recognition Systems.
- J.M. McQueen** (*Max Planck Institute, Nijmegen, The Netherlands*) The Role of Prosody in Human Speech Recognition.
- L. ten Bosch** (*IPO, Eindhoven, the Netherlands*) The Potential Role of Prosody in Automatic Speech Recognition.
- P. Baggia, E. Gerbino, E. Giachin & C. Rullent** (*CSELT, Torino, Italy*) Spontaneous Speech Phenomena in Naive-User Interactions.
- 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.
- K. van Deemter, J. Landsbergen, R. Leermakers & J. Odijk** (*IPO, Eindhoven, The Netherlands*) Generation of Spoken Monologues by Means of Templates
- D. Carter & M. Rayner** (*SRI International, Cambridge, UK*) The Speech-Language Interface in the Spoken Language Translator
- H. Weber** (*University of Erlangen, Germany*) Time-synchronous Chart Parsing of Speech Integrating Unification Grammars with Statistics.
- 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

Eds. J.A. Andernach, S.P. van de Burgt & G.F. van der Hoeven, 124 pages.

Preface and Contents

N. Dahlbäck (*NLP Laboratory, Linköping, Sweden*) Kinds of agents and types of dialogues.

- J.H. Connolly, A.A. Clarke, S.W. Garner & H.K. Palmén** (*Loughborough University of Technology, UK*) Clause-internal structure in spoken dialogue.
- J. Carletta, A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon & A. Anderson** (*HCRC, Edinburgh, UK*) The coding of dialogue structure in a corpus.
- 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. Andernach** (*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 & R. Steetskamp, 263 pages.

Preface and Contents

Teodor Rus (*Iowa City, USA*) Algebraic Processing of Programming Languages.

Eelco Visser (*Amsterdam, NL*) Polymorphic Syntax Definition.

J.C. Ramalho, J.J. Almeida & P.R. Henriques (*Braga, P*) Algebraic Specification of Documents.

Teodor Rus & James, S. Jones (*Iowa City, USA*) Multi-layered Pipeline Parsing from Multi-axiom Grammars.

Klaas Sikkel (*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 Dynamic Programming.

Michael Moortgat (*Utrecht, NL*) Multimodal Linguistic Inference.

R.C. Berwick (*MIT, USA*) Computational Minimalism: The Convergence of the Minimalistic Syntactic Program and Categorical 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.

Wojciech Buszkowski (*Poznan, P*) Algebraic Methods in Categorical Grammar.

- Vladimir A. Fomichov** (*Moscow, R*) A Variant of a Universal Metagrammar of Conceptual Structures. Algebraic Systems of Conceptual Syntax.
- Theo M.V. Jansen** (*Amsterdam, NL*) The Method of ROSETTA, Natural Language Translation Using Algebras.
- C.M. Martín-Vide, J. Miquel-Verges & Gh. Paun** (*Tarragona, E*) Contextual Grammars with Depth-First Derivation.
- Pál Dömösi & Jürgen Duske** (*Kussuth University H, University of Hannover, G*) Subword Membership Problem for Linear Indexed Languages.
- C. Rico Perez & J.S. Granda** (*Madrid, E*) Algebraic Methods for Anaphora Resolution.
- Vincenzo Manca** (*Pisa, I*) A Logical Formalism for Intergrammatical Representation.
-

The eleventh workshop in the series took place on 19-21 June 1996. It focussed 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

- David R. Traum** (*Université de Genève, CH*) Conversational Agency: The TRAINS-93 Dialogue Manager.
- Scott McGlashan** (*SICS, SW*) Towards Multimodal Dialogue Management.
- Pierre Nugues, Christophe Godéreaux, Pierre-Olivier and Frédéric Revolva** (*GREYC, F*) A Conversational Agent to Navigate in Virtual Worlds.
- Anne Vilnat** (*LIMSI-CNRS, F*) Which Processes to Manage Human-Machine Dialogue?
- Susann LuperFoy** (*MITRE, USA*) Tutoring versus Training: A Mediating Dialogue Manager for Spoken Language Systems.
- David G. Novick & Stephen Sutton** (*Portland, USA*) Building on Experience: Managing Spoken Interaction through Library Subdialogues.
- Latifa Taleb** (*INRIA, F*) Communicational Deviation in Finalized Informative Dialogue Management.
- Robbert-Jan Beun** (*IPO, NL*) Speech Act Generation in Cooperative Dialogue.
- Gert Veldhuijzen van Zanten** (*IPO, NL*) Pragmatic Interpretation and Dialogue Management in Spoken-Language Systems.
- Joris Hulstijn, Renè Steetskamp, Hugo ter Doest, Anton Nijholt & Stan van de Burgt** (*University of Twente, NL & KPN Research, NL*) Topics in SCHISMA Dialogues.
- Gavin Churcher, Clive Souter & 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.
- Anders Baekgaard** (*CPK, DK*) Dialogue Management in a Generic Dialogue System.
- Wayne Ward** (*Carnegie Mellon University, USA*) Dialog Management in the CMU Spoken Language Systems Toolkit.

- Wieland Eckert** (*University of Erlangen, D*) Understanding of Spontaneous Utterances in Human-Machine-Dialog.
- Jan Alexandersson** (*DFKI, D*) Some Ideas for the Automatic Acquisition of Dialogue Structure.
- Kristiina Jokinen** (*Nara Institute of Science and Technology, JP*) Cooperative Response Planning in CDM: Reasoning about Communicative Strategies.
- Elizabeth Hinkelman** (*Kurzweil Applied Science, USA*) Dialogue Grounding for Speech Recognition Systems.
- Jennifer Chu-Carroll** (*University of Delaware, USA*) Response Generation in Collaborative Dialogue Interactions.
- Harry Bunt** (*Tilburg University, NL*) Interaction Management Functions and Context Representation Requirements.
- Peter Wyard & Sandra Williams** (*BT, GB*) Dialogue Management in a Mixed-Initiative, Cooperative, Spoken Language System.
- Rolf Carlson** (*KTH, SW*) The Dialog Component in the Waxholm System.
- Laila Dybkjær, Niels Ole Bernsen & Hans Dybkjaer** (*Roskilde University, DK*) Evaluation of Spoken Dialogue Systems.
- Vincenzo Manca** (*Pisa, I*) A Logical Formalism for Intergrammatical Representation.

TWLT 12 took place on 11-14 September 1996. It focussed on 'computational humor' and in particular on verbal humor. TWLT12 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)

Computational Humor: 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.

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

Akira Ito & 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 & Hitoshi Isahara On Computational Processing of Rhetorical Expressions - Puns, Ironies and Tautologies.

Carmen Curc3 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 & Mark Keane *Bad Vibes* Catastrophes of Goal Activation in the Appreciation of Disparagement Humour and General Poor Taste.

Kim Binsted & Graeme Ritchie Speculations on Story Pun.

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

Cameron Shelley, Toby Donaldson & Kim Parsons Humorous Analogy: Modeling 'The Devils Dictionary'.

Michal Ephratt More on Humor Act: What Sort of Speech Act is the Joke?

TWLT 13 took place on 13-15 May 1998. It was the follow-up of the Mundial workshop, that took place in München in 1997. Both the Mundial workshop as TWLT13 focussed on the formal semantics and pragmatics of dialogues. In addition to the three-day workshop in Twente, with invited and accepted papers, on 18 May a workshop titled 'Communication and Attitudes' was organized at ILLC/University of Amsterdam.

Proceedings Twente Workshop on Language Technology 13 (TWLT 13)

Formal Semantics and Pragmatics of Dialogue (Twendial'98)

Eds. J. Hulstijn and A. Nijholt, 274 pages.

Preface and Contents

Nicholas Asher Varieties of Discourse Structure in Dialogue

Jonathan Ginzburg Clarifying Utterances

Steve Pulman The TRINDI Project: Some Preliminary Themes

Henk Zeevat Contracts in the Common Ground

John Barnden Uncertain Reasoning About Agents' Beliefs and Reasoning, with special attention to Metaphorical Mental State Reports

Thomas Clermont, Marc Pomplun, Elke Prestin and Hannes Rieser Eye-movement Research and the Investigation of Dialogue Structure

Robin Cooper Mixing Situation Theory and Type Theory to Formalize Information States in Dialogue

Jean-louis Dessalles The Interplay of Desire and Necessity in Dialogue

Wieland Eckert Automatic Evaluation of Dialogue Systems

Jelle Gerbrandy Some Remarks on Distributed Knowledge

Jeroen Groenendijk Questions in Update Semantics

Wolfgang Heydrich Theory of Mutuality (Syntactic Skeleton)

Wolfgang Heydrich, Peter Kühnlein and Hannes Rieser A DRT-style Modelling of Agents' Mental States in Discourse

Staffan Larsson Questions Under Discussion and Dialogue Moves

Ian Lewin Formal Design, Verification and Simulation of Multi-Modal Dialogues

Nicolas Maudet & Fabrice Evrard A Generic framework for Dialogue Game Implementation

Soo-Jun Park, Keon-Hoe Cha, Won-Kyung Sung, Do Gyu Song, Hyun-A Lee, Jay Duke Park,

Dong-In Park & Jörg Höhle MALBOT: An Intelligent Dialogue Model using User Modeling

Massimo Poesio & David Traum Towards an Axiomatization of Dialogue Acts

Mieke Rats Making DRT Suitable for the Description of Information Exchange in a Dialogue

Robert van Rooy Modal subordination in Questions

Adam Zachary Wyner A Discourse Theory of Manner and Factive Adverbial Modification

Marc Blasband A Simple Semantic Model

TWLT14 was held on 7-8 December 1998. It focussed on the role of human language technology in the indexing and accessing of written and spoken documents, video material and/or images, and on the role of language technology for cross-language retrieval and information extraction. The workshop consisted of a series of accepted papers.

Proceedings Twente Workshop on Language Technology 14 (TWLT 14)

Language Technology in Multimedia Information Retrieval

Eds. D. Hiemstra, F.M.G. de Jong and K. Netter, 194 pages.

Preface and Contents

- Hans Uszkoreit** (*DFKI, Saarbrücken*) Cross-language information retrieval: from naive concepts to realistic applications
- Paul Buitelaar, Klaus Netter & Feiyu Xu** (*DFKI, Saarbrücken*) Integrating Different Strategies for Cross-Language Retrieval in the MIETTA Project
- Djoerd Hiemstra & Franciska de Jong** (*University of Twente*) Cross-language Retrieval in Twenty-One: using one, some or all possible translations?
- David A. Hull** (*Xerox Research Center Europe*) Information Extraction from Bilingual Corpora and its application to Machine-aided Translation
- Arjen P. de Vries** (*University of Twente*) Mirror: Multimedia Query Processing in Extensible Databases
- Douglas E. Appelt** (*SRI International*) An Overview of Information Extraction Technology and its Application to Information Retrieval
- Paul E. van der Vet & Bas van Bakel** (*University of Twente*) Combining Linguistic and Knowledge-based Engineering for Information Retrieval and Information Extraction
- Karen Sparck Jones** (*Cambridge University*) Information retrieval: how far will really simple methods take you?
- Raymond Flournoy, Hiroshi Masuichi & Stanley Peters** (*Stanford University and Fuji Xerox Co. Ltd.*) Cross-Language Information Retrieval: Some Methods and Tools
- Andrew Salway & Khurshid Ahmad** (*University of Surrey*) Talking Pictures: Indexing and Representing Video with Collateral Texts
- Wim van Bruxvoort** (*VDA informatiebeheersing*) Pop-Eye: Using Language Technology in Video Retrieval
- Istar Buscher** (*Südwestrundfunk, Baden Baden*) Going digital at SWR TV-archives: New dimensions of information management professional and public demands
- Arnold W.M. Smeulders, Theo Gevers & Martin L. Kersten** (*University of Amsterdam*) Computer vision and image search engines
- Kees van Deemter** (*University of Brighton*) Retrieving Pictures for Document Generation
- Steve Renals & Dave Abberly** (*University of Sheffield*) The THISL Spoken Document Retrieval System
- Wessel Kraaij, Joop van Gent, Rudie Ekkelenkamp & David van Leeuwen** (*TNO-TPD Delft and TNO-HFRI Soesterberg*) Phoneme Based Spoken Document Retrieval
- Jade Goldstein & Jaime Carbonell** (*Carnegie Mellon University*) The use of MMR, diversity-based reranking in document reranking and summarization
- Michael P. Oakes, Chris D. Paice** (*Lancaster University*) Evaluation of an automatic abstractic system
- Danny H. Lie** (*Carp Technologies, The Netherlands*) Sumatra: A system for Automatic Summary Generation

Marten den Uyl, Ed S. Tan, Heimo Müller & Peter Uray (*SMR Amsterdam, Vrije Universiteit Amsterdam, Joanneum Research*) Towards Automatic Indexing and Retrieval of Video Content: the VICAR system

Anton Nijholt (*University of Twente*) Access, Exploration and Visualization of Interest Communities: The VMC Case Study (in Progress)

Joanne Capstick, Abdel Kader Diagne, Gregor Erbach & Hans Uszkoreit (*DFKI, Saarbrücken*) MULINEX: Multilingual Web Search and Navigation

Klaus Netter & Franciska de Jong (*DFKI, Saarbrücken and University of Twente*) OLIVE: speech based video retrieval

Franciska de Jong (*University of Twente*) Twenty-One: a baseline for multilingual multimedia

TWLT15 was held on 19-21 May 1999. It focussed on the interactions in Virtual World. Contributions were invited on theoretical, empirical, computational, experimental, anthropological or philosophical approaches to VR environments. Invited talks were given by Russell Eames (*Microsoft*), Lewis Johnson (*USC*), James Lester (*North Carolina State University*), Pierre Nugues (*ISMRA-Caen*) and Stephan Matsuba (*VRML Dream Company*)

Proceedings Twente Workshop on Language Technology 15 (TWLT 15)

Interactions in Virtual Worlds

Eds. Anton Nijholt, Olaf Donk and Betsy van Dijk, 240 pages.

Preface and Contents

Riccardo Antonini (*University of Rome*) Let's-Improvise-Together

Philip Brey (*Twente University*) Design and the Social Ontology of Virtual Worlds

Dimitrios Charitos, Alan H. Bridges and Drakoulis Martakos (*University of Athens & University of Strathclyde*) Investigating Navigation and Orientation within Virtual Worlds

M. Cibelli, G. Costagliola, G. Polese & G. Tortora (*University of Salerno*) VR-Spider for (non) Immersive WWW navigation

Bruce Damer, Stuart Gold, Jan A.W. de Bruin & Dirk-Jan G. de Bruin (*The Contact Consortium & Tilburg University*) Steps toward Learning in Virtual World Cyberspace: TheU Virtual University and BOWorld

Russell Eames (*Microsoft Research, Seattle*) Virtual Worlds Applications

Hauke Ernst, Kai Schafer & Willi Bruns (*University of Bremen*) Creating Virtual Worlds with a Graspable User Interface

Denis Gracanin & Kecia E. Wright (*University of Southwestern Louisiana*) Virtual Reality Interface for the World Wide Web

Geert de Haan (*IPO, Eindhoven University of Technology*) The Usability of Interacting with the Virtual and the Real in COMRIS

G. M. P. O'Hare, A. J. Murphy, T. Delahunty & K. Sewell (*University College Dublin*) ECHOES: A Collaborative Virtual Training Environment

Mikael Jakobsson (*Ume University*) Why Bill was killed – understanding social interaction in virtual worlds

W. Lewis Johnson (*Marina del Rey*) Natural Interaction with Pedagogical Agents in Virtual Worlds

Kulwinder Kaur Deol, Alistair Sutcliffe & Neil Maiden (*City University London*) Modelling Interaction to Inform Information Requirements in Virtual Environments

- Rainer Kuhn & Sigrun Gujonsdottir** (*University of Karlsruhe*) Virtual Campus Project – A Framework for a 3D Multimedia Educational Environment
- James C. Lester** (*North Carolina State University*) Natural Language Generation in Multimodal Learning Environments
- Stephen N. Matsuba** (*The VRML Dream Company, Vancouver*) Lifelike Agents and 3D Animated Explanation Generation Speaking Spaces: Virtual Reality, Artificial Intelligence and the Construction of Meaning
- Pierre Nugues** (*ISMRA-Caen*) Verbal and Written Interaction in Virtual Worlds – Some application examples
- Anton Nijholt** (*University of Twente*) The Twente Virtual Theatre Environment: Agents and Interactions
- Sandy Ressler, Brian Antonishek, Qiming Wang, Afzal Godil & Keith Stouffer** (*National Institute of Standards and Technology*) When Worlds Collide –Interactions between the Virtual and the Real
- Lakshmi Sastry & D. R. S. Boyd** (*Rutherford Appleton Laboratory*) EISCAT Virtual Reality Training Simulation: A Study on Usability and Effectiveness
- Frank Schaap** (*University of Amsterdam*) "Males say 'blue,' females say 'aqua,' 'sapphire,' and 'dark navy'" The Importance of Gender in Computer-Mediated Communication
- Boris van Schooten, Olaf Donk & Job Zwiers** (*University of Twente*) Modelling Interaction in Virtual Environments using Process Algebra
- Martijn J. Schuemie & Charles A. P. G. van der Mast** (*Delft University of Technology*) Presence: Interacting in Virtual Reality?
- Jarke J. van Wijk, Bauke de Vries & Cornelius W. A. M. van Overveld** (*Eindhoven University of Technology*) Towards an Understanding 3D Virtual Reality Architectural Design System
- Peter J. Wyard & Gavin E. Churcher** (*BT Laboratories, Ipswich, Suffolk*) Spoken Language Interaction with a Virtual World in the MUESLI Multimodal 3D Retail System
- J. M. Zheng, K. W. Chan & I. Gibson** (*University of Hong Kong*) Real Time Gesture Based 3D Graphics User Interface for CAD Modelling System

TWLT16/AMiLP 2000 is the second AMAST workshop on Algebraic Methods in Language Processing. Like its predecessor, organized in 1995 at the University of Twente in the TWLT series, papers were presented on formal language theory, programming language theory and natural language theory. A common theme in these papers was the use of mathematics, in particular the use of an algebraic approach. AMiLP 2000 was held in Iowa City, Iowa, USA, from May 20–22 2000, just before the AMAST 2000 conference.

Proceedings Twente Workshop on Language Technology 16 (TWLT 16)
Algebraic Methods in Language Processing (AMiLP2000)
 Eds. D. Heylen, A. Nijholt, and G. Scollo, 274 pages.

Preface and Contents

- Peter R.J. Asveld** (*University of Twente*) Algebraic Aspects of Families of Fuzzy Languages
- P. Boullier** (*INRIA-Rocquencourt*) 'Counting' with Range Concatenation Grammars
- D. Cantone, A. Formisano, E.G. Omodreo & C.G. Zarbu** (*University of Catania, University L'Aquila & University of Perugia*) Compiling Dyadic First-Order Specifications into Map Algebra

- Denys Duchier** (*Universität des Saarlandes*) A Model Eliminative Treatment of Quantifier-Free Tree Descriptions
- Theo M.V. Janssen** (*University of Amsterdam*) An Algebraic Approach to Grammatical Theories for Natural Languages
- Aravind K. Joshi** (*University of Pennsylvania*) Strong Generative Power of Formal Systems
- Jozef Kelemen, Alica Kelemenov & Carlos Martin-Vide** (*Silesian University & Rovira I Virgili University*) On the Emergence of Infinite Languages from Finite Ones
- Stephan Kepser** (*University of Tübingen*) A Coalgebraic Modelling of Head-Driven Phrase Structure Grammar
- Hubert Dubois & Hélène Kirchner** (*LORIA-UHP & CNRS*) Objects, Rules and Strategies in ELAN
- Jens Michaelis & Uwe Mönnich & Frank Morawietz** (*Universität Tübingen*) Algebraic Description of Derivational Minimalism
- Rani Nelken & Nissim Francez** (*Dept. of Computer Science, Technion, Israel*) A Calculus for Interrogatives Based on Their Algebraic Semantics
- Gheorghe Păun** (*Institute of Mathematics of the Romanian Academy*) Molecular Computing and Formal Languages: A Mutually Beneficial Cooperation
- G. Reggio, M. Cerioli & E. Astesiano** (*University of Genova*) An Algebraic Semantics of UML Supporting its Multiview Approach
- James Rogers** (*University of Central Florida*) wMSO Theories as Grammar Formalisms.
- Teodor Rus** (*University of Iowa*) Algebraic Definition of Programming Languages.
- Karl-Michael Schneider** (*University of Passau*) Tabular Parsing and Algebraic Transformations
- Edward Stabler & Ed Keenan** (*University of California*) Structural Similarity
- Thomas L. Cornell** (*Cymfony Net, Inc.*) Parsing and Grammar Engineering with Tree Automata.

TWLT 17, *Interacting Agents*, was co-organised with the Centre for Evolutionary Language Engineering (CELE) and as such is also the first workshop in the CELE Workshops on Evolutionary Language Engineering (CEvoLE) series. Together with TWLT18/CEvoLE2 these workshops are jointly titled “Learning to Behave”. The workshops investigate human-agent interaction and knowledge both on the level of agents communicating with the external environment and on the level of the internal agent processes that guide the modelling and understanding of the external sensory input in the brain.

TWLT 17 focussed on the interaction of an agent with the environment. This covers many topics such as the use of conversational strategies, natural language and non-verbal communication, turn-taking, protocols, cross-media references, the effect of context, affect and emotion in agents. A special session is devoted to theatre applications. TWLT 17 was held in Enschede from October 18–20 2000.

Proceedings Twente Workshop on Language Technology 17 (TWLT 17)

Learning to Behave, Workshop I: Interacting Agents

Eds. A. Nijholt, D. Heylen and K. Jokinen, 205 pages.

Preface and Contents

Nadia Magnenat-Thalmann, Sumedha Kshirsagar (*MIRALab, CUI, University of Geneva*) Communication with Autonomous Virtual Humans

- Sabine S. Geldof** (*AI-Lab, Vrije Universiteit Brussel*) Co-habited Mixed Reality and Context
- Zsófia Ruttkay, Jeroen Hendrix, Paul ten Hagen, Alban Lelièvre, Han Noot & Behr de Ruiter** (*Centre for Mathematics and Computer Science, Amsterdam*) A Facial Repertoire for Avatars
- Kristina Höök** (*SICS, Kista, Sweden*) Evaluating Interactive Characters – going beyond body language
- Frédéric Kaplan** (*Sony Computer Science Laboratory, Paris*) Talking AIBO : First Experimentation of Verbal Interactions with an Autonomous Four-legged Robot
- Jan-Thorsten Milde** (*Universität Bielefeld*) The Instructable Agent Lokutor
- Patrizia Paggio & Bart Jongejan** (*Center for Sprogteknologi, Copenhagen*) Unification-Based Multimodal Analysis in a 3D Virtual World: the Staging Project
- Boris van Schooten** (*Parlevink Group, University of Twente, Enschede*) A Specification Technique for Building Interface Agents in a Web Environment
- Bernard Spanlang, Tzvetomir I Vassilev** (*Department of Computer Science, University College, London*) Efficient Cloth Model for Dressing Animated Virtual People
- Luc van Tichelen and Alex Schoenmakers** (*Lernout & Hauspie Speech Products, Ieper*) A Conversational Agent for a Multi-Modal Information Kiosk
- Job Zwiers, Betsy van Dijk, Anton Nijholt & Rieks op den Akker** (*Parlevink/CTIT, University of Twente, Enschede*) Design Issues for Intelligent Navigation and Assistance Agents in Virtual Environment
- Marc Cavazza** (*University of Teesside*) Mapping Dialogue Acts to Narrative Functions for Interactive Storytelling
- Ricardo Imbert & Angélica de Antonio** (*Universidad Politécnica de Madrid*) The Bunny Dilemma: Stepping Between Agents and Avatars
- Patrizia Palamidese** (*CNR-CNUCE, Pisa*) Composing and Editing Structured 3D Stories
- Rachel Price, Chraig Douthier, Mervyn A. Jack** (*Center for CIR, University of Edinburgh*) Using Choreographed Electronic Motion to Create Mood in Virtual Environments
- Nikitas M. Sgouros** (*University of Piraeus*) Empowering the Audience in Virtual Performances
- Elisabeth André** (*DFKI, Germany*) Adding Lifelike Synthetic Characters to the Web
- Roel Vertegaal** (*Queens University, Kingston*) Agents as Attention-based Interfaces

TWLT 18, *Internalising Knowledge*, was co-organised with the Centre for Evolutionary Language Engineering (CELE) and as such is also the second workshop in the CELE Workshops on Evolutionary Language Engineering (CEvoLE) series. Together with TWLT17/CEvoLE1 these workshops are jointly titled “Learning to Behave”. The workshops investigate human-agent interaction and knowledge both on the level of agents communicating with the external environment and on the level of the internal agent processes that guide the modelling and understanding of the external sensory input in the brain.

TWLT 18 focussed on internal aspects of learning and interaction: computation in brain-like systems. The goal is to investigate cognitive models for information processing and coordination, especially how symbolic processing, conceptualising and language learning take place in neural models.

TWLT 18 was held in Ieper (Belgium) from November 22-24 2000.

Eds. K. Jokinen, D. Heylen and A. Nijholt, 221 pages.

Preface and Contents

- Joseph Bruce and Risto Miikulainen** (*University of Texas at Austin, Austin*) Evolving Populations of Expert Neural Networks
- Charlotte K. Hemelrijk** (*University of Zürich*) Social Phenomena Emerging by Self-organisation in a Competitive, Virtual World (“Domworld”)
- Marc Leman** (*IPEM – Dept. Of Musicology, Ghent University*) Spatio-temporal Processing of Musical Texture, Pitch/Tonality and Rhythm
- Erik Robert** (*AZ MM Gent, Belgium; KaHoG, Gent/Foundation “Brain and Behaviour” Erasmus University, Rotterdam*) Psycholinguistic Assessments of Language Processing in Aphasia
- Luc Berthouze and Nadia Bianchi-Berthouze** (*Electrotechnical Laboratory, Tsukuba & Aizu University, Aizu-Wakamatsu*) Dynamics of Sensorimotor Categorization and Language Formation: A Co-evolution?
- L. Andrew Coward** (*Murdoch University, Perth*) Modeling Cognitive Processes with the Recommendation Architecture
- A. Dhunay, C.J.Hinde & J.H.Connolly** (*Loughborough University, UK*) Induction of a Second Language by Transformation and Augmentation
- Thomas Liebscher** (*University of Potsdam*) Modelling Reaction Times with Neural Networks using Leaky Integrator Units
- Jaime J. Dávila** (*School of Cognitive Science, Amherst*) Genetic Algorithms and the Evolution of Neural Networks for Language Processing
- Yvan Saeys & Herwig Van Marck** (*Sail Port Western Europe, Centre for Evolutionary Language Engineering (CELE), Ieper*) A Study and Improvement of the Genetic Algorithm in the CAM-Brain Machine
- Peter beim Graben, Thomas Liebscher & Douglas Saddy** (*University of Potsdam*) Parsing Ambiguous Context-Free Languages by Dynamical Systems: Disambiguation and Phase Transitions in Neural Networks with Evidence from Event-Related Brain Potentials
- Aard-Jan van Kesteren, Rieks op den Akker, Mannes Poel & Anton Nijholt** (*University of Twente*) Simulation of Emotions of Agents in Virtual Environments using Neural Networks
- Régine Kolinsky, Vincent Goetry, Monique Radeau and José Morais** (*Unité de Université Libre de Bruxelles, Fonds National de la Recherche Scientifique*) Human Cognitive Processes in Speech Perception and World Recognition
- Mehdi Dastani, Bipin Indurkha and Remko Scha** (*Vrije Universiteit Amsterdam and Tokyo University of Agriculture and Technology*) Modelling Analogical Projection based on Pattern Perception
- Marc Leman & Bart Verbeke** (*Ghent University*) The Concept of Minimal ‘Energy’ Change (MEC) in Relation to Fourier Transform, Auto-Correlation, Wavelets, AMDF and Brain-like Timing Networks – Application to the Recognition of Reptitive Rhythmical Patterns in Acoustical Musical Signals
- Jakub Zavrel, Sven Degroeve, Anne Kool, Walter Daelemans & Kristiina Jokinen** (*University of Antwerp & CELE, Ieper*) Diverse Classifiers for NLP Disambiguation Tasks. Comparison, Optimization, Combination and Evolution

TWLT 19, *Information Extraction in Molecular Biology*, was organised in the context of the Scientific Programme on Integrated Approaches for Functional Genomics funded by the European

Science Foundation (ESF). In TWLT19, *Information Extraction for Molecular Biology*, the focus is on the published literature. Authors of journal articles identify relations between genes, regulators and proteins, and each such relation is a piece of the puzzle. We do not know most pathways, but we do know that more systematic perusal of the available literature will no doubt uncover many more pieces of the puzzle. It is therefore in order to bring together workers in relevant disciplines to exchange ideas, views and approaches. The main goal of TWLT19 is to further informal and formal collaborations between different disciplines in order to promote better use of existing literature.

TWLT 19 was held in Enschede from November 11-14 2001.

Proceedings Twente Workshop on Language Technology 19 (TWLT 19)

Information Extraction in Molecular Biology

Eds. P. van der Vet, G. van Ommen, A. Nijholt and A. Valencia, 78 pages.

Preface and Contents

Pieter Adriaans (*University of Amsterdam*) Sematic Induction with Emile, Opportunities in Bioinformatics

Sophia Ananiadou, Goran Nenadić & Hideki Mima (*University of Salford, UK & University of Tokyo, Japan*) A Terminology Management Workbench for Molecular Biology

Christian Blaschke, Juan Carlos Oliveros, Alfonso Valencia & Luis Cornide (*Universidad Autónoma, Madrid, Spain & ALMA Bioinformatica, Tres Cantos, Spain*) Biological Function and DNA Expression Arrays

Udo Hahn, Stefan Schulz & Martin Romacker (*Freiburg University, Freiburg, Germany & Freiburg University Hospital, Freiburg, Germany*) 'Deep' Information Extraction From Biomedical Documents in the MEDSYNDIKATE System

Jerry R. Hobbs (*Artificial Intelligence Center, SRI International, California*) Information Extraction from Biomedical Text

U. Reyle & J. Šarić (*University of Stuttgart, Germany & European Media Laboratory, Heidelberg, Germany*) Ontology Driven Information Extraction

B.J. Stapley, L.A. Kelley & M.J.E. Sternberg (*Imperial Cancer Research Fund, London, UK*) Protein Functional Classification by Text Data-Mining

The proceedings of the workshops can be ordered from Leerstoel TKI, Department of Computer Science, University of Twente, P.O. Box 217, NL-7500 AE Enschede, The Netherlands. E-mail orders are possible: bijron@cs.utwente.nl. Each of the proceedings costs EURO 25,=.