Remember, Remember, the 5th of November: A Computational Approach to Cross-Domain Memorability

Ashima Arora* aa3470@columbia.edu
Anusha Balakrishnan* ab3596@columbia.edu
Rituparna Mukherjee* rm3166@columbia.edu

*Department of Computer Science
Columbia University
New York, NY, USA

Abstract

Identifying whether a given utterance is memorable or not has tremendous utility in advertising and education, among other domains. We describe an SVM classifier that uses linguistic cues to predict whether an utterance will be memorable or not. The feature set is based on existing psycholinguistic research, and we describe ways to computationally extract those features. We also created and hand-annotated test corpora from three different domains for memorability. We then trained our classifier on the Cornell Movie-Quotes Corpus, and tested its accuracy on the test sets. Overall, we were able to improve performance over baseline using our complete feature set, with varying results across domains.

1 Introduction

The human mind is impressionable to a very large extent. The idea of memorability deals primarily with the characteristics - of certain messages, images, sounds, etc. - that imprint them onto human memory and allows them to be recalled for a very long time after they are initially encountered. When it comes to text specifically, although we participate in hundreds of interpersonal message exchanges in a day, only a select few of these messages stick to memory and are easily recalled later. The factors determining why those messages are memorable span several syntactic and semantic aspects of the messages. Understanding those factors is an extremely interesting task not only in the field of psychology but also in the field of linguistics. Taking cues from some existing research on Memorability, in this study we try to identify language nuances which can define memorability of text across vaious domains. We primarily focus on three domains; Advertising slogans, Political Slogans and Educational Mnemonics and draw linguistic features which can help distinguish between memorable and non memorable messages based only on the content (text) of these messages. Distinguishing between memorable and non-memorable messages is a difficult problem because in most cases what makes an utterance or a message memorable is the circumstance or the situation around which it occurred. The idea behind this study is to de-link this dependency and come up with linguistic features which can independently help in classifying memorable messages. To understand the problem at hand, let's consider a few examples from different domains.

Show me the money.
Kwan, that's your word.

May be she is born with it. Maybe it's Maybelline.
The watch that made the dollar famous.

My very educated mother just served us nachos.
Raising elephants is so utterly boring.

Yes we can.
All power to the soviets.

In the examples above, each pair consists of a memorable and non-memorable quote from movies, advertising slogans, education mnemonics and political slogans respectively. Some of the above seem very familiar to us, while the others do not. But if we try to consider each of these quotes independently without associating them with circumstances under which they came into play, that's when we see how difficult the task of classifying them truly is. Extending the idea proposed in (Cristian et al., 2012), we aim at understanding and predicting the memorability of messages, from different domains. For this task, we have built a Cross-Domain Memorability Classifier (CDMC) by identifying different linguistic features which can be extracted from texts and which are characteristic to memorable messages.

2 Related Work

The most concrete study on this subject, which also forms the primary reference for this work, is presented in (Cristian et al., 2012). This study considers whether and how phrasing, i.e. choice of words, and sentence structure, affects memorability. Given a pair of quotes, one memorable and the other non-memorable, their model identifies (with reasonable reliability) the memorable quote amongst the two. There have also been some interesting psycholinguistic studies investigating memorability. One of the earliest works in the field of memorability is presented in (Knapp et al., 1981). The authors focus on determining whether certain linguistic cues and sentence structures play a role in making messages memorable. The authors focused on four major aspects of memorable utterances; the structure of the message, the form and organization of the message, the content of the message and the circumstances surrounding the enactment and reception of the message. The study showed that in most cases (75%) memorable messages were expressed in the form of brief oral injunctions that prescribe some rules or advice to the recipient. Taking a cue from this, we also experiment with presence of imperatives in our model. In (Guerini et al., 2012), the authors attempt to understand the role of stylistic and readability features of abstracts on the success of a scientific article. This study is of interest to us because it seems intuitive that the more memorable the abstract of an article is, the higher the chances are of it being cited or bookmarked for future references. This idea transcends to communication on social media as well. For example, textual snippets posted on social networking sites like Twitter (tweets) are prone to receive more attention based on not only their content but also based on how they are written. For our project, it will be an interesting experiment to see if linguistic features such as emotions and sentiments evident in viral messages/text are also evident in memorability. An interesting sociopsychological study by (Heath et al., 2001) aims to understand the factors that affect the spread or virality of urban legends. Specifically, the authors explore the role that emotional selection (particularly selection for disgust) plays in the success of an urban legend in the ecosystem, drawing a parallel with natural selection. In (Resnik et al., 2009) , the authors introduce a new approach to the problem that focuses on the syntactic "packaging" of ideas, which is well suited to investigating the identification of implicit sentiment, or perspective. They establish a strong predictive connection between linguistically well motivated features and implicit sentiment. (Upal, 2005) expands on previous work in the area of memorability in narratives, particularly in investigations of the roles that intuitive, minimally counterintuitive, and maximally counterintuitive concepts play in memorability. In (Boers et al., 2005) , the authors use the results of previous studies of the effectiveness of etymological elaboration to analyze the effect of alliteration (the repetition of at least one consonant/sound in different words in a phrase) on recall. Their findings showed that alliteration definitely improves recall in two groups of students, one with exposure to etymological elaboration and the other without that exposure, students from both groups performed significantly better while recalling alliterative phrases. In (Proctor, 2013), the author claims that certain linguistic features, like rhyme and repetition are effective in increasing consumer memory of an advertisement. The goal of this study is to turn the focus of advertising away from brand naming strategies and instead direct it towards print advertising strategies. The author claims that this aspect of the advertising field is less-researched, but equally relevant in determining the impact that an advertising slogan has on the readers. We plan to use the rich feature set, described in the paper, comprising of repetition, rhymes, alliteration, etc. to augment our classifier's performance, with the intuition that these features might prove useful, at the very least in the advertising domain. For the scope of our project, we intend to experiment with some of the features that these studies suggest, such as brevity, personal focus and simple rule structure, and investigate whether these actually enhance memorability of messages.

3 Motivation

Memorability is an interesting concept in itself, and when coupled with Natural Language Processing, we opine that it poses a very intriguing research topic. Whats most appealing is that the area of memorability, and particularly cross-domain memorability, has not been extensively researched from a computational perspective. Specifically, previous work by (Cristian et al., 2012) aimed to explore features that contribute to memorability, but did so by focusing on a single domain. We think it would be interesting to expore more generic features that can affet memorability across multiple domains. Thus, the motivation behind this project is to develop a computational model that can predict memorability on any kind of message, using just the textual content that it entails.

4 Data

For training our model, we make use of the Cornell Movie Quotes Corpus, available on the website of (Cristian et al., 2012). The corpus contains pairs of 1790 movie quotes, where each pair consists of a memorable quote followed by a non-memorable one. The corpus is designed in such a way that both quotes in a pair are of similar length and were uttered by same protagonist. The quotes are tagged as memorable or not on IMDB , forming the basis of the annotations in the corpus. We use the movie corpus as the training data set for our cross-domain memorability classifier. For testing purposes, we collected data from three different domains as listed below:

1. **Advertising Slogans:** We noticed that many advertising slogans and punch lines are actually very memorable while others are not. Going by this, we collected a total of 139 advertising slogans from across different brands and different categories and these form our first test corpus.

2. **Political Slogans:** Political slogans are built for the sole purpose of attracting voters and therefore are intended to be memorable. The corpus contains a total of 176 political slogans which we have collected from Wikipedia.

3. **Educational Mnemonics:** As school-going children, we all made use of mnemonics to memorize the planets in the solar system or the color code of resistors, using them to recollect information. Thus, our intuition was that mnemonics would form a good corpus for testing memorability of text. This corpus is comprised of 144 mnemonics collected from Wikipedia.

We employed 3 judges to manually annotate the three data sets. We defined some rules for annotations for each of the three categories. These rules are listed in Table 1. We advised the annotators to mark a slogan as memorable if it satisfied at least one of the specified rules. We used the Fleiss Kappa metric to calculate inter- annotator agreement. Not surprisingly, we found that the inter-annotator agreement was really low, which reinforces an important aspect of memorability: its subjectivity. In other words, whether a message is memorable or not is highly subjective to the individual receiving the message. Table 2 1 highlights the inter-annotator agreement scores for all three categories. The political slogan dataset received the highest score of 0.4, which is a moderate score to begin with. Mnemonics received a score of 0.20, while advertising slogans received the lowest score of only 0.11. Because of the subjective nature of this topic, we retained the entire corpus that we designed, and marked quotes as memorable based on majority annotations, despite the low inter-annotator agreement score.

5 Approach

We wanted to develop a good knowledge of memorability and how it manifests in day-to-day lives. We started off by looking at psycho-linguistic and socio-linguistic surveys to determine the impact of the choice of linguistic features on the memorability of slogans and advertising strategies. Our next aim was to figure out how the concept of memorability extends across domains. In particular, we looked at features which manifest in political slogans, advertising slogans and educational mnemonics.

- **Ngrams:** We experimented with lexical, syntactic, mixture of both and pragmatic features. CDMC incorporates lexical ngrams and POS syntactic ngrams.

- **Emotion Strength:**Emotion strength is an interesting feature that captures the degree of emotion (both positive and negative) expressed in a quote. Our intuition tells us that

Corpus	Fleiss Kappa
Political Slogans	0.41
Advertising Slogans	0.11
Education Slogans	0.20

Table 1: Fleiss Kappa Scores for inter-annotator agreement on the three test sets.

Feature	Description
Lexical Ngrams	Word unigrams, bigrams, trigrams
Syntactic Ngrams	POS unigrams, bigrams, trigrams
Emotion Strength	Degree of emotion expressed in the quote
Profanity	Frequency of profane words in the quote normalized by length of quote
Repetition	Repetition of words in the quote
Named-Entity-Recognition	Presence of Named Entities in the quote
Length of Quote	Number of words in quote
Presence of 'you'	Presence of the word you in quote
Use of Imperatives	Use of commanding words in quote
Sentiment Polarity	Positivity/Negativity of sentiment in quote
Alliteration	Presence of alliteration in quote
Rhyme	Presence of rhyming words in quote
Syntactic Distinctness	Syntactic TF-IDF scores compared to Brown corpus
Lexical Distinctness	Lexical TF-IDF scores compared to Brown corpus

Table 2: A complete list of the features considered for use in classification.

the greater the degree of emotion, higher are the chances of the quote being remembered. One of the sentences in the movie corpus is "Well, this looks like the end of a terrible friendship". The sentence turns out to be memorable, having a high negative polarity of 0.86.

- **Profanity** We observed in our movie corpus that majority of the memorable quotes had profane words in them. We use a list of profane words to count the number of profane words in a quote. We normalize this feature value by the length of the quote.

- **Repetition:** It's common knowledge that repetition increases emphasis and consequently improves memorability. CDMC uses word repetition in a quote as a feature. A good example is the advertising slogan for mascara- "More defined. More conditioned. More beautiful lashes".

- **Named Entity Recognition:** The conservative Christian slogan "God made Adam and Eve, not Adam and Steve" has become popular. Although there may be reasons other than the presence of Named Entities responsible for its popularity, our classifier correctly identified it as memorable.

- **Length of Quote:** At times, the length of a quote is seen to affect it's memorability. Longer quotes are harder to remember. Consequently, shorter quotes assist memorability.

- **Presence of 'you':** Our preliminary analysis showed that a slogan became memorable because it connected with the audience. CDMC checks for the presence of 'you' in a sentence and uses it as a binary feature. Mahatma Gandhi's famous quote- "Be the change you wish to see" embodies this notion.

- **Imperatives:** Quotes having imperatives/commanding words tend to be remembered better. For instance, consider this quote from our movie dataset- "A little advice about feelings kiddo; don't expect it always to tickle".

- **Sentiment Polarity:** This ternary feature captures whether the sentence expresses positive (+1), negative (-1) or neutral (0) sentiment. Our analysis showed that quotes exhibiting negative sentiments tend to be more memorable.

- **Alliteration:** "Black is beautiful". Memorable right? This cultural movement started in USA in the 1960s by African Americans gained huge momentum.
- **Rhyme:** Time and again we see that rhyme helps us remember things better. The presence of rhyme was seen to be an important indicator of memorability. A good example is seen in education mnemonics. The famous mnemonic "When two vowels go walking the first does the talking", is commonly used by teachers to help students remember how to spell words with consecutive vowels.
- **Syntactic Distinctness:** In (Cristian et al., 2012), the authors show that generic syntax improves memorability. We use the concept of TF-IDF using POS tags from the Brown corpus and those from our train set to compute the distinctiveness quotient for a quote. "Think global, act local" has high syntactic distinctiveness quotient and therefore, high memorability.
- **Lexical Distinctness:** Contrary to our belief for syntax, we believe that use of unique lexicons/words fuels memorability. If you encounter a quote with a whacky word in it, chances are you'll remember it! We again make the use of the words in the Brown corpus to generate lexical distinctiveness quotients for quotes in our train set.

The full feature set is shown in Table 2 along with the feature descriptions.

6 Experiments

To test the usefulness of our features, we conducted two separate sets of experiments. First, we ran 5-fold cross-validation experiments on the training corpus to evaluate the performance of our classifier. In our second set of experiments, we trained a classifier on the training corpus and tested its performance on each of the test corpora. In each set of experiments, we experimented with different feature sets, to see which set gave us the best results. Overall, we found that the addition of our generic memorability features improved performance across all domains, including the training corpus.

6.1 Cross-Validation

Since the Cornell Movie Quotes corpus is a fully balanced dataset, a naive baseline classifier that simply assigned all instances to either class (memorable or non-memorable) would do so with an accuracy of 50%. Thus, for cross-validation, our baseline accuracy was 50%. We then performed several 5-fold cross-validation experiments with an 80-20 split, plugging in different feature sets in order to analyze the performance of those features. The results of these experiments are shown in Table 3. We then trained our model - an SVM classifier with a linear kernel - on a simple bag-of-words (BOW) feature set, using lexical unigram, bigram, and trigram counts as features. A classifier trained on this feature set significantly improves over the baseline, yielding a cross-validation accuracy of around 60.8%.[1] For our next experiment, we explored the accuracy gain from adding our full feature set (as enumerated in section 5) and syntactic N-gram counts, in addition to BOW features. We found that the addition of our generic cross-domain features slightly improved cross-validation accuracy, yielding a final accuracy of 62.8

6.2 Cross-Domain Testing

The crux of our project - and our main research focus - was to determine the usefulness of the generic feature set described previously in predicting memorability across domains. In order to do this, we performed several experiments in which we trained an SVM classifier on the training corpus, using different feature sets. We then tested the performance of this classifier on our test corpora, noting the F-1 scores and comparing them against the baseline for that domain (a naive classifier that classifies all instances as memorable). At this juncture it's important to recognize that a classifier that improves over this baseline might not necessarily have a *higher* F1 score; since the assumed baseline has a recall of 1.0, an improved F1 might not be extremely easy to achieve, given the subjective nature of the classification problem at hand. A good step in the right direction would be an increase in precision that's accompanied by a reasonable recall.
Going in, we hypothesized that different domains

[1]These results are in agreement with those presented for the same experiment in Danescu-Niculescu Mizil et al. (2012)

Experiment	Feature set	Accuracy
Baseline	-	50.0 %
BOW	Lexical N-gram counts	60.8 %
CDMC	Full feature set	62.4 %

Table 3: Results of cross-validation experiments for linear-kernel SVM classifier trained on various feature sets.

Feature Set	Kernel	Precision	Recall	F1 score
Baseline	-	0.657	1.00	0.793
BOW	Linear*	0.67	0.54	0.6
Generic cross-domain features only	RBF	**0.69**	0.57	0.62
Syntactic N-grams and generic cross-domain features	RBF	0.65	0.52	0.58
BOW and generic cross-domain features	Linear	0.67	**0.64**	**0.66**

* When an RBF kernel is used here instead, the trained SVM behaves exactly like the naive baseline, classifying all instances as memorable.

Table 4: Cross-domain classification results in the advertising domain, using an SVM classifier trained on the movie quotes corpus

might perform better with different feature combination, and that different kernel types might produce better results for different domains. These intuitions proved correct, as can be seen from the results of these experiments in Tables **4, 5** and **6**. Possible reasons for these differences are discussed in section 7. Below, we describe performance in each domain.

6.2.1 Advertising Slogans

The advertising slogan corpus contains about 92 slogans that were annotated as memorable, and 48 that were annotated as non-memorable, for a total of 140 slogans. A naive baseline classifier classifying this test set would thus have a precision of 0.657, with a recall of 1.0, for an overall F1 score of 0.793.

In our experiments, we started by training an SVM with a linear kernel on the movie quotes corpus with only BOW features. This yielded a classifier that, despite a slightly higher precision, performed poorer than baseline overall, since its recall was far lower. Using only generic cross-domain features improves marginally over this model, while the addition of syntactic N-grams actually reduces both precision and recall. The use of BOW and cross-domain features together produces the highest F-1 score of 0.66.

Statistically speaking, the feature set that combines BOW and cross-domain features has the best performance compared to the naive baseline, considering that it has a higher precision. It's also worth noting that the addition of cross-domain features is a significant improvement over the recall when only BOW features are used. Further, it's interesting that this very small feature set (30 features) actually produces the highest precision (see section 7 below for a more detailed discussion of these results).

6.2.2 Mnemonics

The mnemonics test corpus is completely balanced, with 72 memorable and 72 non-memorable quotes. As a result, a naive baseline classifier for this corpus would have a precision of 0.5, a recall of 1.0, and an overall F1 score of 0.67. From the results in Table 5, it is clear that our classifiers could not provide significantly improved performance over a naive baseline classifier. Again, however, we demonstrate that the use of our simple cross-domain feature set provides a significant improvement in precision over a BOW model. We also observe again that the addition of syntactic N-gram counts doesn't impact performance.

Aside from these observations, however, there doesn't appear to be a significant improvement over the baseline using any of these feature set combinations. We discuss possible reasons for this in the discussion section below.

6.2.3 Political Slogans

The corpus comprised of political slogans contains 143 slogans, of which 77 are memorable and 66

Feature Set	Kernel	Precision	Recall	F1 score
Baseline	-	0.5	1.00	0.67
BOW	Linear	0.47	0.39	0.42
Generic cross-domain features only	RBF	0.52	0.9	**0.66**
Syntactic N-grams and generic cross-domain features	RBF	0.52	0.9	0.66
BOW and generic cross-domain features	Linear*	**0.53**	0.82	0.64

* When an RBF kernel is used here instead, the trained SVM behaves exactly like the naive baseline, classifying all instances as memorable.

Table 5: Cross-domain classification results on the mnemonics test corpus, using an SVM classifier trained on the movie quotes corpus

Feature Set	Kernel	Precision	Recall	F1 score
Baseline	-	0.54	1.00	0.70
BOW	Linear	**0.60**	0.51	0.55
Generic cross-domain features only	Linear	0.50	0.47	0.48
Syntactic N-grams and generic cross-domain features	Linear	0.52	0.43	0.47
BOW and generic cross-domain features	Linear	0.59	**0.6**	**0.59**

Table 6: Cross-domain classification results on the political slogans corpus, using an SVM classifier trained on the movie quotes corpus

non-memorable. Thus, a naive baseliner classifier for this corpus would achieve a precision of 0.54. Our next attempt was to use the BOW feature set; surprisingly, this yielded a significantly higher precision than the baseline, while on the other hand, subsequent experiments with cross-domain features and syntactic N-grams produced worse results. This is different from the trend observed while testing on advertising slogans and mnemonics, where the performance of BOW features was similar or worse than the baseline.

However, in keeping with previous observations, we observe again that the use of syntactic N-gram features doesn't affect the performance of the classifier when combined with cross-domain features. Instead, the best results for this domain are again obtained when using BOW features in combination with cross-domain features; this yields an F1 score of 0.59, with a very high recall of 0.6 as compared to the other experiments. Here, our observations are similar to those in the other two domains; the addition of cross-domain features to BOW features increases the precision and/or recall of the model, making it an improvemenet over the BOW features.

7 Discussion

During cross-validation, we obtain an improvement over baseline performance, lending credence to our hypothesis that these cross-domain features are useful predictors of memorability. However, during cross-validation, we found that some features that are less straightforward to compute - like alliteration and rhyme - provided incorrect feature values to the classifier, thus hampering the classifier's performance. We also found that while our original intuition was that the length of a piece of text plays a great role in its memorability, the nature of the training set (wherein it contains memorable and non-memorable quote pairs that are of equal length) renders this feature useless to the classifier. Thus, our final cross-domain feature set includes all the features described in section 5 above, *except the rhyme, alliteation, and length features.*

Our intuition for cross-domain testing, following from the observations during cross-validation, was that the classifier reporting best performance would have to make use of these cross-domain features, perhaps in combination with bag-of-words featurs of syntactic N-gram counts. We also expected that using BOW or syntactic N-gram features in isolation would not improve over baseline performance by much.

For the most part, these expectations held up - classification performance on all three test sets was improved by the use of the generic cross-domain features that we developed, in addition to BOW features. On the other hand, the addition of syntactic N-grams never improved performance, and for both advertising slogans and mnemonics, a classifier trained on BOW features alone reported poorer performance than one that was trained on BOW features and cross-domain features. This was in keeping with our original intuition that the memorability of language has more to do with lexical structure than with syntactic structure.

An important observation that we made repeatedly was that overall recall for all the domains remained consistently suboptimal[2]. This suggests that although our feature set helps to correctly classify several test instances and improve precision on a subset of test instances in each domain, it fails to identify several other instances as memorable. We attribute this to an incompleteness in the feature set that can only be addressed by developing better computational methods to exract features like rhyme, alliteration, humor, etc.

Overall, the poorest cross-domain accuracy performance was on the mnemonics test set, where classifiers trained on all feature sets reported very little gain in precision over baseline performance. We posit that this is because of the nature of mnemonics in general - since they're most often spoken and repeatedly uttered, it's possible that more phonetic features like rhyme, meter, rhythm, etc. need to be analyzed to boost performance on mnemonics. Further, it's entirely possible that lexical distinctness is *not* a characteristic trait of memorable mnemonics - intuitively, one might expect memorable mnemonics to contain common and simple vocabulary in order to aid memorability.

This is in keeping with our general observation throughout cross-domain testing: final performance depended on some way on the nature of the test set itself. Thus, while we achieved our broader goal of identifying features useful in building a cross-domain memorability classifier, the variability in these results seems to suggest that a memorability classifier with maximum performance might actually require some fine-grained knowledge of the domain, in addition to using these cross-domain features.

[2]This wasn't true in the case of mnemonics, where we reported recall values as high as 0.9; however, in those cases, the improvement in precision over the baseline was minimal, which leads us to suspect that the classifier acted more like a naive classifier, classifying a majority of instances as memorable, rather than making more intelligent classifications.

8 Conclusion

In this paper, we propose a cross-domain memorability classifier that employs linguistic features which affect the memorability of a piece of text. We experimented with a variety of lexical, syntactic and pragmatic features. It was exciting to see that some features which intuitively make sense give good performance in the computational framework for recognizing memorability.

9 Future Work

One of the major pitfalls of our work was the poor inter-annotator agreement. We believe a stronger notion of memorability coupled with a large number of annotations should give better performance. In present work, we train on dataset from one domain. We feel that training our memorability classifier across multiple domains will perform well during cross-domain testing. Another drawback is the absence of feature selection. We can implement feature selectionto get an improvement in performance. Future work in this direction could explore the use of contrasting ideas and the presence of humor to build a more wholesome feature set.

Acknowledgement

Our sincerest thanks to Smara for her helpful suggestions and continued guidance throughout the course of the project. We'd also like to acknowledge Danescu-Niculescu-Mizil et al. for providing the intellectual basis for this project.

References

Cristian D.N. Mizil, Justin Cheng, Jon Kleinberg, and Lillian Lee. 2012. You had me at hello: How phrasing affects memorability. *In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics.* Long Papers-Volume 1, pages 892-901. ACM.

Mark L. Knapp, Cynthia Stohl and Kathleen K. Reardon. Memorable Messages. *Journal of Commu-*

nication 31.4 (1981): 2741. Wiley Online Library. Web. 3 Apr. 2014.

Marco Guerini, Alberto Pepe, and Bruno Lepri. 2012. Do Linguistic Style and Readability of Scientific Abstracts Affect their Virality? *ICWSM.*

Leanne Proctor. 2013. Successful stradegies?: linguistic elements used in advertising.

Chip Heath, Chris Bell and Emily Steinberg. Emotional selection in memes: The case of urban legends. 2001. *Journal of personality and social psychology.* Volume 81, Number 6, pages 1028.

Stephan Greene and Philip Resnik. More than words: Syntactic packaging and implicit sentiment. 2009. *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics.* pages 503-511.

M. Afzal Upal. Role of Context in Memorability of Intuitive and Counterintuitive Concepts. 2005. *Proceedings of the 27th Annual Meeting of the Cognitive Science Society.* pages 2224-2229.

Frank Boers and Seth Lindstromberg. Finding ways to make phrase-learning feasible: The mnemonic effect of alliteration. 2005. *Journal of System.* Volume 33, Number 2, pages 225-238. Elsevier.