Tightening up Joke Structure: Not by Length Alone

Christian F. Hempelmann (chempelm@purdue.edu)
CERIAS, Purdue University
656 Oval Dr., West Lafayette, IN 47907-2086 USA

Julia M. Taylor (jtaylor1@purdue.edu)
Department of Computer and Information Technology, Purdue University
401 N. Grant St., West Lafayette, IN 47907-2021 USA

Victor Raskin (vraskin@purdue.edu)
CERIAS, Purdue University
656 Oval Dr., West Lafayette, IN 47907-2086 USA

Abstract
The paper seeks to tighten up the notion of joke structure in the context of the Ontological Semantic Theory of Humor for computational use. The method is testing the prior hypothesis that a minimalist version of a joke, consisting only of the setup and punch line, is the most effective one. A small 'human computation' pilot study casts serious doubt on this hypothesis.

Keywords: humor research; (minimalist) joke structure; setup; punch line; joke versions.

1. Introduction
The human ability to communicate is incomplete without humor. If a computational system is ever to approximate this human communicative ability and act as a competent partner in a conversation with a human, humor must be accounted for. Over the last decades, humor research has become an intense multidisciplinary effort with significant contributions from linguistics, psychology, sociology, neuro- and cognitive sciences (Raskin 1985, 2008; Ruch 1998, Oring 1992, Davies 1990, Attardo 1994, Morreall 1983). Along with theories and analyses of human-generated and perceived humor, since the early 1990s, there have been more explorations of computational humor as well, starting with attempts at humor generation through humor detection to semantically based systems (see Section 2.1 below).

Part of the difficulty in relating computational humor-generation and, to a lesser degree, humor-detection systems to human appreciation is the question how much information has to be present in the text of a joke to ensure a successful setup and the most effective punch line. This is precisely what has not been addressed yet on the computational front—how much information is enough and not too much to carry a joke without risking the opposite extremes of crypticality or verbosity.

Two related goals of this line of research are a) to create an NLP system capable of understanding the mechanism of a joke at a level sufficient for providing a punch line to a human-generated setup (even if unintentionally) and b), conversely, for the computer to react competently to a human-generated punch line that follows a setup, generated by either participant. The first scenario enables the computer to generate humor in reaction to a human cue in human-computer interaction, the second scenario lets the computer identify humor in the same scenario and enables it to react competently to it.

Most existing theories available for humor detection or generation fall short of providing the adequate support for this task. These theories are either too fine-grained to be useful or too coarse to correctly classify any given text as a joke or a non-joke. But our ontological-semantic system provides a sufficiently rich and flexible basis because it operates at the level of human text-meaning processing. In the following, we will summarize the state of the art, introduce our approach, and then discuss a pilot study assessing human appreciation of jokes in variants of different length and types of manipulation.

2. Background

2.1. State of the Art
The usefulness of and motivations for computational humor, along with its feasibility, have been intensely discussed (see Ritchie 2004, Hempelmann 2008, Taylor 2008, Strapparava et al. 2011 and references in all of these sources). The most useful work on computational humor is based on a humor theory and seeks to gain further insights, to validate, and to improve the theory, while taking advantage of its assets. Work on humor theories has a long history, and, to this day, the true multifaceted nature of humor is still being debated (Raskin 1985, Morreall 1983, Oring 1992, Ruch 1998, Davies 1990, Attardo 1994): there is no universally accepted theory of humor that explains "what is funny, why it is funny, how it is funny, when it is funny, and to whom it is funny." (Raskin 1985: 5).

The linguistic theories of humor (Raskin 1985, Attardo & Raskin 1991) have reached a level of formal representation that is adaptable for the computation of any humorous text (Raskin et al. 2009a,b). But the best-known and most-used linguistic theory of humor remains the early Script-based Semantic Theory of Humor (SSTH: Raskin 1985).
According to the SSTH, there are two conditions for a text to be humorous:
• A text has to be compatible, fully or in part, with two different scripts.
• The two scripts with which the text is compatible are opposite, and the text must overlap fully or partially with them.

The compatibility of the text with two scripts is the necessary condition for humor; the oppositeness of the scripts is the sufficient condition. The former was to be detected in the course of normal semantic analysis; the latter was not included at that point.

The central concept, that of a script, is defined as “an enriched, structured chunk of semantic information, associated with word meaning and evoked by specific words. The script is also a cognitive structure internalized by the native speaker, and it represents the native speaker’s knowledge of a small part of the world. […] Formally or technically, every script is a graph with lexical nodes and semantic links between the nodes” (Raskin 1985: 81).

Scripts were further developed, formally and computationally, in Ontological Semantics (Nirenburg and Raskin 2004, Raskin et al. 2003), and the current, third stage of the theory, the Ontological Semantic Theory of Humor (OSTH), has the functionality both to perform the computational semantic analysis that establishes the necessary compatibility of scripts and encompasses their sufficient oppositeness.

The scripts can be linguistic, general knowledge, restricted, or individual. Linguistic scripts are known to any “average,” “standard” native speaker (adult, reasonably educated, mainstream culture, etc). General knowledge scripts, such as crossing the street or going to a store, are known to a large number of people and are not affected by their use of language. Restricted knowledge scripts are known to a smaller number of people and are not affected by their use of language either. Individual scripts are “owned” by one person: an example of an individual script would be a child’s memory of her first swim.

The General Theory of Verbal Humor (GTGH: Attardo & Raskin 1991), is an extended, second-stage multidisciplinary theory of humor that is also built upon the notion of script overlap and script oppositeness. The theory, empirically verified in Ruch et al (1993), describes jokes in terms of six knowledge resources: Script Opposition (SO), informed largely by linguistics, deals with script overlap and oppositeness presented in Script-based Semantic Theory of Humor (SSTH); Logical Mechanism (LM), informed by logic and cognitive psychology, accounts for the way in which the two scripts in the joke are brought together in a faulty, but locally valid way; Situation (SI), informed by many disciplines, contains the “props” of the joke, the textual materials evoked by the scripts of the joke that are not necessarily funny; Target (TA), informed by sociology, represents any individual or group from whom humorous behavior is expected; Narrative Strategy (NS) is the rhetorical structure of the text; Language (LA) is the actual lexical, syntactic, phonological, etc., choices at the linguistic level that instantiate all the other choices. According to the GTGH, each joke can be viewed as a 6-parameter vector (Ruch et al. 1993): Joke = {SO, LM, SI, TA, NS, LA}.

2.2. Ontological Semantic Theory of Humor
Since Raskin’s (1985) definition of scripts and the general conditions for a text to be humorous, the definition and interpretation of script overlap and oppositeness have been debated (Attardo et al. 2002, Hempelmann 2003, Taylor 2008). Basing the GTGH on Ontological Semantic Technology (OST) allows a crisper definition of the necessary and sufficient conditions for verbal humor (Raskin et al. 2009a,b).

At the core of OST (Raskin et al. 2010, Hempelmann et al. 2010, Taylor and Raskin 2011, Taylor et al. 2010) are repositories of world and linguistic knowledge, acquired semi-automatically (or, rather, in hybrid automatic and human computation—see Law and von Ann 2011) and used to disambiguate the different meanings of words and sentences and to represent them comprehensively. These repositories consist of the ontology, containing language-independent concepts and relationships between them; one lexicon per supported language, containing word senses anchored in the ontology which is used to represent their meaning; and the onomasticon, which contains names of people, countries, organizations, etc., and their descriptions, also anchoring them in ontological concepts and interlinking them with its other entries.

The lexicon and ontology are used by the OST Processor, a software that produces Text Meaning Representations (TMRs) from the text that it reads. The format of TMRs conforms to the format and interpretation of the ontology. The processed TMRs are entered into the Information Repository, from which information is used for further
processing and reasoning. Recent implementation of components of the system have produced successful results.

OST progress has enabled enhanced meaning representation of all the components of the joke, shedding light even on such less linguistic ones as the Target and Narrative Strategy (see a much simplified graphic representation of Joke 4 below in Figure 1). OSTH is reconsidering the six GTVH knowledge resources with the additional emphasis on providing the ontological support to tighten and straighten their definitions and conditions of usage. However, one troubling, even if expected result, of the formalization in OSTH is the realization that the SO of GTVH and SSTH was defined inadequately. Because SO constitutes the decisive factor in determining whether a text is a joke, and thus dominates other knowledge resources, the current theories have to be modified and revised to an extent for future research, the rationale and pilot study for which we are presenting here, will help to determine.

3. Joke Variants

One of the recent discoveries, part of gaining new insights into the Narrative Strategy within OSTH, is the apparent need of some extra material right after the punch line (see Taylor 2010, 2011). More generally, by observing the coexistence of different versions of the same jokes, we realized that some extra parts of jokes, in the setup and punch line, may have specific functionalities, while others are pure ballast contributing nothing but verbosity. To the best of our knowledge, the contribution of seemingly inessential information in jokes has never been systematically studied.

Understanding the seeming importance of extra material is needed to detect the essential and necessary information for a joke to make sense and to be effective. Kuipers (2006: 204), for example, found that both in the United States and the Netherlands, longer jokes are considered generally funnier than shorter ones. Such information cannot be measured in the number of words but rather by the tightness/non-redundancy of the underlying conceptual structures. An initial approach along these lines, proposing “meaning density” as a factor in joke funniness was presented in Hempelmann (2011). What allows for testing this assumption is the fact that the same joke often exists in several attested versions. One attractively simple hypothesis may thus be that the essential information of all versions of the joke is the conceptual structure of the minimalist version of the joke. The rationale for computing the essential information is to understand the proliferation of multiple versions, of widely varying lengths and genres, of the same joke—in conversational practice, in print and on the Internet—and to test whether their “common core” can carry the joke on its own.

To illustrate this point, let us compare two versions of the following blind date joke:

(1) Danny sets up Andy to go on a blind date with Shirley, a friend of a friend of his. But Andy is a little worried about going out with someone he has never seen before. "What do I do if she's ugly?" says Andy, "I'll be stuck with her all night." "Don't worry," Danny says. "Just go up to her door and meet her first. If you like what you see, then everything goes as planned. If you don't, just shout Aaaauuuggghhh! clutch your chest and fake a heart attack." So that night, Andy knocks at Shirley's door, and when she comes out he is awe-struck at how beautiful and sexy she is. Andy's about to speak when the girl suddenly shouts, "Aaaauuuggghhh!", clutches her chest and falls to the ground.

(2) Andy is going on a blind date but is worried that she may turn out to be ugly. A friend advises him to fake a heart attack if it turns out to be the case. When Andy arrives, the door is open by a sexy and beautiful woman, who suddenly clutches her chest and falls to the ground.

Both versions contain the same scripts, roughly corresponding to the anticipation of the blind date and the actual event. The second version is minimalistic in that it contains virtually nothing that can be removed from the text without rendering it incomprehensible and useless as a joke. The first version adds much additional detail. The second version is synthesized, and it is possible that it has lost too much, and some supporting detail would actually improve it. This optimality is of crucial significance in generating a joke by a computer. To put it differently, removing information that is redundant for a plain expository text may result in a significant loss for a joke, as demonstrated in (3).

(3) Andy is going on a blind date but is worried that she may turn out to be ugly. A friend advises him to fake a heart attack then. The date turns out to be sexy and beautiful, but she suddenly clutches her chest and falls to the ground.

It has been established in humor theory that the punch line has to be short, and preferably by far to conclude the joke (see, for instance, Attardo et al. 1994). What has not been adequately researched is the punch line parameters, including its boundaries and most effective delivery mode, especially how minimalistic it can and should be. It has been suggested (see, for instance, Giora 2002) that including a familiar element within an innovative stimulus leads to more pleasure for the subjects than a purely innovative stimulus. Our preliminary research seems to indicate that, while, generally, accompanying information can be removed from the setup, some seemingly disposable elements may have to be left in the punch line.

Thus, if we compare (3) above to (4) below, the former reads more like the serious report of a somewhat funny event than as a joke, while the latter is easier to perceive as a joke.

(4) Andy is going on a blind date but is worried that she may turn out to be ugly. A friend advises him to scream and fake a heart attack then. The date turns out to be sexy and beautiful, but she suddenly screams “aaaauuhhh,” clutches her chest and falls to the ground.

The difference between these two versions is presence of the clause she suddenly screams ‘aaaauuuhhh’ in (4). Its precise contribution to the text is something that we are interested in establishing in this pilot.
4. Pilot Study

To test the hypotheses outlined in the previous section and explore general effects between joke variants to generate more hypotheses for more formal future inquiry, we created a small test corpus for a pilot study.

This corpus of fifty stimuli consists of 10 jokes found online in at least two variants differing in length, complemented by another three synthesized variants for each joke: one is the minimalist version, condensing the joke to a summary only mentioning the necessary and, presumably sufficient information for the joke to be operational; another is this minimalist version together with a dialogue element in the Narrative Structure of the joke, since we realized in creating the non-dialogue minimalist version that the joke seemed to us radically decreased in funniness; the third artificial variant was added to be just that, a control version based on the longer real variant of the joke, to see if artificial manipulation in itself affected perceived funniness. The rationale for the final version is based on the fact that jokes are folkloristic creations optimized by iterations of retelling and not owned by individual authors, a characteristic that does not hold for cut-and-paste online joke collections created to generate traffic. In sum, the five variants for each of the ten jokes are:

- long non-manipulated version
- shorter manipulated version
- non-dialogue manipulated minimalist version
- dialogue manipulated minimalist version
- longer manipulated minimalist version with paraphrasing

In this pilot study, we recruited raters for the funniness of these variants using Amazon’s Mechanical Turk, a method generally deemed valid (Buhrmester et al. 2011) and now proclaimed to be a form of ‘human computation (Law and von Ann 2011), with an incentive of $0.10 for participation. The Mechanical Turk aims to filter bots and human responders who don’t pay attention to the instructions in several ways. This includes the researcher’s ability to block certain countries, including those where non-native English issues might affect the research issue, and to select only participants who have had a certain number of approved assignments in the past. In addition to these controls, in a second pilot study we included as the only difference from the first study reported here one additional stimulus that instead of a punch line had the direction to rate it at a given level of funniness. We then excluded all responses who didn’t follow this direction under the assumption that the raters didn’t read the instructions and clicked through the responses randomly. Interestingly, there was no significant difference in the results to the initial study.

In both versions of this pilot study, the participants, of whom we aimed for 200, were directed to a survey in Qualtrics, in which they were presented with a random variant of each of the ten jokes, in random order of joke. Ideally, we would have gathered 40 ratings for each of the 50 variants, but some variation exists because raters who took under 2 minutes for their 10 stimuli or had the same rating for each stimulus were excluded, resulting in a sample of 176 participants for the version without the control stimulus presented here.

Figure 2: Mean funniness for each joke across all variants

An Analyses of Variance (ANOVA) for repeated measurements for the funniness of the ten different jokes across all variants revealed that the jokes were perceived to differ in funniness, $F(8.38, 1558.03) = 11.38$, $p < .001$. Joke 2 (“Matthew Chapter 11”) was considered significantly less funny than the other jokes, while joke 10 (cheap parking in New York City) was deemed funnier than all others (see Figure 2). Low numbers did not allow for the exclusion of these two jokes from further analysis homing in on the variants.

Figure 3: Mean funniness of variant type across all jokes

More importantly for the line of inquiry that we are pursuing here, a second ANOVA for repeated measurements for the funniness of the five variants of jokes revealed significant differences, $F(3.72, 338.48) = 7.43$, $p <$
.001. Figure 3 shows that joke variant type 3, the manipulated minimalist version without dialogue, was rated as the least funny and significantly so against all other variant types. This effect is most interesting in contrast to the other minimalist manipulated variant type 4 that does include a dialogue in its Narrative Structure. The importance of this result, privileging dialogue over exposition, warrants further investigation in our continued research.

Table 1 shows the levels of significance for the effects summarized in Figure 3.

Table 1. Within-subjects contrasts for the pairs of joke variant (v) types.

<table>
<thead>
<tr>
<th></th>
<th>v 2</th>
<th>v 3</th>
<th>v 4</th>
<th>v 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>v 1</td>
<td>2.71</td>
<td>25.91***</td>
<td>4.18*</td>
<td>1.69</td>
</tr>
<tr>
<td>v 2</td>
<td>12.13**</td>
<td>0.41</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>v 3</td>
<td>7.30**</td>
<td>19.68***</td>
<td>4.18*</td>
<td></td>
</tr>
<tr>
<td>v 4</td>
<td>1.06</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Cells contain F-statistics for the contrasts, F(1, 91). * p < .05, ** p < .01, *** p < .001.

Clearly, the minimalist versions fared worst and the longer versions overall were deemed funnier, if not significantly so. But some amount of non-essential information clearly accounts for relevant degrees of funniness of jokes.

Summary and Outlook

With respect to our initial hypotheses, our results confirm that condensing jokes so they only contain the SO-relevant information is not optimizing their funniness. Something is lost in the process and the difference between the dialogue and non-dialogue manipulated variants seems to point at the importance of that NS factor. A further speculation that should be explored based on these results is that the faultiness of the logical mechanism might no longer be sufficiently hinted at to make it retrievable, rendering the oppositeness of those variants too blunt. In terms of a classic linguistic distinction, these initial findings are pointing at the importance of performance-related factors in jokes, in contrast to purely competence-based relation of information. In other words, joke texts have an aesthetic dimension that has yet to be allocated more clearly to a part of the OSTH model in future follow-up studies with further careful manipulations of joke variants.

Acknowledgments

The authors are grateful to Ursula Beermann, Department of Psychology, University of California at Berkeley, for her help with the pilot study. They would also like to thank the IRB of Purdue University for the prompt granting of the exemption.

References


Hempelmann, C. F. (2011) Meaning Density: Explicitness is Proverbially and at This Point in This Title Also Obviously the Death of Wit, or is It? 2011 Annual Meeting of the International Society for Humor Studies—ISHS’11. Boston, MA: Boston University.


she suddenly clutches her chest and falls to the ground. Attack then. The date turns out to be beautiful and sexy, but


Appendix: Minimalist Joke Variants

1 Andy is going on a blind date but is worried that she may turn out to be ugly. A friend advises him to fake a heart attack then. The date turns out to be beautiful and sexy, but she suddenly clutches her chest and falls to the ground.

2 Matthew has been in business for many years, and suddenly the business is going down the drain. When he looked for advice by opening the Bible on a random page. It read, “Matthew. Chapter 11.”

3 A store manager overhears one of his salesmen telling to a customer that the store hasn't had something for awhile and it doesn't look like they'll be getting any soon.

The manager yells after the departing customer to come back next week because surely they'd have it by then. Irrate, he reprimands his salesman for telling a customer they're out of anything and asks what the customer wanted. It was rain.

4 A man gets pulled over by a policeman, who tells the man that his wife fell out of the car about a mile back. The man is relieved because he thought he'd gone deaf.

5 The Pope was finally persuaded by his cardinals to sleep with a woman, so that he could better understand the problems of mankind. The Pope agrees, but insists that she has to have certain qualifications: first, she has to be blind, so she cannot see who is doing it to her; second, she has to be mute, so she can’t tell anyone what happened; and third, she has to have big tits.

6 The door bell rings at the whorehouse. A girl who answers the door, sees a guy with no arms and no legs and asks what he thinks he’s going to do in there.

The guy points out that he was able to ring the bell after all.

7 On a bus, a punk kid with red, green and orange hair notices an old guy staring at him. When he asks the old man if he himself never did anything wild in his time, it turns out that the old man once screwed a parrot and was wondering whether the punk was his son.

8 One day an angel made a male and a female statues that have faced each other in a park for decades come alive to do anything they wanted for thirty minutes. The two dashed for the bushes, whose branches started shaking while there was giggling and laughter. Fifteen minutes later, they emerged with wide grins on their faces and they still had fifteen more minutes. Then the female statue said to the male that this time he should hold the pigeon down and she'll poop on its head.

9 As a woman gets on a bus with her baby, and the driver tells her that hers is the ugliest baby he’s ever seen. Angrily, she complains to a man in the rear of the bus that the driver just insulted her. The man suggests she go and tell the bus driver off and offers to hold her monkey for her.

10 A man walks into a bank in New York City and asks for a $4000 loan. The bank teller agrees to accept the man’s black Porsche parked in the bank’s parking garage as security. A few weeks later the man returns to pay off his loan and the interest of $11 dollars. The manager wonders why the man needed to borrow $4000 dollars, since the bank found out that he was a millionaire. The man replies that nowhere else in New York can he park his car for three weeks for $11 dollars.