

What Humour Tells Us About Discourse Theories

Arjun Karande

Indian Institute of Technology Kanpur

Kanpur 208016, India

arjun@iitk.ac.in

Abstract

Many verbal jokes, like garden path sentences, pose difficulties to models of discourse since the initially primed interpretation needs to be discarded and a new one created based on subsequent statements. The effect of the joke depends on the fact that the second (correct) interpretation was not visible earlier. Existing models of discourse semantics in principle generate all interpretations of discourse fragments and carry these until contradicted, and thus the dissonance criteria in humour cannot be met. Computationally, maintaining all possible worlds in a discourse is very inefficient, thus computing only the maximum-likelihood interpretation seems to be a more efficient choice on average. In this work we outline a probabilistic lexicon based lexical semantics approach which seems to be a reasonable construct for discourse in general and use some examples from humour to demonstrate its working.

1 Introduction

Consider the following :

- (1) I still miss my ex-wife, but my aim is improving.
- (2) The horse raced past the barn fell.

In a discourse structure common to many jokes, the first part of (1) has a default set of interpretations, say P_1 , for which no consistent interpretation can be found when the second part of the joke is uttered. After a search, the listener reaches

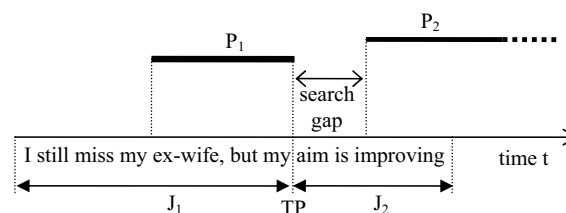


Figure 1: Cognitive model of destructive dissonance as in joke (1). The initial sentence primes the possible world P_1 where *miss* is taken in an emotional sense. After encountering the word *aim* this is destroyed and eventually a new world P_2 arises where *miss* is taken in the physical sense.

the alternate set of interpretations P_2 (Figure 1). A similar process holds for garden path sentences such as (2), where the default interpretation created in the first part (upto the word *barn*) has to be discarded when the last part is heard. The search involved in identifying the second interpretation is an important indicator of human communication, and linguistic impairment such as autism often leads to difficulty in identifying jokes.

Yet, this aspect of discourse is not sufficiently emphasized in most computational work. Cognitively, this is a form of dissonance, a violation of expectation. However, unlike some forms of dissonance which may be constructive, leading to metaphoric or implicature shifts, where part of the original interpretation may be retained, these discourse structures are destructive, and the original interpretation has to be completely abandoned, and a new one searched out (Figure 2). Often this is because the default interpretation involves a sense-association that has very high coherence in the immediate context, but is nullified by later

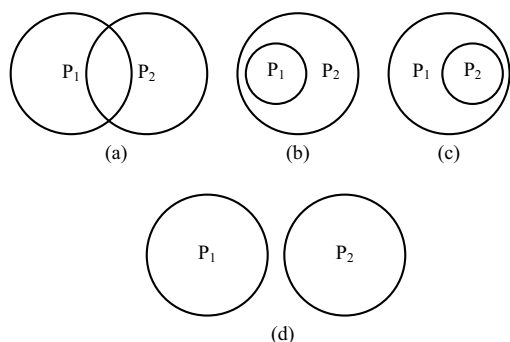


Figure 2: Cognitive Dissonance in Discourse (a-c) can be Constructive, where the interpretation P_1 does not disappear completely after the dissonant utterance, or (d) Destructive, where P_2 has to be arrived at afresh and P_1 is destroyed completely.

utterances.

While humour may involve a number of other mechanisms such as allusion or stereotypes (Shibles, 1989; Gruner, 1997), a wide class of verbal humour exhibits destructive dissonance. For a joke to work, the resulting interpretation must result in an incongruity, what (Freud, 1960) calls an ‘energy release’ that breaks the painful barriers we have around forbidden thoughts.

Part of the difficulty in dealing with such shifts is that it requires a rich model of discourse semantics. Computational theories such as the *General Theory of Verbal Humour* (Attardo and Raskin, 1991) have avoided this difficult problem by adopting extra-linguistic knowledge in the form of scripts, which encode different oppositions that may arise in jokes. Others (Minsky, 1986) posit a general mechanism without considering specifics. Other models in computation have attempted to generate jokes using templates (Attardo and Raskin, 1994; Binsted and Ritchie, 1997) or recognize jokes using machine learning models (Mihalcea and Strapparava, 2005).

Computationally, the fact that other less likely interpretations such as P_2 are not visible initially, may also result in considerably efficiency in more common situations, where ambiguities are not generated to begin with. For example, in joke (1) the interpretation after reading the first clause, has the word *miss* referring to the abstract act of missing a dear person. After hearing the punch line, somewhere around the word *aim*, (the trigger point TP), we have to revise our interpretation to one where *miss* is used in a physical sense, as in

shooting a target. Then, the forbidden idea of hurting ex-wives generates the humour. By hiding this meaning, the mechanism of destructive dissonance enables the surprise which is the crux of the joke.

In the model proposed here, no extra-linguistic sources of knowledge are appealed to. Lexical Semantics proposes rich inter-relations encoding knowledge within the lexicon itself (Pustejovsky, 1995; Jackendoff, 1990), and this work considers the possibility that such lexicons may eventually be able to carry discourse interpretations as well, to the level of handling situations such as the destructive transition from a possible-world P_1 to possible world P_2 . Clearly, a desideratum in such a system would be that P_1 would be the preferred interpretation from the outset, so much so that P_2 , which is in principle compatible with the joke, is not even visible in the first part of the joke. It would be reasonable to assume that such an interpretation may be constructed as a ‘Winner Take All’ measure using probabilistic inter-relations in the lexicon, built up based on usage frequencies. This would differ from existing theories of discourse in several ways, as will be illustrated in the following sections.

2 Models of Discourse

Formal semantics (Montague, 1973) looked at logical structures, but it became evident that language builds up on what is seemingly semantic incompatibility, particularly in Gricean *Implicature* (Grice, 1981). It became necessary to look at the relations that describe interactions between such structures. (Hobbs, 1985) introduces an early theory of discourse and the notion of *coherence relations*, which are applied recursively on discourse segments. Coherence relations, such as *Elaboration*, *Explanation* and *Contrast*, are relations between discourse units that bind segments of text into one global structure. (Grosz and Sidner, 1986) incorporates two more important notions into its model - the idea of intention and focus. The *Rhetorical Structure Theory*, introduced in (Mann and Thompson, 1987), binds *text spans* with *rhetorical relations*, which are discourse connectives similar to coherence relations.

The *Discourse Representation Theory* (DRT) (Kamp, 1984) computes inter-sentential anaphora and attempts to maintain text cohesion through sets of predicates, termed *Discourse Representation Structures* (DRSs), that represent discourse

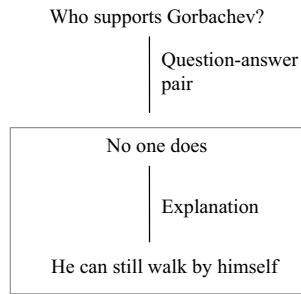


Figure 3: Rhetorical Relations for joke (3)

units. A *Principal DRS* accumulates information contained in the text, and forms the basis for resolving anaphora and discourse referents.

By marrying DRT to a rich set of rhetorical relations, *Segmented Discourse Representation Theory* (SDRT) (Lascarides and Asher, 2001) attempts to create a dynamic framework that tries to bridge the semantic-pragmatic interface. It consists of three components - *Underspecified Logical Formulae (ULF)*, *Rhetorical Relations* and *Glue Logic*. Semantic representation in the ULF acts as an interface to other levels. Information in discourse units is represented by a modified version of DRS, called *Segmented Discourse Representation Structures (SDRSs)*. SDRSs are connected through rhetorical relations, which posit relationships on SDRSs to bind them. To illustrate, consider the discourse in (3):

(3) Who supports Gorbachev? No one does, he can still walk by himself!

The rhetorical relations over the discourse are shown in Figure 3. Here, *Explanation* induces subordination and implies that the content of the subordinate SDRSs work on further qualifying the principal SDRS, while *Question-Answer Pair* induces coordination. Rhetorical relations thus connect semantic units together to formalize the flow in a discourse. SDRT’s Glue Logic then runs sequentially on the ULF and rhetorical relations to reduce underspecification and disambiguation and derive inferences through the discourse. The way inferencing is done is similar to DRT, with the additional constraints that rhetorical relations specify.

A point to note is SDRT’s *Maximum Discourse Coherence (MDC) Principle*. This princi-

ple is used to resolve ambiguity in interpretation by maximizing discourse coherence to obtain the *Pragmatically Preferred* interpretation. There are three conditions on which MDC works: (a) The more rhetorical relations there are between two units, the more coherent the discourse. (b) The more anaphorae that are resolved, the more coherent the discourse. (c) Some rhetorical relations can be measured for coherence as well. For example, the coherence of *Contrast* depends on how dissimilar its connected prepositions are. SDRT uses rhetorical relations and MDC to resolve lexical and semantic ambiguities. For example, in the utterance ‘John bought an apartment. But he rented it’, the sense of *rented* is that of renting out, and that is resolved in SDRT because the word *but* cues the relation *Contrast*, which prefers an interpretation that maximizes semantic contrast between its connectives.

Glue logic works by iteratively extracting subsets of inferences through the flow of the discourse. This is discussed in more detail later.

2.1 Lexicons for Discourse modeling

Pustejovsky’s Generative Lexicon (GL) model (Pustejovsky, 1995) outlines an ambitious attempt to formulate a lexical semantics framework that can handle the unboundedness of linguistic expressions by providing a rich semantic structure, a principled ontology of concepts (called *qualia*), and a set of generative devices in which participants in a phrase or sentence can influence each other’s semantic properties.

The ontology of concepts in GL is hierarchical, and concepts that exhibit similar behaviour are grouped together into subsystems called *Lexical Conceptual Paradigms (LCP)*. As an example, the GL structure for *door* is an LCP that represents both the use of door as a physical object such as in ‘he knocked on the door’, as well as an aperture like in ‘he entered the door’.

In this work, we extend the GL structures to incorporate likelihood measures in the ontology and the event structure relations. The *Probabilistic Qualia Structure*, which outlines the ontological hierarchy of a lexical item, also encodes frequency information. Every time the target word appears together with an ontologically connected concept, the corresponding qualia features are strengthened. This results in a probabilistic model of qualia features, which can in principle determine

that a *book* has *read* as its maximally likely telic role, but that in the context of the agent being the author, *write* becomes more likely.

Generative mechanisms work on this semantic structure to capture systematic polysemy in terms of type shifts. Thus *Type Coercion* enforces semantic constraints on the arguments of a predicate. For example, ‘He enjoyed the book’ is coerced to ‘He enjoyed reading the book’ since *enjoy* requires an activity, which is taken as the telic role of the argument, i.e. that of *book*. *Co-composition* constrains the type-shifting of the predicate by its arguments. An example is the difference between ‘bake a cake’ (creating a new object) versus ‘bake beans’ (state change). Finally, *Selective Binding* type-shifts a modifier based on the head. For example, in ‘old man’ and ‘old book’, the property being modified by *old* is shifted from physical-age to information-recency.

To accommodate for likelihoods in generative mechanisms, we need to incorporate conditional probabilities between the lexical and ontological entries that the mechanisms work on. These probabilities can be stored within the lexicon itself or integrated into the generative mechanisms. In either case, mechanisms like *Type Coercion* should no longer exhibit a default behaviour - the coercion must change based on frequency of occurrence and context.

3 The Analysis of Humour

The General Theory of Verbal Humour (GTVH), introduced earlier, is a well-known computational model of humour. It uses the notion of scripts to account for the opposition in jokes. It models humour as two opposing and overlapping scripts put together in a discourse, one of which is apparent and the other hidden from the reader till a trigger point, when the hidden script suddenly surfaces, generating humour. However, the notion of scripts implies that there is a script for every occasion, which severely limits the theory. On the other hand, models of discourse are more general and do not require scripts. However, they lack the mechanism needed to capture such oppositions. In addition to joke (3), consider:

(4) Two guys walked into a bar. The third one ducked.

The humour in joke (4) results from the polyse-

mous use of the word *bar*. The first sentence leads us to believe that *bar* is a place where one drinks, but the second sentence forces us to revise our interpretation to mean a solid object. GTVH would use the DRINKING BAR script before the trigger and the COLLISION script after. Joke (3), quoted in Raskin’s work as well, contains an obvious opposition. The first sentence invokes the sense of *support* being that of political support. The second sentence introduces the opposition, and the meaning of *support* is changed to that of physical support.

In all examples discussed so far, the key observations are that (i) a single inference is primed by the reader, (ii) this primary inference suppresses other inferences until (iii) a trigger point is reached.

To formalize the unfolding of a joke, we refer back to Figure 1. Let t be a point along the timeline. When $t < TP$, both P_1 and P_2 are compatible, and the possible world is $P = P_1 \cup P_2$. P_1 is the preferred interpretation and P_2 is hidden. When $t = TP$, J_2 is introduced, and P_1 becomes incompatible with P_2 , and P_1 may also lose compatibility with J_2 . P_2 now surfaces as the preferred inference. The reader has to invoke a search to find P_2 , which is represented by the *search gap*.

A possible world $P_i = \{q_{i1}, q_{i2}, \dots, q_{ik}\}$ where q_{mn} is an inference. Two worlds P_i and P_j are incompatible if there exists any pair of sets of inferences whose intersection is a contradiction. i.e.

P_i is said to be incompatible with P_j iff $\exists \{q_{i1}, q_{i2}, \dots, q_{ik}\} \subseteq P_i \wedge \exists \{q_{j1}, q_{j2}, \dots, q_{jl}\} \subseteq P_j$ such that $\{q_{i1} \wedge q_{i2} \wedge \dots \wedge q_{ik} \wedge q_{j1} \wedge q_{j2} \wedge \dots \wedge q_{jl}\} \Rightarrow F$. They are said to be compatible if no such subsets exist.

We now explore in detail why compositional discourse models fail to handle the mechanisms of humour.

3.1 Beyond Scripts - Why Verbal Humour Should Be Winner Take All

An argument against the approach of existing discourse models like SDRT concerns their iterative inferencing. At each point in the process of infer-

encing, SDRT’s Glue Logic carries over all interpretations possible within its constraints as a set. MDC ranks contending inferences, allowing less preferred inferences to be discarded, and the result of this process is a subset of the input to it. Contrasting inferences can coexist through underspecification, and the contrast is resolved when one of them loses compatibility. This is cognitively unlikely; (Miller, 1956) has shown that the human brain actively retains only around seven units of information. With such a limited working memory, it is not cognitively feasible to model discourse analysis in this manner. Cognitive models working with limited-capacity short-term memory like in (Lewis, 1996) support the same intuition. Thus, a better approach would be a *Winner Take All (WTA)* approach, where the most likely interpretation, called the winner, suppresses all other interpretations as we move through the discourse. The model must be revised to reflect new contexts if they are incompatible with the existing model.

Let us now explore this with respect to joke (3). There is a Question-Answer relation between the first sentence and the next two. The semantic representation for the first sentence alone is:

$$\exists x(\text{support}(x, \text{Gorbachev})), x = ?$$

The $x = ?$ indicates a missing referent for *who*. Using GL, it is not difficult to resolve the sense of *support* to mean that of political support. To elaborate, the lexical entry of *Gorbachev* is an LCP of two senses - that of the head of government and that of an animate, as shown:

$$\left[\begin{array}{l} \text{Gorbachev} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = x: \text{man} \\ \text{ARG2} = y: \text{head_of_govt} \\ \text{D-ARG3} = z: \text{community} \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{human.president.lcp} \\ \text{FORMAL} = p(x, y) \\ \text{TELIC} = \text{govern}(y, z) \end{array} \right] \end{array} \right]$$

The two senses of *support* applicable in this context are that of physical support and of political support. We use abstract support as a generalization of the political sense. The analysis of the first sentence alone would allow for both these possibilities:

$$\left[\begin{array}{l} \text{support}_{\text{abs}} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = x: \text{animate} \\ \text{ARG2} = y: \text{abstract_entity} \end{array} \right] \\ \text{EVENTSTR} = \left[\text{E}_1 = e_1 : \text{process} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{FORMAL} = \text{support}_{\text{abs}}\text{-act}(e_1, x, y) \\ \text{AGENTIVE} = \dots \end{array} \right] \end{array} \right]$$

$$\left[\begin{array}{l} \text{support}_{\text{phy}} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = x: \text{physical_entity} \\ \text{ARG2} = y: \text{physical_entity} \end{array} \right] \\ \text{EVENTSTR} = \left[\text{E}_1 = e_1 : \text{process} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{FORMAL} = \text{support}_{\text{phy}}\text{-act}(e_1, x, y) \\ \text{AGENTIVE} = \dots \end{array} \right] \end{array} \right]$$

Thus, after the first sentence, the sense of *support* includes both senses, i.e. $\text{support} \in \{\text{support}_{\text{abs}}, \text{support}_{\text{phy}}\}$.

We then come across the second sentence and establish the semantic representation for it, as well as establish rhetorical relations. We find that the sentence contains $\text{walk}(z)$. SDRT’s Right Frontier Rule resolves the referent *he* to *Gorbachev*. Also, the clause ‘no one does’ resolves the referent x to *null*. Thus, we get:

$$\text{walk}(\text{Gorbachev}) \wedge \text{support}(\text{null}, \text{Gorbachev})$$

Now consider the lexical entry for *walk*:

$$\left[\begin{array}{l} \text{walk} \\ \text{ARGSTR} = \left[\text{ARG1} = x: \text{animate} \right] \\ \text{EVENTSTR} = \left[\text{E}_1 = e_1 : \text{process} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{FORMAL} = \text{walk.act}(e_1, x) \\ \text{AGENTIVE} = \text{walk.begin}(e_1, x) \end{array} \right] \end{array} \right]$$

The action *walk* requires an animate argument. Since $\text{walk}(\text{Gorbachev})$ is true, the sense of *support* in the previous sentence is restricted to mean physical support, i.e. $\text{support} = \text{support}_{\text{phy}}$, since only $\text{support}_{\text{phy}}$ can take an animate argument as its object - the *abstract_entity* requirement of $\text{support}_{\text{abs}}$ causes it to be ruled out, ending at a final inference.

The change of sense for *support* is key to the generation of humour, but SDRT fails to recognize the shift since it neither has any priming mechanism nor revision of models built into it. It merely works by restricting the possible inferences as more information becomes available. Referring to Figure 1 again, SDRT will only account for the refinement of possible worlds from $P_1 \cup P_2$ to P_2 . It will not be able to account for the priming of either P_i , which is required.

4 A Probabilistic Semantic Lexicon

We now introduce a WTA model under which priming could be well accounted for. We would like a model under which a single interpretation is made at each point in the analysis. We want a set

of possible worlds P such that:

$$J_1 \xrightarrow{WTA} P = \{p : p \text{ is a world consistent with } J_1\}$$

WTA ensures that only the prime world P is chosen by J_1 . When J_2 is analyzed, no world $p \in P$ can satisfy J_2 , i.e:

$$\forall p \in P, \neg J_2 \longrightarrow p$$

In this case, we need to backtrack and find another set P' that satisfies both J_1 and J_2 , i.e:

$$(J_1, J_2) \xrightarrow{WTA} P'$$

In Figure 1, $P = P_1$ and $P' = P_2$.

The most appropriate way to achieve this is to include the priming in the lexicon itself. We present a lexical structure where senses of compositional units are attributed with a probability of occurrence approximated by its frequency count. The probability of a composition can then be calculated from the individual probabilities. The highest probability is primed. Thus, at every point in the discourse, only one inference emerges as primary and suppresses all other inferences. As an example, the proposed structure for Gorbachev is presented below:

$$\left[\begin{array}{l} \text{Gorbachev} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = x: \text{man} \\ \text{ARG2} = y: \text{head_of_govt} \\ \text{D-ARG3} = z: \text{community} \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{FORMAL} = p(x, y) \\ p(\text{man}) = p_1 \\ p(\text{head_of_govt}) = p_2 \\ \dots \end{array} \right] \end{array} \right]$$

Instead of using the concept of an LCP as in classical GL, we assign probabilities to each sense encountered. These probabilities can then facilitate priming.

To add weight to the argument with empirical data, we use WordNet (Fellbaum, 1998), built on the British National Corpus, as an approximation for frequency counts. We find that

$$P(\text{support}_{abs}) = 0.59 \text{ and} \\ P(\text{support}_{phy}) = 0.36.$$

Similarly, for the notion of *Gorbachev*, it is plausible to assume that *Gorbachev* as head of

government is more meaningful for most of us, rather than just another old man. In order to make an inference after the first sentence, we need to *search* for the correct interpretation, i.e. we need to find $\text{argmax}_{i,j}(P(\text{support}_i/\text{Gorbachev}_j))$, which intuitively should be $P(\text{support}_{abs}/\text{head_of_govt})$. Making a similar analysis as in the previous section, the second sentence should violate the first assumption, since $\text{walk}(\text{Gorbachev})$ cannot be true (since $P(\text{abstract_entity}) = 0$). Thus, we need to revise our inference, moving back to the first sentence and choosing $\text{max}(P(\text{support}_i/\text{Gorbachev}_j))$ that is compatible with the second sentence. This turns out to be $P(\text{support}_{phy}/\text{animate})$. Thus, the distinct shift between inferences is captured in the course of analysis. Cognitive studies such as the studies on Garden Path Sentences strengthen this approach to analysis. (Lewis, 1996), for example, presents a model that predicts cognitive observations with very limited working memory.

Storing the inter-lexical conditional probabilities is also an issue, as mentioned earlier. Where, for example, do we store $P(\text{support}_i/\text{Gorbachev}_j)$? One possible approach would be to store them with either lexical item. A better approach would be to bestow the responsibility of calculating these probabilities upon the generative mechanisms of the semantic lexicon whenever possible.

Let us now analyze joke (1) under the probabilistic framework. Again, approximations for probability of occurrence will be taken from WordNet. The entry for *wife* in WordNet lists just one sense, and so we assign a probability of 1 to it in its lexical entry:

$$\left[\begin{array}{l} \text{wife} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = x: \text{woman} \\ \text{D-ARG2} = y: \text{man} \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{FORMAL} = \text{husband}(x) = y \\ \text{AGENTIVE} = \text{marriage}(x, y) \\ p(\text{woman}) = 1 \end{array} \right] \end{array} \right]$$

The humour is generated due to the lexical ambiguity of *miss*. We list the lexical entries of the two senses of *miss* that apply in this context - the first being an abstract emotional state and the other being a physical process.

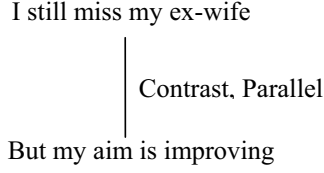


Figure 4: Rhetorical relations for joke (1)

$$\left[\begin{array}{l} \text{miss}_{\text{abs}} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = x: \text{animate} \\ \text{ARG2} = y: \text{entity} \end{array} \right] \\ \text{EVENTSTR} = \left[\begin{array}{l} \text{E}_1 = e_1: \text{state} \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{FORMAL} = \text{miss}_{\text{abs_act}}(e_1, x, y) \\ \text{AGENTIVE} = \dots \end{array} \right] \end{array} \right]$$

$$\left[\begin{array}{l} \text{miss}_{\text{phy}} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = x: \text{physical_entity} \\ \text{ARG2} = y: \text{physical_entity} \\ \text{D-ARG1} = z: \text{trajector} \end{array} \right] \\ \text{EVENTSTR} = \left[\begin{array}{l} \text{E}_1 = e_1: \text{process} \\ \text{E}_2 = e_2: \text{state} \\ \text{RESTR}_2 = <_{\alpha} \\ \text{HEAD}_2 = e_2 \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{FORMAL} = \text{missed}_{\text{phy_act}}(e_2, x, y, z) \\ \text{AGENTIVE} = \text{shoot}(e_1, x, y, z) \end{array} \right] \end{array} \right]$$

The Rhetorical Relations for joke (1) are presented in Figure 4. After parsing the first sentence, the logical representation obtained is:

$$\exists e_1 \exists e_2 \exists e_3 \exists x \exists y (\text{wife}(e_1, x, y) \wedge \text{divorce}(e_2, x, y) \wedge \text{miss}(e_3, x, y) \wedge e_1 < e_2 < e_3)$$

To arrive at a prime inference, note that the semantic types of the arguments of both senses of *miss* are exclusive, and hence $P(\text{physical_entity}/\text{miss}_{\text{phy}}) = 1$ and $P(\text{entity}/\text{miss}_{\text{abs}}) = 1$. Thus, using Bayes Theorem, to compare $P(\text{miss}_{\text{abs}}/\text{entity})$ and $P(\text{miss}_{\text{phy}}/\text{physical_entity})$, it is sufficient to compare $P(\text{miss}_{\text{abs}})$ and $P(\text{miss}_{\text{phy}})$. From WordNet,

$$P(\text{miss}_{\text{abs}}) = 0.22 \text{ and} \\ P(\text{miss}_{\text{phy}}) = 0.06.$$

Thus, the primed inference has $\text{miss} = \text{miss}_{\text{abs}}$. The second sentence has the following logical representation:

$$\exists x (\delta \text{goodness}(\text{aim}(x)) > 0)$$

This simply means that a measure of the aim, called *goodness*, is undergoing a positive change. The word *but* is a cue for a *Contrast* relation between the two sentences, while the discourse suggests *Parallelism*. The two senses of *aim* compatible with the first sentence are aim_{abs} , which is synonymous to *goal*, and aim_{phy} , referring to the physical sense of missing. We now need to consider $P(\text{aim}_{\text{abs}}/\text{miss}_{\text{abs}})$ and $P(\text{aim}_{\text{phy}}/\text{miss}_{\text{phy}})$. The semantic constraints of the rhetorical relation *Contrast* ensures that the second is more coherent, i.e. it is more probable that the contrast of physical aim getting better is more coherent with the physical sense of *miss*, and we expect this to be reflected in usage frequency as well. Therefore $P(\text{aim}_{\text{abs}}/\text{miss}_{\text{abs}}) < P(\text{aim}_{\text{phy}}/\text{miss}_{\text{phy}})$, and we need to shift our inference and make $\text{miss} = \text{miss}_{\text{phy}}$.

As a final assertion of the probabilistic approach, consider:

(5) You can lead a child to college, but you cannot make him think.

The incongruity in joke (5) does not result from a syntactical or semantic ambiguity at all, and yet it induces dissonance. The dissonance is not a result of compositionality, but due to the access of a whole linguistic structure, i.e. we recall the familiar proverb ‘You can lead a horse to water but you cannot make it drink’, and the deviation from the recognizable structure causes the violation of our expectations. Thus, access is not restricted to the lexical level; we seem to store and access bigger units of discourse if encountered frequently enough. The only way to do justice to this joke would be to encode the entire sentential structure directly into the lexicon. Our model will now also consider these larger chunks, whose meaning is specified atomically. The dissonance will now come from the semantic difference between the accessed expression and the one under analysis.

5 Conclusion

We have examined the mechanisms behind verbal humour and shown how existing discourse models are inadequate at capturing the mechanisms of humour. We have proposed a probabilistic

WTA model based on lexical frequency distributions that is more capable at handling humour, and is based on the notion of expectation and dissonance.

It would be interesting now to find necessary and sufficient conditions under this framework for humour to be generated. Although the above framework can identify incongruity in humour discourse, the same mechanisms are used and indeed are often integral to other forms of literature. Poems, for example, often rely on such mechanisms. Are Freudian thoughts the key to separating humour from the rest, or is it a result of the intentional misleading done by the speaker of a joke? Also, it would be very interesting to find an empirical link between the extent of incongruity in jokes in our framework and the way people respond to them.

Finally, a very interesting question is the acquisition of the lexicon under such a model. How are lexical semantic models learned by the language acquirer probabilistically? An exploration of the question might result in a cognitively sound computational model for acquisition.

References

- Salvatore Attardo and Victor Raskin. 1991. Script theory revis(it)ed: Joke similarity and joke representation model. 4(3):293–347.
- Savatore Attardo and Victor Raskin. 1994. Non-literalness and non-bona-fide in language: An approach to formal and computational treatments of humor. volume 2, pages 31–69.
- Kim Binsted and Graeme Ritchie. 1997. Computational rules for generating punning riddles. *HUMOR - International Journal of Humor Research*, 10(1):25–76.
- Christiane Fellbaum. 1998. *WordNet - An Electronic Lexical Database*. MIT Press.
- Sigmund Freud. 1960. Jokes and their relation to the unconscious. *The Standard Edition of the Complete Psychological Works of Sigmund Freud*.
- Herbet Paul Grice. 1981. Presupposition and conversational implicature. In *Radical Pragmatics*, pages 183–197. New York: Academic Press.
- Barbara J. Grosz and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.
- Charles R. Gruner. 1997. *The Game of Humor: A Comprehensive Theory of Why We Laugh*. NJ: Transaction Publishers.
- Jerry R Hobbs. 1985. On the coherence and structure of discourse. Technical Report CSLI-85-37, Center for the Study of Language and Information, Stanford University.
- Ray S. Jackendoff. 1990. *Semantic Structures*. MIT Press.
- H. Kamp. 1984. A theory of truth and semantic representation. In J. Groenendijk, T. M. V. Janssen, and M. Stokhof, editors, *Truth, Interpretation and Information: Selected Papers from the Third Amsterdam Colloquium*, pages 1–41. Foris Publications, Dordrecht.
- Asher Lascarides and Nicolas Asher. 2001. Segmented discourse representation theory: Dynamic semantics with discourse structure. *Computing Meaning*, 3.
- Richard L. Lewis. 1996. Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*, 25(1):93 – 115.
- William C. Mann and Sandra A. Thompson. 1987. Rhetorical structure theory: A theory of text organization. Technical Report ISI/RS-87-190, Center for the Study of Language and Information, Stanford University.
- Rada Mihalcea and Carlo Strapparava. 2005. Making computers laugh: Investigations in automatic humor recognition. In *Joint Conference on Human Language Technology / Empirical Methods in Natural Language Processing (HLT/EMNLP)*.
- George A. Miller. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63:81–97.
- Marvin Minsky. 1986. *The Society of Mind*. Simon and Schuster.
- Richard Montague. 1973. The proper treatment of quantification in ordinary English. In *Approaches to Natural Language*, pages 221–242. D. Reidel.
- James Pustejovsky. 1995. *The Generative Lexicon*. MIT Press.
- Warren Shibles. 1989. Humor reference guide. <http://facstaff.uww.edu/shiblesw/humorbook>.