

Towards Simulation of Semantic Generation and Detection of Humorous Response

Julia M. Taylor¹ and Vitaliy L. Rayz²

¹ Computer and Information Technology, Purdue University,
401 N Grant Street, West Lafayette, IN 47907, USA
jtaylor1@purdue.edu

² College of Engineering and Applied Science, University of Wisconsin – Milwaukee,
3200 N Cramer Street, Milwaukee, WI 53211, USA
rayz@uwm.edu

Abstract. This paper explores headlines that are so obvious that they are somehow funny. We develop a model that uses ontological representation as the basis for retrieving such obvious information. Using this model, we generate jokes in a narrow domain of fluid dynamics.

Keywords: Restricted joke modeling · Ontology · Computational humor

1 Introduction

Throughout social media, every once in a while, there are sparks of posts about headlines that are funny for one reason or another. In this paper, we are investigating a mechanism for creating and understanding headlines that state information that is so obvious that the headlines themselves become ridiculous. Such headlines have been a source of attention not just by social networking sites, such as Facebook or Twitter, but they also serve as a topic for newspaper or blog publications [1, 2] or even chapters in books [3].

These headlines, arguably, are mildly funny, as the examples below demonstrate:

- (1) Diana was still alive hours before she died.
- (2) Statistics show that teen pregnancy drops off significantly after age 25.
- (3) Federal Agents Raid Gun Shop, Find Weapons.
- (4) Homicide victims rarely talk to police.
- (5) Study Shows Frequent Sex Enhances Pregnancy Chances.
- (6) Bridges help people cross rivers.
- (7) Healthy diet lowers death risk for women.
- (8) High heels lead to foot pain.

Some of the examples are taken from satirical newspapers – for example, according to Reddit post by barmonkey, (1) originally appeared in *Private Eye*, a British satirical newspaper, making fun of the *Daily Express* – while others may be from a legitimate quote that deserve this much attention, or can be blamed on journalists not having enough time to read them carefully enough. An example of a headline that appeared from a quote

is (2), which, according to [4] is attributed to a then Colorado State Senator quoted in *The Denver Post* on 5/14/1995.

We are interested in such headlines for two reasons: one, they attract attention and comments from people, thus they are suitable for some sort of a dialogue; two, they are based on a violation of ontological defaults [5] and thus should be possible to model computationally. Interestingly, these headlines (or sentences, if we are to move into dialogues) contain the same script overlap/oppositeness (SOs) [6], obvious/non-obvious, and, thus, they don't require a full-blown implementation of a humor theory, such as the Ontological Semantic Theory of Humor (OSTH) [7].

With the exception of (2) and – possibly – (7), each of the headlines contains one event that serves as the anchoring point of a script. The event is described by some of the properties, filled with information from the sentence. Each of these properties also has the ontological default, described by the OST ontology. It is confirmation of this default that is so unusual in normal speech [8] that serves as a mechanism for getting attention from the readers and, possibly, a humorous response.

It is possible for a sentence to describe multiple events, as demonstrated by most of the examples above. The only sentence that has only one event is (6) – in the papers we are using the standard notion that is employed by OST that an event can be a verb or a noun, and not all verbs can serve as events. A careful examination of the examples shows that even when there are multiple events in the sentence, they are strongly connected to the main one. Moreover, if the event in question is to serve a role of a script, most of the supporting events would become a necessarily components of the script.

2 Scripts

A notion of script is perhaps most familiar to a knowledge representation audience from [9], which describes a restaurant script with all details that can be expected from people that visit restaurants on a regular basis and are comfortable with navigating the procedure of walking in, talking to whoever greets them, following to a table, placing an order, etc. It should be noted that some of these sub-events are optional, depending on the type of a restaurant – all of which [9] describes.

To a humor community, however, the scripts are known from Script-based Semantic Theory of Humor (SSTH) [6]. According to SSTH, a text is joke-carrying if it is compatible fully or in part with two different scripts and these scripts must somehow oppose. Moreover, the oppositeness must be unexpected.

A script can be seen as any situation that can be easily understood/described by a human being. A restaurant is only one of these. The label is not important in SSTH, it is the actual description of the situation that is of interest. Thus [6] describes a so-called doctor/lover joke, also known as patient/lover, where one of the scripts is visiting a doctor's office and the other one is having an affair (with a doctor's wife). The analysis of the joke is described in [10].

The SSTH itself is not clear on how to calculate the oppositeness, but several proposals, mostly throughout OSTH, have been made. For the purposes of this paper we will assume that oppositeness has to follow a salient property of a script or event.

Moreover, we will assume that these salient properties can be marked, either in the ontology as Ontological Semantic Technology [11–13] does, frames as FrameNet [14] does, or any other system that a reader may choose to use.

2.1 Federal Agents Raid Gun Shop, Find Weapons

Headline (3) is based on the script of RAID. We can look at this event from the point of view patrolling. According to FrameNet, the frame of patrolling describes “An individual or group, the Patrol moves through and examines a Ground in order to ensure that it is in a generally Desired-state-of-affairs, particularly that it is safe and contains no dangerous Unwanted-entity.” The core elements of the frame are:

- Desired-state-of-affairs, which the Patrol hopes to ensure by visiting Ground
- Ground, which is the area that the Patrol inspects to insure its safety
- Patrol, which is the person or group who inspects the Group to see that it is safe
- Purpose, which is the desired outcome of patrolling
- Unwanted-entity, which is an entity whose presence would impair the desirability or safety of the ground.

In addition to the core frame elements, there are some non-core ones:

- Circumstances, under which the Patrol examines the Ground
- Co-participant, that patrols along with the Patrol
- Degree, or the extent to which the examination is done
- Descriptor, which describes one of the participants of the patrolling event
- Duration, or the length of patrol
- Instrument, which is the entity that is used to scrutinize the ground
- Location, which is the position of Patrol during the act of perception
- ...

We have used FrameNet to describe the event of RAID since it is available to a general public. One can just as easily use OST for the description, however since one requires username and password to access its online resources we will proceed with the analysis from FrameNet while possible.

Let us now compare the frame of Patrolling with the information that is in the headline (3), based on the event of RAID. The intended outcome of a RAID is some finding – this finding is the purpose of a RAID. Thus, the fact that AGENTS of a RAID found something, in this case WEAPONS, is a necessary part of the script. In this case it corresponds to the Unwanted-entity element of the Patrolling. What is interesting here is the LOCATION of a RAID, which happens to be a SHOP that sells GUNS, and thus it must contain them.

The frame of Shopping contains only two core elements:

- Goods, or the entity that the Ground may contain
- Shopper, or the person who attempts to find the goods

The non-core elements are:

- Co-participant
- Degree, which identifies the amount of effort put into shopping
- Depictive, which describes a participant of shopping scenario
- Ground, which is the entity to which the Cognizer pays attention
- Purpose, which is an action that the Shopper intends to accomplish
-

By definition, a RAID is a sudden activity (not reflected in FrameNet, but would be in OST), and the purpose of a RAID is to find something that is possibly illegal or at least frowned upon. However, there is nothing surprising in gun shops to sell guns (it is the Goods that the Ground should contain), which is where the obvious/non-obvious SO comes in. It should be noted, that while it is not necessary for a computer, in this case, to detect that finding illegal weapons maybe newsworthy, it may add an extra layer of appreciation for a human, and, possibly, adds an LM [15] resource to a joke.

It is possible to explain the oppositeness of the scripts of Shopping and Raid not in terms of obvious/non-obvious, but rather in terms of expected and legal vs. illegal. However, to connect to a potentially illegal activity, one must completely ignore the location of the Patrolling frame, which happens to be a gun shop. Nevertheless, the Unwanted_entity of the Patrolling frame (with negative connotation) happens to be the desired entity of Goods in Shopping (with positive connotation), which, without a doubt, is in oppositeness with each other (see Fig. 1).

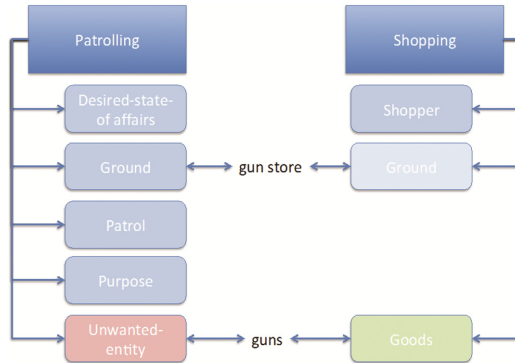


Fig. 1. Frames in the headline (3)

2.2 Homicide Victims Rarely Talk to Police

We will analyze another headline before generalizing the principle. The event that is in focus of this headline is homicide, which corresponds to the Killing frame of FrameNet. This frame has the following core elements:

- Cause, which is an inanimate entity or process that causes death
- Instrument, which is the device used to bring death about
- Killer, which is the person that causes death
- Means, or the method or action that was performed resulting in death
- Victim, which is the entity that dies as a result of the killing

We are not interested in most of the non-core elements on the frame, but what is of interest is that the result of this frame is the Victim being dead. Thus, brings the Frame of Death into the picture. The frame of Death is a subframe of Cycle_of_life_and_death which uses Biological entity that, in turn, have “naturally occurring biological processes and functions.” Since Death is the termination of the Cycle_of_life_and_death, we can assume that all processes and function have been terminated after Death.

Another frame that is activated in this headline is that of Telling. According to FrameNet, the definition of Telling is “a Speaker addresses an Addressee with a Message, which may be indirectly referred to as a Topic.” The following core elements are used in Telling:

- Addressee, which receives the Message from the Speaker
- Medium, in which the Message is expressed
- Message, or the communication produced by the Speaker
- Speaker, which is the sentient entity that produces the Message
- Topic, which is a general description of the content of the Message

Another possible frame here is Statement, which used “communicate the act of a Speaker to address a Message to some Addressee using language.” It has four core element, Message, Medium, Speaker and Topic. The Addressee is a non-core element, but for our purposes, the frames work very similarly.

The oppositeness here is, again, is in the obvious/non-obvious, rather, stated obvious vs. understood one. However, it could be argued that the opposition between the Speaker of Telling/Message, which must be alive and the Victim of Killing which must be dead. Perhaps this second pair is easier for computer to detect (see Fig. 2).

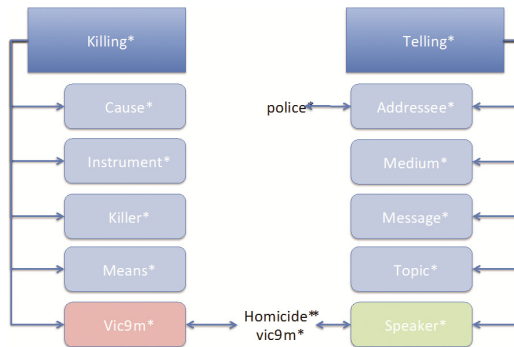


Fig. 2. Frames in the headline (4)

3 Default-Based Oppositeness

It should be noted though that very few details are filled for the core elements in both examples. Thus, it is the so-called defaults [8] that are coming into play here. Defaults are defined as obvious knowledge that does not have to be explicitly specified for the reader to be aware of it. Some example of defaults are unlocking the door (implied: with a key), talking to somebody (implied: a person, unless specified otherwise), patrolling the neighborhood (implied: police or armed forces, depending on a location/situation).

We are using the notion of defaults in processing joke-like structure that these headlines generate. We are assuming that most of the information to process the headlines is taken from the defaults, and it's with the help of the defaults that the oppositeness can be found. While the “model” that we outline doesn't cover all cases of headlines, we will use it to generate some of the sentences that could be used for a specific domain. The “model” (more accurately template) is described in Fig. 3.

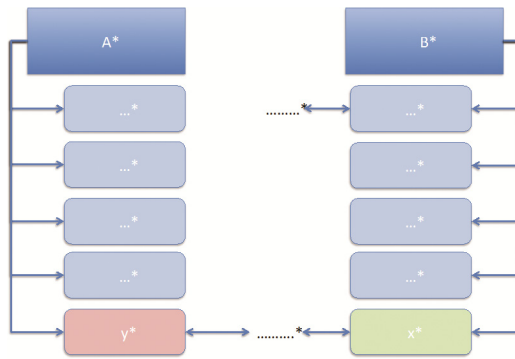


Fig. 3. Script defaults used to generate joke-like headlines

In the full paper, we will analyze different paths that the above jokes take and generalize it to different types of jokes that are possible within this type of humor, thus creating an algorithm for detecting them. We will then generate several jokes that are based on the found algorithm and describe the results.

3.1 Defaults in the Fluid Mechanics

We assume that any definition that is used in a field can serve as a default for the professionals of that field. Thus, for people who are familiar with the linguistics of humor, jokes, by default, are analyzed as scripts. For people that are familiar with computing, arithmetic operations are done in binary code (at the lowest level). For people that are familiar with security, any system should come protected.

We are using the following terms and properties from fluid dynamics to generate jokes based on our model:

- Incompressible fluid – fluid with a constant density
- Shear stress – frictional force exerted by a moving fluid on a surface (frictional force between layers of fluid)
- Pressure increases with depth in the fluid
- In incompressible flow the inlet flow must balance the outlet flow
- Conservation of mass principle: mass cannot be created or destroyed.

The domain of fluid mechanics was selected opportunistically – to allow one of the co-authors to test the jokes on the student population.

FrameNet doesn't cater to such restricted domains, but, as we previously stated, any semantic knowledge base system is capable of representing information described in FrameNet or in the definitions above. We can thus use Ontological Semantic Technology to create concepts necessary to represent our domain of interest.

We then use the model outline to generate the following jokes:

- (1) Fluid dynamics researcher stirred tea; discovered milk can be mixed into it.
- (2) Fluid dynamics researcher went diving; discovered pressure increases with depth in the fluid.
- (3) Studies of incompressible flow showed if you want to drink a cup of milk, you have to swallow it too.
- (4) Study conservation of mass! You will learn that you can't add more water into a bottle than it's capacity.

Each of the jokes is based on every day situations that we are all familiar with, then explained from the basics of fluid dynamics. The following information might be helpful to understanding of first joke: there are two mechanisms of mixing – molecular diffusion, when we rely on chaotic motion of the molecules and advection, when we rely on the flow transport. Since the flow transport is a much more effective mechanism, it is obvious that stirring can enhance mixing. The joke here is that it is intuitively clear that one should stir milk into a cup of tea without any knowledge of the underlying physics of transport and mixing. The underlying scripts are that of pouring tea (by a regular person) vs. flow transport.

The second joke is based on knowledge that in a non-moving fluid, the pressure is balanced by the gravity and thus increases linearly with the depth. As the diver is going down, the weight of the column of fluid above increases resulting in higher pressures. Here again, the result is so obvious from a common experience that once does not have to think about the balance of the pressure gradients and gravity forces.

The third joke is based on the fact that if the flow is incompressible, the flow in must be balanced by the flow out, it cannot be compressed like the air in a car's tire. Thus, once the mouth is completely filled with milk, I cannot add more to it without swallowing since I can't compress it into a smaller volume inside the mouth.

Finally, in the fourth joke, since mass cannot be just created or destroyed, it is clear that there can be no sources or sinks in the bottle. The joke implies that a liquid in the bottle cannot be compressed, hence once the bottle is filled we can't push more liquid into it. Thus, a fancy term of mass conservation describes a trivial fact.

It is clear that in order for an ontology to capture the meaning of these jokes, the domain of fluid mechanics (or at least information representation the basics) should be

very well developed. However, once the scripts B in Fig. 3 are well represented, the results of these scripts is what triggers the green box of B.

The A scripts of Fig. 3 is the basic knowledge scripts – those that should be easily available to a general knowledge system. Most of the time, however, such systems lack detail even for a general knowledge domain. For example, FrameNet anchors the verb *dive* in a *Path_Shape* frame (describing “the fictive motion of a stationary Road”). Thus, while a *Direction*, as a core element, can describe the event of going down, the water is not present in the frame as a default. Similarly, the frame of *Ingestion*, where *drink* is anchored, is not connected to swallowing.

References

1. White, B.: Teen pregnancy drops off significantly after age 25. *The Morning Call*, 13 December 2013
2. McGlynn, K.: The Most Obvious Headlines of All Times. *Huffington Post*, 23 July 2010
3. Kyff, R.: *English Goes on a Toot*. Creators Publishing (2015)
4. Lindsell-Roberts, S.: *Technical Writing for Dummies*. Wiley Publishers, New York (2001)
5. Taylor, J.M., Raskin, V., Hempelmann, C.F., Attardo, S.: An unintentional inference and ontological property defaults. In: *International Conference on Systems, Man and Cybernetics*, Istanbul, Turkey (2010)
6. Raskin, V.: *Semantic Mechanisms of Humor*. D. Reidel, Dordrecht, Boston, Lancaster (1985)
7. Raskin, V., Hempelmann, C.F., Taylor, J.M.: How to understand and assess a theory: the evolution of the SSTH into the GTVH and now into the OSTH. *J. Literary Theory* **3**, 2 (2009)
8. Hickman, L., Taylor, J., Raskin, V.: Direct object omission as a sign of conceptual defaultness. In: *FLAIRS* (2016)
9. Schank, R., Abelson, H.: *Scripts, Plans, Goals and Understanding*. Erlbaum, Hillsdale (1977)
10. Raskin, V.: Ambient scripts in humor and beyond. In: *HCI 2016*, Toronto, Ontario, Canada (2016)
11. Raskin, V., Hempelmann, C., Taylor, J.: Guessing vs. knowing: the two approaches to semantics in natural language processing. In: *Dialogue* (2010)
12. Taylor, J., Hempelmann, C., Raskin, V.: On an automatic acquisition toolbox for ontologies and lexicons in ontological semantics. In: *International Conference in Artificial Intelligence*, ICAI 2010 (2010)
13. Hempelmann, C., Taylor, J., Raskin, V.: Application-guided ontological engineering. In: *International Conference in Artificial Intelligence*, ICAI 2010 (2010)
14. Ruppenhofer, J., Ellsworth, M., Petruck, M.R.L., Johnson, C.R., Scheffczyk, J.: *FramNET II: Extended Theory and Practice* (2010)
15. Attardo, S., Raskin, V.: Script theory revis(it)ed: joke similarity and joke representation model. *HUMOR: Int. J. Humor Res.* **4**, 3–4 (1991). (83pp.)