



Theses and Dissertations

---

2018-04-01

## Pun Strategies Across Joke Schemata: A Corpus-Based Study

Robert Nishan Crapo  
*Brigham Young University*

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Linguistics Commons](#)

---

### BYU ScholarsArchive Citation

Crapo, Robert Nishan, "Pun Strategies Across Joke Schemata: A Corpus-Based Study" (2018). *Theses and Dissertations*. 6739.

<https://scholarsarchive.byu.edu/etd/6739>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

Pun Strategies Across Joke Schemata:

A Corpus Based Study

Robert Nishan Crapo

A thesis submitted to the faculty of  
Brigham Young University  
in partial fulfillment of the requirements for the degree of

Master of Arts

Dallin D. Oaks, Chair  
Deryle Lonsdale  
Mark Davies

Department of Linguistics

Brigham Young University

Copyright © 2018 Robert Nishan Crapo

All Rights Reserved

## ABSTRACT

### Pun Strategies Across Joke Schemata: A Corpus Based Study

Robert Nishan Crapo  
Department of Linguistics and English Language, BYU  
Master of Arts

In the linguistic study of humor, research has largely been centered around the formulation of models and theories or the dissecting and categorization of jokes. Because of the often difficult-to-categorize aspects of verbal jokes, much time has been spent trying to create taxonomies for humor types and mechanisms. Linguists such as Raskin and Attardo have sought to categorize all verbal humor according to various functional elements (Attardo & Raskin, 1991). Such elements include, but are not limited to, the logical mechanism that drives the humor in the joke or the situation where the joke takes place. These categorizations are helpful in understanding the potential components of a given joke. However, relatively few studies have sought to quantify and qualify the distribution of these components across real-world data. This study seeks to understand the distribution of some of these categorizations laid out by Raskin and Attardo across joke topics, namely pun wordplay and narrative strategy. To do this, an original 100,000 word joke corpus was designed and compiled consisting of four joke topics: Marriage, Politics, Animals, and Food. Through some manual sorting and Python programming, jokes were labeled according to wordplay strategy and narrative structure. A subsequent statistical analysis was carried out to determine whether there exists a pattern of specific joke strategies when dealing with children's humor versus adult humor.

Keywords: puns, corpus, joke, schemata

## ACKNOWLEDGEMENTS

In the writing of this thesis I am grateful for those who have helped me grow in this program and foster my enjoyment of linguistic research. I am grateful to the faculty at BYU who have shown me the many facets of linguistics in both theoretical and applied forms.

I want to thank Mark Davies for stoking my interest in corpus linguistics. Research, honestly, wasn't enjoyable for me until I learned the versatility and power of constructing a corpus and searching the data.

I want to also thank Deryle Lonsdale for answering my many questions. I am grateful for his vast repository of knowledge and his dedication to getting things right. I appreciate his thoroughness and his wisdom.

Finally, I want to thank my chair Dallin D. Oaks for the many conversations about using wordplay in varying contexts of life. The opportunity to work with him and his software fueled my desire to learn computational linguistics and break out into applied linguistics. I will never forget the practical applications of linguistic research that he has imparted to me on almost a daily basis. He has been a conscientious mentor in my graduate experience and has been helpful every time I have approached him for assistance. My experience at BYU and this program was the perfect platform for my future pursuits.

Thank you all.

## Table of Contents

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iii
Table of Contents .....	iv
List of Tables .....	vi
List of Figures.....	vii
1. Introduction.....	1
2. Literature Review .....	4
2.1 Humor Theories.....	4
2.1.1 Philosophical Theories of Humor .....	4
2.1.2 Linguistic Theories of Humor.....	6
2.2 Pun Classification.....	10
2.2.2 Ambiguity.....	10
2.2.3 Homonymous Puns.....	12
2.2.4 Near Puns .....	14
2.2.5 Polysemous Puns.....	15
2.3 Perspectives on Puns .....	18
2.3.1 Pun Usage.....	19
2.4 Humor Research with Corpora.....	24
3. Methodology .....	27
3.1 Corpus Design.....	27

3.2 Corpus Construction.....	28
3.2.1 Tools .....	28
3.2.2 Procedure.....	29
4. Results.....	45
4.1 Quantitative.....	45
4.1.1 Chi-square Testing.....	47
4.1.2 Content Word Frequency .....	51
4.2 Qualitative .....	52
5. Conclusion .....	56
5.1 Answers to Research Questions.....	56
5.1.1. Pun and Topic Connection.....	56
5.1.2 NS and Pun Connection.....	57
5.1.3 NS and Topic Connection .....	58
5.1.4 Characteristics of Joke Schemata: Child versus Adult.....	58
5.2 Limitations.....	59
5.3 Implications.....	61
5.4 Future Research .....	62
6. References.....	64
7. Appendix .....	68
7.1 Data Sources for Corpus and In-Text Examples .....	68
7.2 Survey of Children’s Joke Books.....	71

**List of Tables**

Table 1 Research Questions.....	3
Table 2 Sub Corpora Totals .....	45
Table 3 Observed Versus Expected Pun/Topic Totals .....	48
Table 4 Pun Prevalence (percentage) among Topics .....	48
Table 5 Observed versus Expected Pun/NS Totals.....	49
Table 6 Pun Proportions (percentage) per NS .....	50
Table 7 Observed Versus Expected NS/Topic Totals.....	50
Table 8 Most Common Words per Sub Corpus.....	51
Table 9 Research Questions Revisited.....	56

**List of Figures**

Figure 1 Food Pun Comic .....	21
Figure 2 Trump Comic.....	22
Figure 3 Homonymous Puns across Topics .....	46
Figure 4 Polysemous Puns across Topics .....	47



## 1. Introduction

Within the field of linguistics, semantic and pragmatic (Raskin, 1984; Attardo & Raskin, 1991) as well as structural analyses (Pepicello & Green, 1984) have been important in delineating joke components. These types of studies have led other researchers to build and expand off these taxonomies (Hempelmann, 2004; Oaks, 2010). These approaches have done much in helping us understand the ways humor can arise from language, but there are few empirical studies that trace these strategies to their natural contexts (Attardo, Raskin, & Ruch, 1993; Attardo, Attardo, Baltes, & Petray, 1994). What determines the strategies for a given joke? Are joke strategies simply left to the artistic license of the joke teller, or is there a more innate internal connection between joke strategy and content? Using real-world data, this study seeks to determine whether there are prominent objective strategies that are tied to certain scripts or topics.

Within humor a joke teller has various available approaches to take. The joke teller may use positive and negative types of humor. He or she may use satire, parody, irony, wordplay and so on. These strategies provide the main impetus or mechanism behind the joke (Raskin, 2008). Notable among the available humor strategies is the pun. Punning takes advantage of the semantic, structural, or phonemic ambiguity found in English (and other languages) in order to subvert an established expectation. Common in joke formats, puns are a divisive topic among joke tellers and audiences as some people greatly enjoy puns while others detest them.

Pun taxonomies across linguistic studies lack uniformity. Although some researchers have investigated pun structure (Attardo, 1994; Hempelmann, 2004; Partington, 2009) there is still work to be done in understanding puns as a humor strategy.

In addition to the central mechanism behind a joke, there are different ways a joke can be told. Jokes can offer a quick punchline in a one-line quip or a simplistic riddle. Jokes can also take the form of stories or humorous conversations. These different mediums provide different flavors of comedy to the audience. Within these mediums the joke teller can control the degree of complexity the humor offers as he or she seeks to gain laughs from the audience. Narrative strategies such as riddle, knock-knock, or story formats are independent from the central mechanism strategy in the sense that different mechanisms can interact with different narrative forms (Attardo & Raskin, 1991). However, it may be that these strategies follow certain patterns. The question is, if joke strategies do follow a pattern, what is dictating or at least influencing that pattern?

Perhaps the most central aspect of a joke is the content itself. Jokes are often built on real life experiences and events that are shared by various groups of people. It is through humorous tellings of these experiences that people can connect and relate to one another (Knight, 2013). These common experiences known as scripts or schemata provide the meat of the joke and the basis on which joke strategies function. Joke tellers seek to find the right scripts that are relevant to their audience and apply joke strategies to those scripts in order to incite laughter (Hempelmann & Ruch, 2005). The difficulty in this task comes from knowing which scripts to talk about. Understanding what is meaningful and relevant to the audience is key but often elusive. For example, a football joke may be effective for a football fan demographic but maybe not for soccer fans or people that only possess a casual interest in sports. Moreover, the specificity of content may affect the effectiveness of the joke overall. Familiarity with a certain joke topic may lead to laughter, but not always. This can be seen when exploring the differences between jokes fashioned for children and jokes fashioned for adults.

In children's humor certain topics center around a child's experience, interest and imagination. The ideas that occupy the mind of a child are different from what occupies the minds of adults in cognitive complexity as well as focus. These jokes also tend to be simplistic as children are still learning about the scripts and social norms of their respective culture. With a lack of knowledge for social scripts, children find humor through other means. In this situation wordplay can be used to elicit laughs while not requiring a great deal of set-up or script background knowledge. Indeed, the majority of jokes constructed by children uses lexical humor (Sutton-Smith, 1976). This leads us to ask a broader question concerning humor types. If jokes are fashioned towards a particular audience, how do the strategies and content involved vary based on different audiences? In this study the following questions will be explored:

Table 1 Research Questions

- 
1. Is there a connection between pun usage and joke topics, and what is that connection?
  2. Is there a connection between narrative strategy and pun usage?
  3. Is there a connection between narrative strategy and joke topics?
  4. Other than their content, do jokes for children generally follow different strategies than jokes for adults?
- 

To help answer these questions, an original corpus was designed and compiled to understand the strategic characteristics of jokes according to joke topic, with an emphasis on puns. Before making any analysis, an exploration of the taxonomies and past linguistic research in humor must be explored.

## 2. Literature Review

In this discussion we will explore the evolution of humor studies and do this across disciplines. We will start with the largely philosophical theories that discuss the emotional and psychological motivations for humor. We will then look at the leading linguistic theories and mechanisms that comprise verbal humor. Finally, we will explore the classifications and usages of puns.

### 2.1 Humor Theories

For the purposes of this thesis, the discussion of humor theories will focus mostly on linguistic perspectives, although it is important to mention some psychological and philosophical paradigms. Rather than giving an exhaustive review of humor theories, this thesis will focus more on the specific historical path that has led to the study of puns. For extensive surveys of humor across disciplines and time, see the major reviews found in *The Psychology of Laughter: A Study in Social Adaptation* (Piddington, 1933) and *Laughter and the Sense of Humor* (Bergler, 1956). Much has been added to these works since being published, but the earlier works act as landmark examples of humor research. For more modern surveys see *The Primer of Humor Research* (Raskin, 2008) and *Comic Relief: A Comprehensive Philosophy of Humor* (Morreall, 2009).

#### 2.1.1 Philosophical Theories of Humor

Among the most prominent philosophical theories for humor there are three that stand out: the *superiority theory*, the *relief theory*, and the *incongruity theory* (Raskin, 2008). As different philosophers, psychologists, and religious figures have weighed in on the role of humor in society, the perception of humor has evolved. Largely shaped by western thought, the use of humor has shifted from negative types to more positive/neutral types.

Perhaps the most negative and primitive view of humor came in the form of the *superiority theory*. Philosopher John Morreall cites western philosophy and the Bible as sources that view laughter as a decidedly indulgent and immoral act (2009). As a stoic, Plato was a decided opponent of humor and laughter, believing that it was a loss of control that led to more indulgent vices. Additionally, various Christian groups such as monks and Puritans forbade joking one with another. Known for his plain speech discourses and letters, Archbishop of Constantinople, John Chrysostom said the following concerning humor:

Laughter often gives birth to foul discourse, and foul discourse to actions still more foul. Often from words and laughter proceed railings and insult; and from railing and insult, blows and wounds. (Chrysostom, 1889)

Such words equate laughter with aggression that destroys interpersonal relationships and personal morality. This view of humor prevailed for hundreds of years up until the end of the 19<sup>th</sup> century when two new theories emerged.

In the 1800's the perception of humor in the Western world moved away from the negative aspects of humor in society. Perhaps the driving force of this movement came from the conception of two new humor theories, with one of them being the *relief theory*. Relief theory took on the metaphor that the pressures of life beget nervous energy that spills over psychologically and physically to humor and laughter (Morreall, 2009). In this theory humor does not require a target to mock. Rather, humor was an intellectually and emotionally freeing tool that helped people deal with the constraints placed on them by laws and societal norms (Freud, 1901). Thus, through relief theory, humor was seen as a more positive way to cope with the challenges of day-to-day life.

The final prominent and most relevant theory to this study is *incongruity theory*. The second theory developed in the 1800's, incongruity theory states that humor arises from

situations, utterances, or events that break from logic or rational thought (Attardo, 1994; 2008). Often this is exemplified in humor that is ridiculous or absurd. This is also the basis for many joke forms (Hempelmann, 2004; Partington, 2009). Jokes often contain two simple components that capitalize on this type of humor: the setup and the punchline. The setup first creates a certain expectation in the mind of the audience. Then, when the punchline is reached, there is a moment when the original expectation is flouted. However, the mind realizes that the incongruity actually is relevant to the initial set up, and the mind perceives this as funny. The perception that incongruity can form the basis of humorous reactions further supported the idea that humor did not necessarily have to be negative. Incongruity theory has formed the foundation of prominent modern linguistic theories of humor.

### **2.1.2 Linguistic Theories of Humor**

Whereas past philosophers and psychologists sought to find the emotional mechanisms behind humor, linguists or those attuned to linguistic or language-related approaches have sought to delineate and label the formalized components of humor, namely the components in verbal jokes. Jokes require a deliberateness in the sense that a humorous situation or anecdote is constructed with an audience in mind (Raskin, 1984; Beard, 2008). Incongruous happenstance may create collateral humor, but a joke is fashioned with intent and purpose. As a side note, *verbal* does not refer to the oral performance of jokes but rather the linguistic aspects of a joke. Verbal jokes often contain an incongruity or rather a difference between what is expected that will happen in the joke and what actually occurs (Raskin, 2008). This is then followed by a resolution in the joke that elucidates the initial incongruity while validating its relevance. This incongruity resolution process has been investigated generally across disciplines (Morreall,

2009). Among the leading linguistic researchers, Victor Raskin has built upon the theory of incongruity with his theory of semantic scripts.

Raskin's theory states that a joke receives its power from the overlapping of social scripts. Scripts or schemata are mental representations, formed through experiences and memories of things, people, or situations that culminate in a "typical experience" in the mind. These schemata are widespread and common to human experience (Cook, 1994). Raskin posits that these ubiquitous societal scripts provide a basis for verbal jokes. Coined as the "Script based Semantic Theory of Humor" (SSTH), this theory states that the overlapping of scripts in a text can generate an incongruity type humor given the following stipulations:

- (i) The text is compatible, fully or in part, with two different [semantic] scripts.
- (ii) The two scripts with which the text is compatible are opposite [...]. The two scripts with which the text is compatible are said to overlap fully or in part on this text. (1984)

When two scripts or experiences that are familiar to a person converge with grammatical or pragmatic similarity, a connection is made in the mind (Raskin, 1984). However, these scripts must also oppose each other either semantically or pragmatically. The incongruity and resolution occur when a reader processes a text with the primary script in mind. When the punchline of the joke is reached, there is a sudden shift in the mind of the reader to the opposing script. This sudden incongruity and resolution create the humor as the reader realizes there were two scripts all along.

To illustrate this theory, Raskin gives the following joke as an example:

"Is the doctor at home?" the patient asked in his bronchial whisper. "No," the doctor's young and pretty wife whispered in reply. "Come right in."

The text given in this example contains the two scripts of DOCTOR and LOVER. The initial part of the joke seems to indicate that a patient is in need of seeing the doctor. The "bronchial

whisper” points to a potential ailment that requires tending. However, the script DOCTOR shifts to LOVER in the punchline. The wife’s response reveals that the purpose of the patient’s visit is not to receive medical care. The illicit affair is revealed when the expectation that the patient should be sent away is unmet. Additionally, looking back through the whole text, it is evident that there are clues to this affair and that both scripts took place the entire time. The whisper given by the patient could mean he is suffering from an ailment, but it could alternatively mean that he is seeking to be secretive. This goal is supported and reciprocated when the wife responds with a whisper as well. Both stipulations of Raskin are found within this joke as it involves a compatibility of two scripts in one text that oppose each other.

When these inconsistencies are bound together with the similarity of the scripts (or language), humor is created (Raskin, 1984; Hempelmann, 2004). Jokes can be short or elaborate, but many of them rely on this principle. This is often illustrated with the setup: “What’s the difference between...” or “What do X and Y have in common?” The humor comes with either an explicit or implicit comparison. This opposition of scripts found in the SSTH would soon become the foundation for another defining theory in linguistic humor research.

#### ***2.1.2.1 Linguistic Components of Jokes***

In the forefront of linguistic research, Raskin and Attardo have created the General Theory of Verbal Humor or GTVH (Attardo & Raskin, 1991), which is an expansion of Raskin's SSTH (Raskin, 1984). Their work seeks to organize in great detail the structural mechanisms of jokes. The tension or set up of a joke can be created in multiple ways—much like various physical mechanisms can create both potential and kinetic energy. Under the GTVH these mechanisms are *Script Opposition* (SO), *Logical Mechanism* (LM), *Situation* (SI), *Target* (TA), *Narrative Strategy* (NS) and *Language* (LA) (Attardo, 1994). In this theory the Script Opposition



(SO) remains the central driving force of the joke, while the five aforementioned additional components or “Knowledge Resources” (KR) act as auxiliary support for the execution of the joke.

The Logical Mechanism refers to the way multiple scripts are overlaid. This mechanism can range in complexity and usually relies on a punchline. This punchline can take the form of a single phrase that can be interpreted as figurative or literal or simply an utterance that can apply to both introduced scripts. Often this mechanism can simply set up a single word that acts as the pun of the joke. In other words, all wordplay rests on the logical mechanism of the joke.

The last four knowledge resources of the GTVH are simple in their role. The Situation refers to the setting that the joke creates (e.g. a park, grocery store, etc.). The Target refers to the butt of the joke, like an ethnicity, social class, gender, or specific person and is meant for mocking purposes (not all jokes contain a TA). Jokes with targets are usually classified as negative humor that is often associated with the superiority theory of humor. The Narrative Strategy is the way the joke is presented according to narrative format. The joke can take on such forms as a story, riddle, knock-knock setup, and so on. This is the strategy that the joke teller uses to get to the punchline of the joke. Finally, the Language component of the joke dictates all of the details necessary for the audience to understand the joke. For jokes with structural ambiguity, this component plays an important role in creating the specific wording of the joke that makes the wordplay viable. The importance of wording in jokes can be found with this falsely attributed Confucianist admonition:

“Man who put head in fruit drink get punch in nose” (Oaks, 2010)

The reduced syntactical structure of this joke, where a count noun meaning occurs without a preceding determiner, allows for the ambiguity of *punch* to function as both a party drink as well

as assault by a disgruntled party host. There is much more to say concerning these KRs, but such is not requisite for this study.

These knowledge resources make up the components used in verbal jokes. They serve as a means to categorize each type of verbal joke. All of these components can be found in the corpus created for this thesis, but the main study and analysis will center on the Logical Mechanism (puns), with some discussion of Narrative Strategies.

Now that we have explored the major components of jokes in general, we can turn our attention to the main issue of puns. The past research on puns is vast, and the number of pun taxonomies are great. However, these studies lack consistency as each researcher provides an additional nuance to pun classification. Thus, before we can confidently label puns across thousands of real-world jokes, we must create a standard for our classification. This starts with the question, what is a pun?

## **2.2 Pun Classification**

Under the GTVH, puns are categorized as logical mechanisms. As mentioned above, in jokes that use puns, puns are the driving force of the incongruity and resolution that elicits the humorous response. In this section we will discuss the characteristics of puns and how they differ from other logical mechanisms that are similar in form. We will also explore and delineate the different types of puns in order to create a labeling standard for the corpus. To do this, we will first investigate the ambiguities in English that make pun jokes linguistically viable.

### **2.2.2 Ambiguity**

The foundation of incongruity humor and pun jokes is ambiguity. The fact that intent, situations, or words themselves can be ambiguous allows the overlap of two scripts. Simple script overlap and opposition can use an audience's initial unawareness of one script possibility

to lead them down a single interpretive path until they are surprised with an additional one. The most straightforward instance of this (at least structurally) is punning.

The mechanism of the pun comes through the overlaying of scripts and lexical, structural or phonemic ambiguity. Such an example can be seen in the following joke:

By sheer coincidence all the sheep looked the same. (Funny Puns and One Liners, 2018)

The script of shearing sheep is perfectly overlaid with the common collocation of “sheer” with “coincidence”. This ambiguity ties the two scripts together and allows for two parallel meanings. The identical sound of both words creates this basic example of punning. However, not all puns follow such a clean structure. Notice the contrast with this joke:

What do sea monsters eat for lunch? Fish and ships. (Funny One-liner #1251, 2018)

In terms of phonemic structure the punchline does not perfectly coincide with both scripts.

However, there is enough similarity in the words “ships” and “chips” phonemically to elicit a mental connection, especially within the semantic association of “fish and chips”. Often jokes will contain punchlines that do not contain perfect homophonic alternates in punchlines but rely on the pragmatic background to make the connection. This slightly more (phonemically) complex pun is merely one example of the variations and structures that a pun can take on.

Sometimes the grammatical structure is also imperfectly matched between the two meanings as in the earlier Confucius joke about punch in the nose. Beyond this there are many more linguistic environments that can provide the foundation for puns and wordplays. However, exploring all environments would be a lengthy process that would not add to the central focus of this thesis.

For a much greater examination of these environments see Oaks’ *Structural Ambiguity in*

*English: An Applied Grammatical Inventory* (2010) wherein he lists various grammatical formulas that produce possible wordplays.

Now that we have examined the linguistic source underlying puns, we turn to the issue of taxonomy. Taxonomies are rarely exhaustive in nature and thus remain in dispute across linguistic disciplines. Although the classification used in this thesis will be informed by various approaches, ultimately the final taxonomic discrimination will be unique to this analysis. Additionally, it must be noted that the main point of this thesis is not to create an exhaustive taxonomy of all puns. Rather the purpose is to create working definitions that can be used as standards for labeling the data. Ultimately this will lead to three categories according to logical mechanism: *homonymous puns*, *polysemous puns*, and a broad category labeled *non-pun* that contains all other types of jokes.

### **2.2.3 Homonymous Puns**

Historically, some controversy has existed about the origin, classification, and desirability of puns. The word *pun* does not have a clear etymology or meaning, as the *Oxford English Dictionary* states that it is “of uncertain origin”, with a speculative connection to the Italian *punctilio*. Additionally, the definition is vague in terms of its linguistic identity:

The use of a word in such a way as to suggest two or more meanings or different associations, or of two or more words of the same or nearly the same sound with different meanings, so as to produce a humorous effect; a play on words (2017)

The word “different” becomes problematic when dealing with wordplays. There are several ways in which identical words can be “different”. Although various linguists and literary philosophers have weighed in on this subject, ultimately the delineation must be taken on a case-by-case basis and explicitly outlined as to what is a pun and what is not a pun.

In discussing wordplay many linguists agree that homonymy is central to the identity of puns (Attardo, 1994; Bell, Crossley, & Hempelmann, 2011; Partington, 2009). However, this distinction is still unclear when labeling specific jokes. In one study this fact is evident as the homonyms are assumed to be pun sources as illustrated by the following joke:

A: John found a shell on the beach. (Where shell means “discarded marine carapace”)

B: That’s a coincidence. Yesterday, I found a hand grenade. (Forcing shell to mean “artillery round”) (Partington, 2009)

The dual meaning of *shell* as both artillery and a shellfish exterior are different in superficial meaning, but they fundamentally contain similar characteristics (hard and hollow container that protects a sensitive interior). This example illustrates a difficulty in taxonomy as there can either be a confusion for what theoretically constitutes a homonym, or there can be difficulty in identifying true homonyms.

At times homonyms can be difficult to identify because of the historical knowledge required for each instance. For example, the word “pine” can take on two different meanings: a noun meaning “(a)ny of the numerous evergreen coniferous trees” or a verb form meaning “to endure pain”(OED). These two entries are apparently different, but how can we tell that they are not related through some sort of historical association? The *Oxford English Dictionary*, shows that they do not share a common etymological root. The meaning that refers to a tree comes from the Latin *pīnus*, whereas the meaning pertaining to pain comes from a Germanic root evident in words such as *pijnen* (Dutch), *pína* (Old Icelandic), and *pīnōn* (Old High German). Both the respective meanings and origins of the word “pine” are unrelated. These differences along with the fact that these two iterations of *pine* have identical orthographic and phonemic structure, identify these two counterparts as homonyms. Thus, through accidental means these words coincide in form, allowing the opportunity for a homonymous pun. For the purposes of this

thesis, “homonymous puns” will refer to jokes that hinge on either an acoustically or orthographically ambiguous word that also originates from a different etymological root from its counterpart form.

Beyond word-for-word homonymy, jokes in this corpus were labeled as puns if they approximated a word even with added or deleted phonemes as well as word boundary changes. Also, compounds and/or word blends that contained homonymous components were deemed as puns. This can be seen with the following joke:

Knock Knock Who's there? Artichokes. Artichokes, who?  
Artichokes when he eats too fast! (Food Jokes, 2017)

This joke uses “artichoke” as a punchline that doubles as the name of the vegetable as well as the name of a person with a corresponding action (Artie chokes). This variation leads us to discuss another categorization of puns.

#### **2.2.4 Near Puns**

One subset of homonymous puns are paronyms or ‘near puns’. These puns, unrelated in meaning or origin, sound similar but are not exact (Attardo, 1994). These differences can be due to a single phonemic difference such as a simple minimal pair (“bin” vs “pin”) or an additional phoneme such as with “bed” versus “bread”. These wordplays will often involve forceful manipulation of words to resemble other words that have pertinent meanings to the main scripts. This can come in the form of switching out a word for another existing word with similar phonetic features in order to create the script connection. One such instance can be seen here:

Q: What do squirrels give each other for Valentine’s Day?  
A: Forget-me-nuts. (Joke #11654, 2018)

The punchline given here is reliant on the script of giving flowers for Valentine’s Day. Forget-me-nuts is a phrasal invention whose humor is reliant on the flower name (forget-me-nots) and

squirrels-eat-nuts script, but is specifically engineered for the joke. The phrasal invention sets up a phonological structure that doesn't line up with the form that directly relates to either script.

Because of this compromise, the punchline lacks a certain phonemic elegance.

Additionally, some jokes will create new words to accommodate a merging consonant or vowel that ties the two scripts together:

Q: What kind of cars do cats drive?

A: Catillacs! (Kids Jokes - Animal Jokes, 2018)

In these types of jokes many setups and punchlines are contrived on a phonemic level and pragmatic level. In the example above, the two words "cat" and "Cadillac" coincide as a blend.

Although "cat" and "cad-" are a minimal pair in their own right, in order to produce a final punchline that is viable according to pun conventions, the joke teller must invent a new word.

Additionally, the premise of the joke is forced since cats do not drive cars. Not only is the phonemic wordplay forced, but so is the combination of scripts. It gives the impression that the joke was engineered with the simple goal of demonstrating the ability to arbitrarily tie together two phonemically similar words or word parts. Although this type of wordplay may not be as effective as perfect puns, its manifestations are still seen as puns. As mentioned before, near puns will be classified along with homonymous puns in this thesis.

### **2.2.5 Polysemous Puns**

Although the differences between polysemous words and homonymous words is commonly known within the linguistics community, it is surprising that there is not a clear distinction between these two phenomena in academic treatments of puns. Perhaps it seems pedantic to make this distinction, but the difference between these two phenomena can potentially make a difference in the overall effect of the joke (the reason for this being because of a humor factor embedded within the SSTH posited by Raskin (1984)). The premise of verbal

humor is that two scripts are paradoxically intertwined while still conflicting one with another. The schism created by lexically ambiguous words or situations allows for this mechanism to function with varying degree of comedic effectiveness. The important factor comes in the form of semantic distance. By nature, true homonyms are similar through accidental means, whereas polysemes do have a semantic connection. Thus, if jokes derive their humor from the opposition or tension of two scripts, wouldn't varying levels of semantic difference in the pun affect the overall opposition and execution of the joke? Indeed, it has been found that the degree of incongruity in the joke affects its humor, with a medium degree producing the greatest comical effect (Giora, 2002). Whether homonyms or polysemes achieve the "right" amount of incongruity given the scripts is a matter outside of the scope of this research. Now that we have established a motivation for this demarcation, we must explore what a polysemous pun is.

As noted, polysemy is a phenomenon in which two words, though identical in phonemic and orthographic form, differ somewhat in their meaning while still containing a remnant of connectivity through language roots and usage. Wordplays that use polysemy often play between the literal and figurative usage of a particular word:

Did you hear about the man that fell into the upholstery machine?  
He's fully recovered! (Jokes for Happy Harry, 2018)

The play on words in this example draws from the dual meaning of the word "recover". The literal meaning of the word in this setting has ties to the script involving "upholstery", while the figurative meaning has ties to the script of "accident and injury". Additionally, these words are polysemous for two reasons. Again the OED shows that the French root *couvrir* is present in both entries of "recover". Additionally, both definitions elicit the idea of gaining something back. In one sense of the word, "recover" could mean that a damaged piece of furniture gains back a material cover that may have been torn or stripped. In another sense it elicits the idea of



an injured person gaining back their physical form or health with time. Both senses of the word “recover” are linked through etymology as well as a general meaning of “gaining back something that was once lost”. With these connections these words are polysemous in their relationship. Polysemes can often vary in their literal and figurative meaning, but we must be careful to delineate them from other linguistic phenomenon. Polysemes can also be metaphorical, as we can see with the figurative usage of *pig*, which could serve as an insult. However, polysemes are not metaphors in their own right.

A metaphor creates connections from a concrete source domain to a more abstract target domain (Lakoff & Johnson, 1980). For example, calling politics a playground for spoiled children is an application of a simple concrete image being applied to a complex idea. A metaphor takes two seemingly different ideas and ties them together to clarify one of those ideas. However, Lakoff and Johnson make it clear that metaphors are not solely mappings from one term to another (1980). Metaphor construction requires a concrete idea being mapped to an abstract idea. Polysemes on the other hand are a system of meanings that connect to one etymological root. Polysemous puns must have two separate words that have different established meanings that are connected through historical means (Ravin & Leacock, 2000; Apresjan, 2009). If a word is mapped to another and used continually, this may extend the system of definitions that make up a polysemous word net. This word net is often recorded in etymological dictionaries. Conversely, metaphors do not have a formalized word net and are difficult to tack in their entirety. Thus, although metaphorical language may give way to polysemy, polysemy is not conceptually interchangeable with metaphor.

### 2.3 Perspectives on Puns

Puns are an international phenomenon (Guidi, 2012; Otake & Cutler, 2013) that are inserted into various social situations and contexts. Traditionally, puns have been at the center (for good or ill) of social joke telling. However, this phenomenon has found its way into advertising (Oaks, 1995; Beard, 2008), education (Vande Kopple, 1995; Lems, 2011), and even religious contexts (Bell, Crossley, & Hempelmann, 2011). Puns are ubiquitous in usage and recognition across generations and topics. Although puns are a universal linguistic tool, they are not universal in their usage. That is to say that not everyone appreciates them in all contexts. Just as there are nuanced rules to humor in general, appropriate and inappropriate usages of puns also exist. Additionally, some authorities debate whether puns should be used at all.

As an artistic form, puns have historically been a divisive topic. Although used throughout the works of renowned authors such as Dickens, Donne, Shelly, and Shakespeare, puns sometimes receive fierce derision from authors and grammarians alike. Puns are often seen as low-brow or ridiculous. Samuel Johnson believed that punning was detrimental to self expressions (Bates, 1999). As if there is a hierarchy of quality, puns are sometimes seen as a poor use of creative or linguistic skills, often lacking complexity or sophistication. Despite this, some researchers have found that wordplay enhances perceived comedy (Lippman, Sucharski, & Bennington, 2001). Although puns may be seen as low-brow, this is not to say that they are always easy in execution.

One important aspect to note about puns comes from the execution of the pun. The effectiveness of the punchline is dependent on the elegance of the merging of the two scripts. We could speculate that the reason puns may incur such contempt comes from the fact that many puns do little to effectively integrate opposing scripts except in fantastical or contrived

circumstances. Situations like animals talking to people or witches playing baseball are prevalent in punning wordplay. Indeed, words pairs that share no core meaning or origin would require greater fashioning on the part of the joke maker to establish a proper incongruity to resolve. Researchers have found that jokes that are given additional context to an overarching story found puns to be funnier than when they were present in shorter jokes (Sutton-Smith, 1976; Lippman, Sucharski, & Bennington, 2001). Puns that had more carefully constructed setups received higher marks for humor than puns with haphazard setups.

### **2.3.1 Pun Usage**

Contextual appropriateness affects our approach when dealing with humor. In other words, just as a presidential speech requires a certain register, so also does a social situation dictate the congruent verbal strategy for humor. In advertising, businesses will often use humor to attach a positive association with their product or service. However, the use of humor is not indiscriminate. Certain studies have shown that advertisers will use humor in their messages when promoting minimal risk or “fun” products such as snacks or drinks (Beard, 2008; Gulas & Weinberger, 2006), whereas products and services that are high risk or more expensive are not as well suited for humorous messaging. This leads back to the question of strategy based on domain knowledge and content. In what circumstances can puns be used? Are they limited by topic or tone?

As mentioned in section 2.1.1, humor can take on both positive and negative forms. The idea of positive and negative humor centers around the way humans cope with inconsistencies and injustices in the world. A person that employs positive humor uses the challenges of life as a platform to find common ground with others and build connections with them (Samson & Gross, 2012; Knight, 2013). Conversely, a person who employs negative humor looks at the same

difficulties and assigns blame to other people. Negative humor is often characterized by its aggressive and divisive nature. Where positive humor seeks to build connections, negative humor tears such connections down. As a part of understanding general perceptions of puns in verbal humor, we will explore the real-world contexts in which puns are used.

### ***2.3.1.1 Positive puns***

One positive way that puns are used occurs in a language teaching context. Although not a ubiquitous practice in the field, using puns to teach English has been a fruitful practice. For many ESL teachers seeking to help students understand the complexities of English, puns have become a way to encourage collaboration in the classroom as well as to improve verbal ability, particularly in trouble spots (Cekaite & Aronsson, 2005; Lems, 2011). Pun humor is used as a positive polarity humor that is inclusive to various social and cultural groups. The innocuous nature of puns allows teachers to inject humor into a potentially stressful situation. The release of humor can help students to relax and participate without fear of being judged. With greater participation, students learn more quickly and gain greater English proficiency over time.

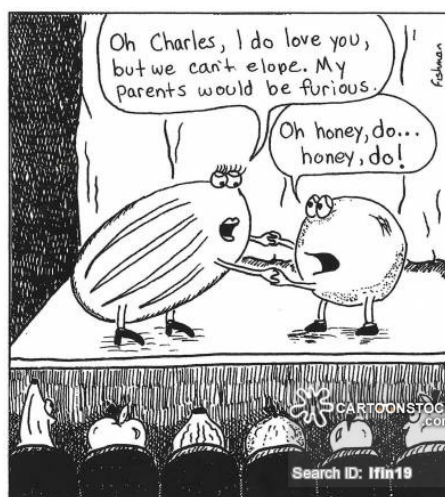
Puns can also inject positive humor into relatively precarious social situations. Historically, as mentioned in section 2.1.1, humor was seen as an opposing force to religious and moral betterment. However, as cultural notions about humor and as social perceptions have evolved, positive humor has been increasingly injected into religious contexts. In a study performed by Bell, Crossley and Hempelmann (2011) it was found that punning was prominent in the usage of church marquees. Where religious humor has historically been irreverent (Raskin, 1984), religious puns can inject humor into religious lessons and contexts without offense. Because of the humor generated through linguistic and semantic incongruity, puns need not be hostile or subversive in the goal of eliciting a humorous reaction. The puns found in the

marquees often used positive humor that avoided targeting any group as well as any divisive doctrinal stances:

A smile adds to your face value.

The positive yet vague stances on religious matters creates an inclusive message to lapsed Christians, while not alienating true believers.

Wordplay utilizes incongruity as its main humor mechanism, so there is flexibility in the way it is used. Because of this, jokes can be created from a variety of scripts, even absurd ones not based in reality (as demonstrated in the section on paronyms). Puns can easily be molded to fit different scripts or topics of interest suited for any audience. A good example of this using puns and near puns can be seen in the following comic<sup>1</sup>:



Melodrama

Figure 1 Food Pun Comic

Connecting the script of hammy theatrical performances to food words, the comic's humor is mostly driven by the wordplay. Additionally, the premise is ridiculous, and the wordplay does not negatively target any particular group.

---

<sup>1</sup>Retrieved from <https://www.cartoonstock.com/directory/w/wordplay.asp>

### 2.3.1.2 Negative Puns

Although often used for fun and games, punning also has a history of being used in negative polarity humor. Such examples can be found in jokes that depict violence, crime, and crude sexual acts (Lillo, 2006; Lillo, 2008). Often in explicitly negative scripts, taboo notions are converted into euphemistic slang in order to not be detected by linguistic/cultural outsiders or figures of authority. For example, the word “Bruce” can mean an ounce of drugs among dealers and addicts. Additionally, a “koala” can mean a man that is only interested in one night stands (Lillo, 2008). These seemingly innocent words act as fodder for mocking jokes that would be unrecognized by the uninitiated. In this sense innocent words are rebranded for taboo notions, yielding greater opportunity for negative uses of punning.

Aside from euphemistic puns being used for joking about disreputable notions, puns are also commonly used in open public insults and derision, especially within political debate or critique. Since America’s independence (and far earlier within the English language), politicians have been criticized through artful wordplay usage. Some usages are playful jabs at political gaffes, while others, such as the example in figure 2, are much darker and controversial<sup>2</sup>:



Figure 2 Trump Comic.

<sup>2</sup> Retrieved from <https://www.dailykos.com/stories/2017/1/18/1621634/-Cartoon-Trump-s-health-care-plan>

Utilizing a polysemous pun, this comic takes a definite and definitively negative stance. There is a clear target: the president and his proposed healthcare plan. The lexical ambiguity of “covered” alludes to both the meaning of healthcare covering all American citizens’ needs, as well as covering up a dead patient that the health plan presumably has failed. The wordplay crosses the script of political promises and assurances with a deep mistrust of the president’s intentions and/or credibility. Such is a clear example of divisive humor.

From this information we see wordplay strategies according to sentiment. Wordplay can be used for collaboration or denigration, but we still do not have an understanding of the prevalence of punning wordplay among various topics. Is there something specific about certain topics that would cause a joke teller to employ wordplay over a different type of humor strategy?

### ***2.3.1.3 Schema appropriateness***

It is important to remember that the fodder available to create a joke lies in its respective script or source material (Raskin, 1984). The larger or deeper the script, the greater the resources available. This entails a greater versatility to interact with other scripts or a greater variety of available humor strategies. One must also consider that certain scripts may have natural associations and feelings tied to them. For example, with political situations there may be a natural cynicism and mistrust that the general public have for politicians, despite good will political campaigns and promises.

Although personal style is an important aspect of humor delivery, the schema or social script plays a central role to what is funny and what is not. This thesis aims to determine correlations between humor strategies (specifically wordplay) and schemata. However, in order to do this, there must be a way to categorize and divide schemata. One way is to divide them according to psychological maturity. More specifically, jokes are not divided according to

“adult” or “kid-friendly” content but rather schemata that have natural appeal to one demographic or another. This is to say that some scripts are innate to certain stages in human life, and these scripts are only understood through experience. These categories are widely recognized and regarded as self-evident throughout society. The application of this idea will be discussed further in the methodology section.

## **2.4 Humor Research with Corpora**

From a descriptivist standpoint, linguists are concerned with gathering data that is true to real world language usage. Corpus research has become an increasingly more valued resource in achieving this goal. As a methodology it gives a relatively accurate look into how language is actually used in day-to-day contexts (Biber & Reppen, 2015). Within the social sciences there is always a concern that an experimental environment may influence the way humans act or communicate. Human subjects may censor themselves or modify behavior when they know that someone is analyzing them. Thus, experimental conditions may affect the external validity of a study or rather the extent that the results can be generalized to the human population as a whole (Taylor, 2013). When working with human subjects face to face, social science researchers will often create elaborate design models in order to hide the true purpose of study from the subjects. This is done with the hopes of lowering the chance that subjects modify their target behavior, thus exhibiting more natural behavior and increasing external validity.

Within corpus research, however, data collection is often performed away from experimental environments. Because of this, the language data remains closer to the natural way that humans communicate. This factor becomes increasingly more important in humor research as joke telling is influenced greatly by delivery and audience demographics. Additionally, corpus research helps to drill down on the complex factors of humor that have been debated over



centuries. Certain words or phrases can be searched with relative ease in a collection of texts (Biber & Reppen, 2015). These search results in conjunction with statistical calculations can reveal real empirical patterns across genre, dialect, ethnolect, and other dimensions.

Humor research has already produced some corpus applications to understanding joke structures. One study, using a 2000 joke corpus, found that lexically ambiguous wordplay jokes (puns) outnumbered all other types of verbal jokes (Attardo, Attardo, Baltes, & Petray, 1994). Topic and schemata were not emphasized or sorted in that study, however. It would be difficult to measure the prevalence of humor strategies across schemata, due to the inherently difficult-to-categorize pragmatics of humor. Additionally, there is little consensus as to the normalization of humor research. Joke structures and strategies are rarely mutually exclusive and thus difficult to parse. There has been some cutting-edge research conducted in finding computational models of verbal humor (Kao, Levy, & Goodman, 2015; Valitutti, Doucet, Toivanen, & Toivonen, 2016). But any objective research in humor requires extensive labeling and sorting of joke corpora. Such is the challenge and opportunity for corpus linguists within humor.

With greater online support (library creation and online Q & A forums) for linguistic programming, corpus research is becoming increasingly more accessible to all linguists. Packages such as nltk<sup>3</sup> allow for quicker sorting and processing of phrases and other linguistic features. Additionally, packages such as Seaborn<sup>4</sup> and Scipy<sup>5</sup> allow researchers to visualize data and determine significance across categorical and continuous data types. Together, these packages help researchers to conduct both quantitative and qualitative studies with relative ease.

---

<sup>3</sup> See <https://www.nltk.org/>

<sup>4</sup> See <https://seaborn.pydata.org/>

<sup>5</sup> <https://www.scipy.org/>

Such packages are examples of the tools used in processing and sorting the data from the sub corpora in this thesis.

### 3. Methodology

#### 3.1 Corpus Design

The corpus that was designed for this thesis is web-based, containing jokes from a variety of joke sites. Jokes were limited to four topics that highlight adult and child schemata: Marriage, Politics, Animals, and Food. This was done to see whether pun and narrative strategies could be connected to joke topic. In terms of restrictions though, topic was the only delimiter for this joke corpus.

The definition for what constitutes a joke is broad in scope for the purposes of this analysis. For this thesis a joke is a humorous text that is intentionally crafted to be told to an audience. These involve riddles, anecdotes, humorous observations, one-line witticisms, etc. The key idea is that a joke is an invention of the joke teller and not something that occurred through happenstance. Thus, real-life events where people accidentally become the central character of a humorous situation are not included. Most text samples pulled from joke websites were used, except in cases where samples did not meet the aforementioned criteria. Some websites receive their content from user submissions and thus require some vetting of non-joke material from the data. It must be noted that in order to pursue external validity, jokes were not censored or removed from the corpus due to immoral or explicit content. However, jokes were limited to one of four topics or schemata, as will be discussed later on. Utterances that consisted merely of insults or complaints were not included in the corpus.

However, jokes were cleaned according to grammar and spelling. Grammar cleaning mostly consisted of deleting repeated words or adding missing prepositions. The purpose of this cleaning process was to assist in the quantitative and qualitative analysis. If words are misspelled and syntactic structures are needlessly flawed, finding specific phrases is more difficult. Certain

features such as capitalization can be accounted for in terms of word variation, but misspelled words are lost from the aggregate when performing a frequency distribution function or a concordance search. However, jokes that were intentionally spelled using texting or internet speech were not modified, to preserve the original voice of the joke. Additionally, for the same reason, spelling variations were also preserved if the alternate spelling denoted an alternate dialect or accent, except in certain cases where alternate spellings were not deemed necessary. This issue and other additional cleaning and filtering of the corpus will be discussed in section 3.2.2.2.

## **3.2 Corpus Construction**

### **3.2.1 Tools**

In the compilation, sorting, cleaning (grammar and spelling), and labeling of this corpus, multiple technical tools were used. Microsoft Word and Excel were used to edit and store the corpus respectively. The labeling process of wordplay was manually performed, but some labeling of the narrative strategy (NS) was done with Python scripts.

In the analysis of this corpus the Python packages (pandas<sup>6</sup>, seaborn, numpy<sup>7</sup>, scipy) and the Natural Language Tool Kit (nltk) were used to extract and calculate the pertinent data. Pandas provided the Python interface to read the Microsoft Excel file corpus as a data frame, and collect the pun counts for each joke topic. These counts were then visualized into count plots (bar graphs) through the seaborn package. These observed counts were then processed through numpy to obtain the expected values in preparation for a chi-square independence test. Chi-square tests were then performed using scipy. The level of statistical significance for this test was

---

<sup>6</sup> <https://pandas.pydata.org/>

<sup>7</sup> <http://www.numpy.org/>

set at  $p=0.01$  for the wordplay distribution among topics. For supplementary illustrations of the content of the text, nltk was used to extract the top ten content words from the corpus. Using the concordance function, these content words were searched as a basis for the qualitative analysis.

### 3.2.2 Procedure

For various reasons discussed in this section, this joke corpus was sourced entirely from the Web. Jokes were selected from various joke websites using the Google search engine. The idea behind compiling a corpus based on internet jokes comes from the anonymity factor of the web. Because of this, joke contributors are less likely to censor themselves in both the form and content of their jokes. This allows the researcher greater access to joke forms in all their complexity (and often simplicity). Joke books are compiled with a sense of cohesion or even agenda in mind. These factors can be influenced by the bias of the compiler of the book or the publishing company. Conversely, internet joke sites are often democratized and receive their jokes from multiple contributors. This allows for a body of jokes with little cohesion and greater amounts of competing perspectives and agendas. Ultimately, this helps to create a corpus that is based on topics and not particular stances on those topics. Additionally, by selecting jokes from the top website search results, this ensures that the corpus contains jokes that are actually being viewed and appreciated by the world. In other words, not only are these jokes real in the sense that they are produced by a variety of joke tellers, but they are also evidently being appreciated by a global audience.

Another simple reason that this corpus is internet-based is ease of compilation. Although requiring a great deal of data cleaning (grammar, spelling etc.), copying and pasting internet jokes is a much quicker process than scanning or digitizing print text, especially for a single researcher.

### *3.2.2.1 Child Versus Adult Schemata*

To decide which joke topics would be included in the joke corpus, several factors were considered. To determine which schemata would be classified as adult, the experiences would naturally follow societal and legal norms. For instance, the two topics of Politics and Marriage were deemed adult because these topics are more relevant to the interests of mature human beings and because both of these experiences have legal age restrictions that preclude children from fully participating in such activities. Although it is reasonable to say that precocious children may have interests in these topics, they cannot have a primary knowledge of the nuances contained in these topics.

To classify schemata for children, there is a problem with the aforementioned methodology. To better understand what a child's schemata consist of, further reasoning and research were conducted. Adults were once children, and thus they have had similar experiences to the generations that follow them. Additionally, fewer legal restrictions exist on adults than on children. These factors required an assessment of joke relevance for children. In other words, there needed to be a method of identifying joke topics that were relevant to children. To carry this out, a survey of ten children's books (see appendix) was taken to determine common topics for jokes. If a certain topic was common to all the books, that topic would be admissible as a child schema for the corpus.

For this survey, each book contained in its title some form of the words "kid" and "joke". None of the titles would contain the words "pun" or "riddle". The assumption in selecting books with these conditions was that all of these joke collections were assembled for children but without deliberately targeting wordplays or puns for selection in these collections. The idea was to find joke topics that are ubiquitous to children and that do not have an explicit association

with wordplay. Additionally, none of the joke books contained the phrase “Knock-Knock”. Often times these types of jokes also lend themselves to wordplays.

Upon sifting the joke books, certain themes emerged. Topics such as dealing with authority figures like school teachers, parents, and doctors could be found in a few of the books. Other interests such as sports and fantasy characters were also prevalent. However, two topics ubiquitous among all the joke books were animals and food. Although these topics are not inherently tied to childhood experience, they are seen as tied to basic human experience. Another interesting pattern with these topics comes in the contrast with the adult topics. With the schemata of marriage and politics, communication is crucial for success and daily function. Conversely discussion and communication are peripheral to interacting with food and animals. Although children may talk to their pets during play time or talk with their family during meal time, such communications are not fundamental to these activities. Linguistically and psychologically, the schemata for marriage and politics are more complex than for animals and food. This brings back the question of whether the linguistic complexity of schemata influences the likelihood of wordplay usage.

### ***3.2.2.2 Extraction and Cleaning Process***

For the joke categories of Marriage, Animals, Food, and Politics, the exact google searches were “Marriage jokes”, “Animal jokes”, “Food jokes”, and “Political jokes” respectively. Jokes were then selected from the top websites by their respective topic and search input until each category contained jokes reaching 25,000 words (totaling 100,000 words for all four topic categories). As a side note, it must be stated that these website results are determined by Search Engine Optimization factors that are dictated by Google’s algorithms. This is relevant to the discussion because these algorithms are regularly refined, and thus, the webpage results for

the same key word searches will shift over time. Consequently, anyone seeking to replicate this corpus compilation using the given string searches above will most likely pull their jokes from different sites than ones used in this thesis. However, those seeking to know the exact sites used for this corpus can find them in the appendix section of this thesis.

Another factor that affects the consistency of this procedure is the creation and popularity of new joke sites that emerge over time. To help with the standardization of this process, all top website results were recorded on the day of first access and recorded in the appendix section. This ensured that all joke extraction was consistent across multiple days. Joke extraction was performed in the order of the appearance of jokes on each website until either all jokes from the site were exhausted or until the corpus was filled.

It may be said that pulling jokes from a site until exhaustion may not be the best sampling process. Some joke websites may have 50-100 jokes in total while other sites have thousands of jokes. This means that with the chosen methodology some sites will be disproportionately represented over others. Indeed, it may have been better to extract a set amount of jokes from each website to have a more even distribution. However, the method chosen for this thesis was done so for pragmatic reasons. A uniform extraction process would be lengthy to perform manually. That would require the aid of a web scraper that could automatically search and extract jokes according to certain given parameters. However, this researcher did not have access to such a web scraper and did not have sufficient time to learn web scraping through Python. Thus, the extraction process was performed manually.

Where duplicate jokes were found in multiple parts of the corpora, the first appearance of the joke was kept, while the remaining copies were deleted to maintain joke diversity. Because many jokes consist of two or more scripts, some jokes could be categorized under multiple topics



and corpora. Additionally, it is difficult to clearly discern if a joke definitively encompasses one topic more than another. If a joke deals with both animals and politicians, how would one determine the dominant script for categorization purposes? To simplify this organization and filtering process, corpora were created and filtered in the following order with the first corpora receiving priority in keeping duplicates: 1. Marriage 2. Animal 3. Food 4. Politics. The ordering of this filtering process is arbitrary, but the standardized hierarchy allowed an unbiased method to delete duplicates. Otherwise, for example, if a joke containing a wordplay could be found within two different corpora, it may be easier to leave that wordplay in a corpus that naturally contains more puns, thus doctoring the final statistical significance for each corpus.

It should also be noted that within each topic, variation versions of a single joke exist within and across websites. Thus, jokes that were deemed “too similar” were considered duplicates. Factors that were considered were the scripts, setup, and punchline of each joke. Two jokes that contained the same scripts, setup, and punchline, even when differing in character or location names, were deemed as duplicates and thus filtered out. However, if one of the three aspects of the joke differed between two jokes, then both jokes would be kept. To illustrate this process here is an example where one of the two jokes would be filtered:

Q: Where did the artichoke go to have a few drinks?

A: The Salad Bar! (Artichoke Jokes, 2017)

Q: Where did the bell pepper go to have a few drinks?

A: The Salad Bar! (Bell Pepper Jokes, 2017)

These jokes are essentially the same joke, differing only in the type of vegetable that was used in the setup. Conversely, here is example of two jokes that would be deemed different enough to both remain in the corpus:

Q: What do bees do if they want to use public transport?

A: Wait at a buzz stop! (Animal Jokes, 2017)

Q: What's a bee-line?

A: The shortest distance between two buzz-stops! (Animal Jokes, 2017)

These two jokes share the same script and pun. However, the setups of each joke are different. Although both jokes get to the same punchline, each joke takes a slightly different route to get there.

Admittedly, this filtering process favors pun preservation in the first few corpora, adding a potential skew to the overall data. However, the purpose of this method was to gather unique data. Many duplicates were found across corpora, especially between the food and animal corpus. Because of this the delineation between animal and food jokes became unclear (as many animals can be processed for eating). Thus, the duplicate cleaning process was employed to strengthen the identity of each corpus and provide unique content.

Microsoft Word was also used to track the word count for each corpus. Once they were cleaned according to grammar and spelling, the jokes were divided and separated into tables which were then inserted into Microsoft Excel. From here the data frame for analysis and labeling was constructed. The filtering process for duplicates was performed at this stage with search and replace as well as manual scanning.

A minor note must be made concerning the cleaning process. Several British spellings of words were found in the jokes such as “colour” or “honour”. Evidently, either such jokes were sourced from one or more Commonwealth countries, or the website administrator comes from one of these countries. However, no wordplays or other type of joke used prominent slang from one of these countries as a humor device. Because the purpose of this thesis was not to source jokes from American culture specifically, the alternate spellings or word usage were not an issue in the overall analysis. Although the phraseology seems to use standard American English, and the topics insofar as can be determined center around elements that can be found within

American culture, there is no certainty that the corpus represents American humor. Rather, it can simply be said that the corpus represents English language based humor.

### ***3.2.2.3 Pun Categorization***

To delineate the jokes in each corpus according to wordplay, jokes were separated into three categories: Homonymy, Polysemy and Non-pun. Wordplay is a broad term linguistically, and all types of wordplay other than homonymous and polysemous puns such as anagrams, spoonerisms, simple rhymes, etc. are often difficult to objectively separate. These were not categorized further in this thesis and were placed with other non-punning jokes in the non-pun category. The categories are meant to be mutually exclusive, but at times there are overlaps with other types of jokes. To simplify the categorization, there was a hierarchy of priority established in the labeling process. If a joke contained homonymy as well as rhyming, the joke was sorted to the “Homonymy” group. Likewise, if a joke contained homonymy as well as polysemy, the joke was also labeled “Homonymy”. Because the conventional definition for puns lies with homonyms, these jokes were given priority in the labeling process. Here is an example of a joke containing both polysemy and homonymy:

A panda walks into a bar, sits down, and orders a sandwich. He eats, pulls out a gun, and shoots the waiter dead. As the panda stands up to go, the bartender shouts, "Hey! Where are you going? You just shot my waiter and you didn't pay for the food!" The panda yells back, "Hey man, I'm a panda. Look it up!" The bartender opens his dictionary to panda, "A tree climbing mammal of Asian origin, characterized by distinct black and white coloring. Eats, shoots, and leaves." (Joke #4925, 2017)

The punchline of this joke has two puns: “shoots” and “leaves”. The two counterparts of *shoots* as in the verb *firing a bullet* and the noun *stem of a plant* are polysemous. Both of these counterparts come from the verb *to shoot*. On the other hand, the second pun *leaves* has two counterparts that are homonymous. “Leaves” can refer to either the plural noun

form meaning *foliage* or the third person singular form of the verb form meaning *to depart*. This joke contains differing pun categories, but because it contains a homonymous pun, it receives the label of “Homonymy” according to the decided hierarchy.

To effectively sort these different jokes into their respective categories, several resources were used. In seeking to discern a wordplay, some examples could be approached intuitively, but others could not. Here is one such example where the wordplay was not intuitive to the researcher:

Q: What's Clinton's Economic forecast?  
A. A "Bare" Market (Political Jokes, 2017)

This joke relies on two scripts: 1) the Clinton scandal with Monica Lewinsky and 2) the knowledge that a “bear market” refers to a downturn in the U.S. economy. The researcher did not originally understand the reference to the second script. Thus, this required the researcher to investigate terminology having to do with economic markets to fully understand the joke. After doing this, the homonymous pun involving the counterparts *bear* and *bare* became apparent, allowing the researcher to correctly label the joke.

In order for a wordplay to be perceived correctly, the context of the joke must give the sufficient clues needed to yield two meanings in a script and punchline. Indeed, the most effective jokes employ a medium amount of incongruity and information that is optimal for wittiness as well as clarity (Giora, 2002). This means that effective jokes require some degree of the audience’s effort in resolving the incongruity while still leaving just enough clues to allow for an ambiguity. Some jokes are brief and contain few priming clues (such as the example given in the previous paragraph). These jokes rely more on the audience’s domain knowledge than the

overt clues in the setup itself. In such cases, the wordplay in jokes was not always readily available to the researcher as the punchline required a script reference, such as a specific quote or policy put forth by a particular politician. To correctly discern these wordplays as well as what type of pun was used, general Wikipedia and news site searches were conducted to understand the two separate meanings in a punchline.

As previously mentioned, homonymous puns are deemed to be words that phonemically allude to another word (either partially or exactly) in the given script overlap, while not carrying a related meaning. Although polysemy and homonymy are clearly defined in their differences, being able to discern one over the other in a given pair of words is not always intuitive. For example, the word *bank* has multiple meanings that come to mind. When comparing the referents “financial establishment” with a “mound or slope” (as in a river bank), these words seem to lack even a tenuous semantic connection, despite the fact that they share the same orthography and phonemic structure. Upon closer inspection in the *Oxford English Dictionary*, however, it is discovered that both words are connected through the Middle French *banc*, meaning “bench”. The OED gives further elucidation as this word refers to both the shape of the dirt mound for one referent and the “market stalls” or “sales counter” for the other. In this situation the phonemic connection between the two referents of *bank* are not serendipitous, but rather rooted in historical usage.

Before delving into the technical methods used for separating homonymy and polysemy, a potential concern about the methodology must be addressed. It must be said that perhaps the technical separation of homonymy and polysemy may not be relevant for all audiences. As shown in the example above, jokes that employ polysemy may not have a recognizable semantic connection between pun counterparts. Some may ask, if a semantic connection is not readily

apparent (regardless of technical classification), why make this distinction? If humor is based on the degree of semantic opposition between scripts, why make such a technical delineation when the semantic distance is not readily apparent to an audience? To this it can simply be said that the proposed method provides a pragmatic means of providing consistency in the classification and labeling process of this corpus. Determining the etymological roots of a pun in order to separate polysemy and homonymy is a diachronic approach and a relatively objective method to track semantic connections between roots. Alternatively, a survey method or synchronic approach that surveys people about their awareness of current semantic connections between words could determine semantic connectivity between two words, but there are a host of other issues involved with such a method, which will be discussed later in section 5.2.

The diachronic approach was used over a synchronic approach because it provided the simplest empirical means for drawing a distinct line between polysemy and homonymy. If this sorting were done synchronically, it would require a combination of surveying people and making guesses about the degree to which people are aware of a semantic relationship between two forms. The survey of various word pairs would have been impractical and perhaps even inconclusive, forcing the researcher to make an arbitrary judgement anyway. Indeed, it is because of this complication that diachronic approaches were often used to discover polysemy (Ravin & Leacock, 2000). To practically and empirically discern homonymous over polysemous jokes, several resources were used to find the etymological roots of the words or morphemes in question. Resources such as the *Oxford English Dictionary* (OED), the *American Heritage Dictionary*, and [www.etymonline.com](http://www.etymonline.com) were used to identify and compare roots.

In a few cases the wordplays in question were based on proper names. For these situations [www.ancestry.com](http://www.ancestry.com) was used to find the etymology of surnames. Although the

namesake of a person or a company is difficult to determine, the assumption made was that surnames have the same root as their common noun counterpart. For example, it cannot be assumed that two people with the last name *Smith* are related, but it is reasonable to say that the origin of the surname is consistent with certain associations (blacksmith, silversmith etc.). These roots were then compared to the OED or other aforementioned references to establish consistency.

In order to separate homonymous from polysemous puns, a wordplay was divided into its two phonemically similar parts. Each word was then searched in the OED, and the respective etymological roots of each word were compared. In order for word pairs to be deemed polysemous, both entries had to contain the exact root from the same language or language family. In some cases, the OED would make express connections between two etymological entries, including a link from one word to its counterpart. In this comparison process, the researcher did not consider similar word stems from sister languages to be polysemous unless one of the resources used identified them as cognates. Although there were some instances that an argument could be made for possible cognate matches, such arguments would require a much deeper study of the English language and may involve conjecture. Such allowances would also compromise the standardized method for labeling, making the methodology significantly more difficult to replicate. In situations where the OED lacked etymological data, the *American Heritage Dictionary* (<https://ahdictionary.com/>) was used as a supplementary resource.

For wordplays that involved proper names such as last names of presidents or names of companies, the search on [www.ancestry.com](http://www.ancestry.com) helped clarify their semantic connection to common nouns or other types of words. For example, one joke in the corpus involves a fictitious interchange between Moses of the Old Testament and George W. Bush. The wordplay of the

joke hinges on the scripts of Moses wandering in the desert for 40 years as well as America's discontent or lack of confidence in the American president. Because of this, Moses is reluctant to listen to the counsels of a *Bush*. In searching *www.ancestry.com* for the etymology of the last name *Bush*, the following entry is found:

English: topographic name for someone who lived by a bushy area or thicket, from Middle English bush(e) 'bush' (probably from Old Norse buskr, or an unrecorded Old English busc); alternatively, it may derive from Old Norse Buski used as a personal name. Americanized spelling of German Busch. (Bush Family History, 2018)

It is evident that the proper name and the word for shrubbery are directly connected, thus signaling a clear polysemy. In situations of a wordplay hinging on a company name, the same steps were applied. If the origin of any proper name could not be secured, the wordplay between the two counterparts was deemed homonymous.

Once homonymous puns were labeled, jokes were sorted for possible polysemy. Jokes that were deemed as polysemous hinged on a single word that had at least two meanings that made sense within the priming of the joke and that were related by etymological root or had alternate meanings based within the same dictionary entry (the OED is organized in such a way that each entry contains multiple definitions, but all definitions in that entry have the same etymological root). In addition to this factor, other criteria factored into the labeling process.

Certain nuances exist in the way this corpus was labeled. Jokes that used compound words were not labeled as puns. The term "compound word" is used along the lines of the following definition found in the *Cambridge Dictionary Online*:

A compound word is two or more words linked together to produce a word with a new meaning (2018)



Compound words were not seen as homonymous or polysemous for this thesis since the ambiguity of meaning is not self-contained within one word. A pun is a single word that has two counterparts, and thus a compound does not fit this description. For example, the compound verb *to put down* has alternate meanings (literal and figurative) that could produce a wordplay. It could mean to insult, to physically place onto the ground, or to euthanize. Each of these specific meanings exists solely within the compound verb. Neither *put* nor *down* can individually convey the meanings given above. And since we have established that puns are single word wordplays for this thesis, compounds will not be labeled as puns. Additionally, jokes that employed longer idiomatic expressions that could also be interpreted literally (e.g. a person using the phrase “that’s the way the cookie crumbles” while eating a cookie) were not categorized as polysemous since the ambiguity or alternate meaning is distributed across the entire idiom and not solely on one word.

Although jokes were not deemed puns if the meaning of the punchline spanned more than one word, there was one caveat made for this thesis. Jokes that had a single word counterpart in one sense, but a multiword counterpart in another were labeled as puns:

Q: What did the mama cow say to the baby cow?  
A: "It's pasture bedtime." (Animal Jokes, 2017)

Because of phonemic similarity and contextual clues, the single word “pasture” can also be interpreted as “past your”. Jokes like the example above were labeled as puns because the orthographic realization itself is a single word by one of the interpretations. Thus, this technically falls under our established definition of a pun. Additionally, other similar examples like this could fall under *near puns* and in turn would be labeled as homonymy in the corpus.

### *3.2.2.4 Narrative Strategy Categorization*

As a means of understanding the associations of puns, it was necessary to explore the context of the puns in greater detail. To do this, the narrative strategies (NS) of each wordplay were also recorded. However, because these narrative strategies are not mutually exclusive, certain judgements were required as to what constitutes one category versus another. For instance, when does a dialogue become a narrative? Is there a certain word count on the sequence of events that constitutes one or the other?

As two colleagues of Raskin point out, little research has targeted the actual language of each type of NS as no one method exists to “enumerate and define” the complexities of each strategy (Hempelmann & Ruch, 2005). One way to categorize jokes within this thesis according to NS could have involved a panel method in which subjects would intuitively categorize each joke’s category. With this, an average from the responses could serve to give a final label for each joke. However, for a corpus of 3500 jokes, this would be a lengthy process for any panel of people to execute. It would also be difficult for an individual to label all 3500 jokes. More likely, it would require the researcher to break down and distribute smaller chunks of the corpus. Even if each participant were given 100 jokes to label, this could be a substantial task. Each one of these sections would require multiple raters of 5-10 people. If executed this way, the number of participants (simply for NS labeling) could surge into the hundreds. Rather than conduct such a large study, it makes sense pragmatically to simply pick two narrative strategies and explicate their respective characteristics in reliable detail. For this the forms of riddles and narratives will be discussed.

The form of riddles in joke contexts have been discussed in some detail among linguists (Pepicello & Green, 1984; Dienhart, 1998). Riddles are often characterized by a juxtaposition of

two or more elements combined with a question to the audience. The punchline often delivers the resolution of the seemingly incongruent components. Generally, the setup for a riddle joke would include phrasing like “What’s the difference between \_\_\_\_ and \_\_\_\_?” or “What do you call a \_\_\_\_\_ that plays \_\_\_\_\_?”. Such a form can be seen in the following example:

Q. What's the difference between a bagpipe and an onion?  
A. No one cries when you chop up a bagpipe. (Bagpipe Jokes, 2018)

Riddle jokes are relatively short, and thus it is easy to remember all aspects of the joke. As a result, riddles are stereotypically seen as the “canned joke” that is recycled and shared among friends and family. In addition to this, riddle jokes in Modern English have historically been seen as childish in their form (Espy, 1971; Augarde, 1986). Such jokes are seen as low-brow or unsophisticated in the realm of humor. It will be interesting to see whether riddle jokes predominantly correlate with puns and/or children’s scripts/topics.

Some joke forms looked like riddle jokes, but were not included in the riddle category. Jokes that took on the structure of “Did you hear the one about \_\_\_\_?” or “Have you heard about \_\_\_\_?” were not labeled as riddles. Although each form uses a question for the NS, the question does not include an interrogative that is asking the audience to divine an answer, given the clues. Rather, the question asks the audience what their past experience has been with a similar joke type. In this latter type of inquiry, the question is largely rhetorical, and this subtle pragmatic difference between these and the aforementioned forms should be noted.

Another prominent NS found in jokes involves a format in which a joke has components of a narrative, but no story is fully expressed. See the following example:

Two peanuts were walking down the road. One was assaulted (Peanut Jokes, 2018)

This joke creates a setting with characters, but there is almost no succession of events, building of events, or reaction given to the incident. The setting is merely placed to set up a single wordplay. Thus, in this sense, this joke is more of a one-liner. In seeing whether jokes were dialogues, the single determining factor was whether there were at least two characters in the joke and both characters say at least one utterance to each other.

For this thesis each joke was also labeled according to Narrative Structures discussed above. Specifically, jokes were given the label of *Riddle*, *Narrative/Dialogue*, or *Other*. The third category of *Other* contained an assortment of various types of strategies including, but not limited to, observational humor, one-liners, knock-knock jokes, pick up jokes (e.g. “Hey girl are you tired? Because you have been running around in my head all day.”), yo mamma jokes, and so forth. This extra categorization serves as a possible means of understanding the social perceptions of puns in general.

## 4. Results

After compiling and labeling the corpus, the researcher analyzed jokes both quantitatively and qualitatively according to each topic. Basic counts for jokes, words, pun strategies, and narrative strategies were taken. After taking these measurements, chi-square tests were conducted in relation to narrative and pun strategies in order to determine the significance of their respective distributions across joke topics. Subsequently, concordance searches were performed of the top content word tokens for each sub corpus in order to understand additional characteristics of the joke schemata.

### 4.1 Quantitative

Each sub corpus comprised as close to 25,000 words as possible without going under. Thus there is a slight discrepancy between the word count of each corpus due to joke length.

Table 2 shows the statistical counts of the sub corpora:

Table 2 Sub Corpora Totals

Corpus	Total Word Count	Total Joke Count	Average Word Count per Joke
Animal	25001	916	28.07
Food	25006	1260	20.42
Marriage	25034	424	60.82
Politics	25020	918	28.05
	100,061	3518	34.34

One detail that is evident from Table 2 is the variance in the average word count per joke per sub corpus. The average word count in the marriage corpus is 2-3 times larger than any of the other corpora. The reason for this will be explored later in section 4.1.2, but this variance must be noted as it significantly reduces the number of marriage jokes in proportion to the other topics. Although statistical analysis can accommodate for this discrepancy in terms of distribution significance, fewer marriage jokes could potentially influence the degree of

confidence for discerning patterns in joke strategies. Admittedly, designing a corpus based on word count may have not been the optimal method in designing this corpus, but there are still noteworthy patterns we can discern from each sub corpus.

In the classification of each category, significant patterns were found both in joke structure and humor strategy. To calculate these results after the labeling process, Python scripts, as described earlier, computed counts and calculated statistical significance. For the sorting of a joke in the corpus, each feature received a binary label of 1 or 0 to signify whether the joke contained a particular feature or not. For example, if a joke contained a homonymous pun, then the joke received a 1 for the homonymy feature. Conversely, if the joke used any other pun strategy, the homonymy feature would receive a 0.

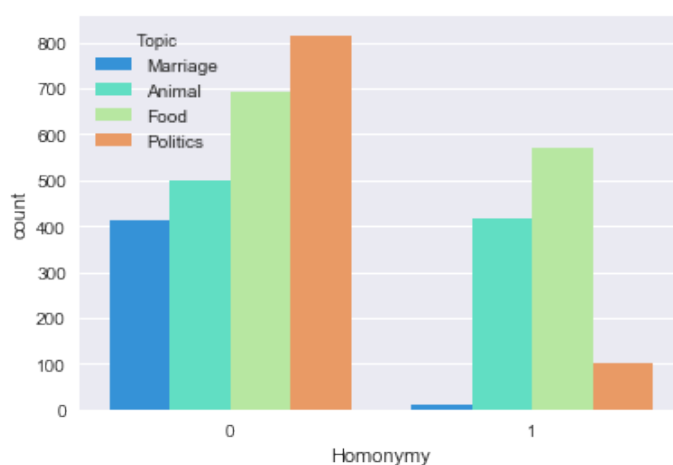


Figure 3 Homonymous Puns across Topics

Figure 3 shows that the use of homonymous puns is significantly weighted towards non-adult scripts both within and across categories. Figure 4 shows a similar result for polysemous puns, though not to the same extent:

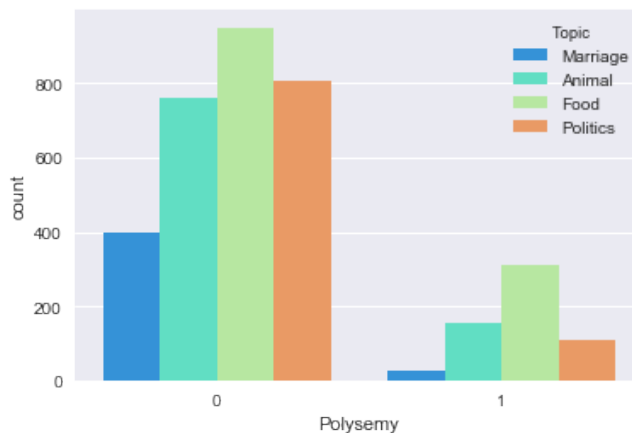


Figure 4 Polysemous Puns across Topics

These two figures show that puns in general are used less among adult schemata jokes. However, in order to determine the statistical significance of distribution for the entire corpus, we must turn to chi-square testing.

#### 4.1.1 Chi-square Testing

To determine correlations among strategies and scripts, three chi-square tests were performed: 1) comparing pun strategies to topics, 2) comparing narrative strategies to pun strategies, and 3) comparing narrative strategies to topics. The Pearson's chi-square test for independence ( $\chi^2 = \sum (O - E)^2 / E$ ) (where "O" means *observed* values and "E" means *expected* values) is used on categorical data to determine how likely a certain distribution is due to chance. If a chi-square test yields a statistically significant result, then it can be said that the distribution is not due to chance. By then measuring the observed values against the expected distribution for the data, relationships between categories can be ascertained.

The equation  $E = (t_r \times t_c) / N$  computes the expected values for each cell in a contingency table (matrix), where  $t_r$  = the total value for the given row,  $t_c$  = the total value for the given column, and  $N$  = the grand total for the table. These expected values provide the baseline for the

observed values to clarify how far from the mathematical norm the observed values diverge. The observed and expected values are then used to compute the chi-square value  $\chi^2 = \sum((O-E)^2/E)$ . Once the chi-square value is calculated, we need to find the *degrees of freedom* of the contingency table. The value for the degrees of freedom and the *p score* will be measured against a *critical value table* to determine the significance threshold for the chi-square value. For this thesis a p value of 0.01 will be assigned to determine significance for the following chi-square tests.

#### 4.1.1.1 Pun Strategies and Joke Topics

Table 3 shows the values for the pun strategies as they are distributed across joke topics. The observed values are in black while the calculated expected values are parenthetically rendered in blue:

Table 3 Observed Versus Expected Pun/Topic Totals

	Homonymy	Polysemy	Non-pun	Topic Totals
Animal	418 (286.15)	156 (157.01)	342 (472.84)	916
Food	568 (393.62)	310 (215.97)	382 (650.42)	1260
Marriage	10 (132.45)	27 (72.68)	387 (218.87)	424
Politics	103 (286.78)	110 (157.35)	705 (473.87)	918
Pun Totals	1099	603	1816	3518

$$\chi^2(DF = 6, N = 3518) = 841.75, p < 0.01$$

The result is highly significant, yielding a strong connection between joke topics and wordplay strategies. Table 4 shows the proportion of pun usage according to joke topic:

Table 4 Pun Prevalence (percentage) among Topics

Topic	Pun	Non-Pun
Animal	62.62%	37.37%
Food	69.59%	30.41%
Marriage	8.73%	91.27%
Politics	23.04%	76.97%



The majority of the jokes in the two adult schema corpora do not contain much punning wordplay, with marriage jokes containing the fewest examples of puns. Additionally, there is one more thing to note about the pun strategies of the marriage corpus. In Table 3 it is shown that of the 37 usages of puns in this sub corpus, 27 are due to polysemy. Although this is not enough data to draw definitive conclusions, an almost 3 to 1 ratio of polysemous puns to homonymous ones is notable.

#### *4.1.1.2 Narrative Strategies and Pun Strategies*

Table 5 shows the NS categorical breakdown (as delineated above) of all jokes and their respective pun strategies according to observed and expected values.

Table 5 Observed versus Expected Pun/NS Totals

	Homonymy	Polysemy	Non-pun	NS totals
Narrative/Dialogue	50 (208.05)	62 (114.16)	554 (343.79)	666
Riddle	902 (632.60)	444 (347.09)	679 (1045.31)	2025
Other	147 (258.35)	97 (141.75)	583 (426.90)	827
Pun Totals	1099	603	1816	3518

$$\chi^2(\text{DF} = 4, N = 3518) = 661.78, p < 0.01$$

In comparing the observed values versus the expected values, we notice that the great majority of puns coincided with riddle forms, with 1346 riddle jokes containing puns. Of all the narrative strategies, the Narrative/Dialogue form showed the fewest puns overall. Additionally, this form is also unique in that it is the only form that contains more polysemous puns than homonymous puns.

Overall, it appears that at least punning wordplay seems to gravitate towards riddle structures. Perhaps this is one reason why puns are often seen as childish in their form. For a clearer depiction of the results, see the following condensed breakdown of pun prevalence according to narrative strategy among all jokes:

Table 6 Pun Proportions (percentage) per NS

	Riddles	Narrative/ Dialogue	Other
Pun Joke Counts	1346	112	244
% of Total Puns	79.08%	6.58%	14.34%

#### 4.1.1.3 Narrative Strategies and Joke Topics

It was previously mentioned that certain narrative strategies seem to gravitate towards children or adult jokes. However, for greater triangulation it was deemed necessary to run a comparative empirical analysis of narrative strategies and topics. Table 7 shows the observed and expected values for this comparison:

Table 7 Observed Versus Expected NS/Topic Totals

	Narrative	Riddle	Other	Topic Totals
Animal	177 (173.41)	671 (527.26)	68 (215.33)	916
Food	104 (238.53)	935 (725.27)	221 (296.20)	1260
Marriage	312 (80.27)	15 (244.06)	97 (99.67)	424
Politics	73 (173.79)	404 (528.41)	441 (215.80)	918
NS Totals	666	2025	827	3518

$$\chi^2(\text{DF} = 6, N = 3518) = 1502.49, p < 0.01$$

We can say with confidence that narrative strategy is not independent of joke topic.

By comparing the observed and expected values, we notice that riddles indeed correlate with the non-adult topics as has been stated before. On a more specific level, the marriage corpus shows a correlation with narrative structured jokes. Perhaps this is indicative of the fact that many marriage jokes involve interchanges between husband and wife. One unexpected pattern arises in the political corpus. The political jokes displayed fewer narrative strategies than expected. Additionally, the political corpus displayed a greater proportion of *other* NS types. Because this type was not delineated in detail, these results will be investigated further in the qualitative analysis.

### 4.1.2 Content Word Frequency

In addition to wordplay labeling, a simple frequency distribution was conducted to find the top ten content words within each corpus. The intent of this was to bring greater empirical insight to the content of the jokes themselves. These word counts are meant to be a primer for the additional qualitative analysis of the jokes that will be performed later. This analysis was performed through some Python scripts that utilized the nltk package. Using the list of stop words (function words) provided by nltk, functional words such as *the* and *of* were filtered from the corpus, and a basic frequency distribution was applied to the remaining content words:

Table 8 Most Common Words per Sub Corpus

<b>Marriage</b>	<b>Word Counts</b>	<b>Politics</b>	<b>Word Counts</b>
'wife'	266	'q'	213
'husband'	238	'president'	153
'said'	174	'trump'	138
'man'	129	'cheney'	124
'one'	112	'said'	101
'woman'	92	'clinton'	100
'says'	92	'donald'	100
'asked'	87	'one'	95
'married'	84	'obama'	93
'like'	69	'dick'	82
<b>Food</b>	<b>Word Counts</b>	<b>Animal</b>	<b>Word Counts</b>
'q'	596	'q'	564
'call'	153	'call'	128
'get'	142	'dog'	123
'man'	116	'says'	115
'like'	105	'one'	113
'say'	91	'get'	110
'eat'	84	'man'	101
'one'	78	'bear'	93
'said'	74	'bee'	85
'says'	70	'bird'	64

Table 8 shows one particular token that stands out. The 'q' token stands as the highest frequency token for all corpora except the marriage corpus. Upon performing a concordance search of the

data, we can see that the ‘q’ is revealed to be the beginning of a riddle joke—as in “Q: What do you get when you...”. Additionally, the high frequency of the iterations of the tokens ‘say’, ‘get’ and ‘call’ in the animal and food corpora are supportive of the riddle joke formats like “Q: What do you call a...”, or “Q: What did the \_\_\_\_ say to the \_\_\_\_?”

The Marriage sub corpus contains words such as “say” and “ask” in addition to the words “husband”, “wife”, “man”, and “woman”. Together with the average word count of each joke in this section (60.82), these characteristics are supportive of the idea that the marriage jokes predominantly use a story format. After performing concordance searches on these top words, we find that many of the marriage jokes come in the form of conversations between married couples. Perhaps the reason why most jokes about marriage are presented as conversations comes from the frustration and humor found in marital misunderstandings. Conversations often unveil the way of thinking of each participant, and thus provide a prime medium to highlight the (comical) differences between men and women.

Political schemata show a high prevalence of names of political figures. Rather than focusing on policies and philosophies, the people are emphasized in the jokes. Searching concordance lines explains why this is the case. Many of the political figure references were connected to mocking type jokes. Presidents and government employees are targeted, often by late night comics, for their lack of intelligence or for immorality. These findings are only some of the joke characteristics found in the corpus. For a greater understanding of the essence of each sub corpus, I performed a follow up qualitative analysis.

## 4.2 Qualitative

Upon examining the concordance lines of each quantitatively significant pattern found above, we notice additional characteristics of the corpora according to topic. An examination of

the *Other* NS type jokes within the political corpus reveals a high prevalence of jokes that come from late night talk show hosts. An example of this can be seen below, concerning the Dick Cheney hunting incident:

It turns out now that Dick Cheney did not have a license to hunt, and coincidentally, turns out we didn't have a license to go into Iraq. -David Letterman (Political Jokes, 2017)

Whereas each corpus seems to have more canned jokes, the political corpus has multiple instances where these hosts make humorous observations about political figures. This coincides with the high frequency of political names found in the distribution frequency counts above.

Additionally, when looking at the content of jokes, there are interesting patterns that occur within the scripts. The humor used in the non-adult scripts almost wholly derived its comedic strategy from the wordplay itself. With the adult topics the wordplay presented was often more strategic than the non-adult topics. There was a greater sense that the wordplay was used to complement the natural inconsistencies found in the scripts. This contrast can be found within the following examples:

Adult script (Marriage):

Soon after we were married, my husband, Paul, stopped wearing his wedding band. "Why don't you ever wear your ring?" I asked. "It cuts off my circulation," Paul replied. "I know," I said. "It's supposed to. (170 LOL-Worthy Jokes About Marriage, 2018)

The polysemy of "circulation" in this joke heightens the humor of the script that in marriage there is a constant power struggle dynamic. The two scripts that oppose each other are COMFORT and INDEPENDENCE. The husband simply wants a ring that fits well, while the wife questions his commitment. The joke speaks to the fact that husbands and wives are sometimes suspicious of the fidelity of their spouses. In this example the wordplay acts as a supporting role to the humor found in the very real frustrations of marriage. Thus, the wordplay

strategy is actually paired with the power of cathartic humor. Conversely, here is an example from a non-adult joke:

Q: What did the pasta say to the tomato?  
A: Don't get saucy with me! (Tomato Jokes, 2018)

Again, there is a use of polysemy, but the joke does not rest much on another type of humor strategy. The audience presumably knows the simple script that tomatoes and pasta often go together in meals, but there is no established script of an anthropomorphic hostility from one food group to another. We could say that adult script jokes are not only more complex in their scripts, but also in some of the other features that have been mentioned.

Another interesting instance of the differences of joke composition was the degree of script knowledge needed for the jokes. Many of the political jokes required a knowledge of current events for the intended wordplay, as seen in the following example:

What is Donald Trump's favorite song?  
ICE ICE Baby..... (Donald Trump Jokes, 2018)

The humor of this joke relies on multiple pieces of knowledge: 1) Who Donald Trump is, 2) His political affiliation and policy platform, and 3) The meaning of ICE outside of the context of the song. In this sense the joke alludes to the government institution of Immigration and Customs Enforcement. It is a commentary on Trump's marked opinion and policy against illegal immigrants in the United States. The joke also depends on the audience to know that "ICE ICE Baby" was a self-aggrandizing hip hop song created in the 1990's. Although containing both a brief setup and punchline, there is a depth of knowledge required for the resolution of this joke. This is not to make a quality assessment of the joke itself or who would find it funny. Rather, it is a simple analysis of the underlying scripts and script overlay itself. Conversely, here is an example from the animal corpus:

Q: What is a cow's favorite place?

A: The mooseum. (Animal Jokes, 2018)

The setup is analogous to the previous example, but there is a clear lack of sophistication in the script. There is no social commentary on museums or cows. There is no usage of sarcasm or targeting involved in the setup or punchline. There is no relevance to real life situations, as cows do not travel to museums, nor do they voice preferences. Indeed, one might say that this joke is lacking a proper overlapping of scripts. There is no real pragmatic connection. Rather the humor of the joke resides solely in the phonemic similarity of the onomatopoetic utterance of a cow and the pronunciation of the word *museum*.

## 5. Conclusion

### 5.1 Answers to Research Questions

The purpose of this thesis was to determine the way puns were used in real world data. Specifically, I sought to discover the distribution of pun and narrative strategies according to joke topic/schema:

Table 9 Research Questions Revisited

- 
1. Is there a connection between pun usage and joke topics, and what is that connection?
  2. Is there a connection between narrative strategy and pun usage?
  3. Is there a connection between narrative strategies and joke topics?
  4. Other than their content, do jokes for children generally follow different strategies than jokes for adults?
- 

Using quantitative and qualitative methods we were able to answer these questions with some success.

#### 5.1.1. Pun and Topic Connection

Using a chi-square test for independence, we noticed that the pun usage and joke topics across the corpus were not independent. Thus, there is a connection between pun strategy and joke topics. As shown in Table 4, puns predominantly correlated with the child schemata found in Animal and Food jokes with 62.62% and 69.59% of jokes in each respective sub corpus containing puns. Conversely, puns were much less common among adult schemata. For the Marriage and Politics schemata, pun jokes comprised only 8.73% and 23.04% of the respective sub corpora.

Additionally, Table 3 showed us that the majority of the puns across the entire corpus were homonymous puns, while polysemous puns only comprised 35.43% of all puns (a percentage that can be extrapolated from the raw data in the table). However, within the marriage



corpus, the majority of puns were due to polysemy, with almost a 3 to 1 ratio of polysemous puns to homonymous puns. This is a curious finding that pushes us to ask why.

An important distinction between homonymy and polysemy is the degree of similarity between respective counterparts. Homonymous pairs share only a phonemic similarity, while polysemous pairs share both a phonemic and semantic similarity. This added similarity in polysemy requires greater maneuvering on the part of the joke creator to bind the two scripts together. The joke must not only accommodate the phonemic similarity, but also the semantic one as well. A joke that uses a polysemous pun requires a more carefully constructed setup in order to execute a meaningful punchline. It is also more impressive and clever to execute a polysemous pun that delivers on multiple linguistic levels (phonemic and morphological). We might wonder whether jokes based on more mature schemata like marriage are more likely to display greater linguistic prowess.

### **5.1.2 NS and Pun Connection**

A chi-square analysis also found that the narrative strategies were often related to the pun usages in each sub corpora. The majority of riddle type jokes contained puns, with Narrative/Dialogue jokes comprising only 6.58% of all pun jokes. Additionally, Table 6 showed that 79.08% of all pun jokes in the corpus were riddle jokes. It seems that the simple structure of riddles do lend themselves to pun strategies overall.

This analysis also yielded an interesting result concerning polysemous jokes. Of all the pun jokes the majority contained homonymous puns. However, the narrative/dialogue jokes were the only category to have more polysemous puns than homonymous puns with a ratio of 62 to 50. This result may coincide with the previous discussion on polysemous puns. Polysemy requires more care in the setup of the joke in order to provide an effective punchline. Perhaps

there is a correlation between narrative/dialogue structures and polysemy because the added length of the joke provides the sufficient amount of detail in the setup to execute the punchline properly. However, this speculation lacks sufficient data to make any firm conclusions. Such research would require greater numbers of jokes that employ narrative or story type strategies.

### **5.1.3 NS and Topic Connection**

As an additional perspective of the characteristics of the sub corpora, a chi-square analysis was conducted to see the relationship between narrative strategies and joke topics. In this final statistical analysis, it was found that the narrative strategies had proportionally the greatest relationship with the marriage corpus. Extrapolating the data from Table 7, we notice that 73.58% of marriage jokes followed the Narrative/Dialogue strategy.

The majority of the riddle jokes fell into the Animal and Food schemata, which was consistent with the previous research and analysis above. However, one surprising result was found within the political corpus. Only 7.95% of the political jokes used a Narrative/Dialogue strategy, discrediting the idea that this NS form is more related to adult schemata. Additionally, 44.01% of the political corpus used a riddle format. The rest of the political jokes showed a high level of one liner and stand-up comic observational humor. These characteristics seem to point to complexities among the schemata as we saw in the qualitative analysis.

### **5.1.4 Characteristics of Joke Schemata: Child versus Adult**

Combining all of these results together along with the qualitative analysis of the corpora, we find certain patterns beyond content emerge between the child and adult schemata. The child schemata seemed to favor homonymous puns with simple riddle structures. Conversely, the adult schemata used fewer puns with a higher proportion of polysemous puns to homonymous puns.

Although the marriage corpus was comprised mostly of narrative/dialogue type jokes, the political corpus did not follow this trend.

The political corpus diverged from the patterns found in the other corpora. Many of the political jokes came from late night comedy hosts, often in the form of mocking humor. This differs from the other corpora in the sense that the jokes generally were canned, with no attributed author. Negative or mocking type humor was also found often within the riddle jokes of the political corpus. These jokes targeted political figures and ideologies.

Overall the child and adult schemata differ in their approach to the joke. Although not uniform in their pattern differences, the adult schemata did diverge in pun and narrative strategy. These patterns may point to reasons for the negative perceptions of puns as a whole, but this will be explored in greater depth in the implications section. Before deriving any lessons from these patterns, we must address the limitations of the methodology of this thesis.

## **5.2 Limitations**

Some limitations with regard to the methodology of this thesis should be acknowledged. One important note comes from classification. Because taxonomies differ so much across authors and disciplines, it is difficult to make broad generalizations concerning humor. Some linguists may not agree with the classification of homonymous puns versus polysemous puns. Although etymology can provide strong empirical support to categorization, an audience will often lack an understanding of etymological roots. If the audience perceives a semantic connection between two similar words, the humor may be seen as more sophisticated regardless of the historical connection. A survey methodology that records public perceptions about the semantic connections between similar words may prove to be more useful in determining

wordplay types. However, as mentioned above, this would also be difficult to do pragmatically speaking.

Furthermore, the resources for finding the etymologies presented some problems of their own. As with most etymological resources, some roots or word forms simply lack the necessary data. Some entries would hedge, saying that a word “probably” had connections to the root of another word. In instances where a common root could not be tied to two similar words, the wordplay was deemed homonymous. Additionally, some wordplays were based on proper nouns whose origin remained dubious.

Another drawback comes in the form of search engine optimization (SEO) algorithms and joke selection. SEO algorithms can favor different styles of humor over others. Although web sites were selected for data collection to avoid some of the bias issues of printed collections, some bias can still remain. Some sites may favor narrative jokes over one-liners, and thus there is a degree of homogeneity in the selection. Some website managers may have a particular taste in humor and may choose to select jokes that include more or less wordplay humor according to personal preference. Practically speaking, it would be difficult to do a random sampling of jokes across many more websites, taking only a handful of jokes from each. Such efforts to attain greater diversity as well as reduce preprocessing in jokes would require greater use of spider bots or web crawlers that would automatically pull the data from websites on a grand scale.

Finally, there is the issue of equal comparison. Traditionally, corpora are compiled with the word count in mind as a delimiter. This was the method used in compiling the joke corpus for this thesis. However, upon reexamination of the methodology, it seems that it would be more productive to build corpora based on joke counts instead of word counts. This would make more sense as the analysis is based on a joke by joke basis. A corpus based on word count makes sense

when a linguistic analysis is based on vocabulary or phrase frequency. This allows the researcher to find multiple instances of said words or phrases within and across documents. In other words, corpora based on word count work well when dealing with discrete continuous data. Conversely, this thesis dealt with categorical data (e.g. did the joke contain a pun or not?) with jokes that varied in length. Using a word count approach with this type of data becomes an issue as corpora that contain long jokes (like the marriage corpus) receive less representation overall as the word count is reached sooner with fewer jokes. Thus, when dealing with joke analysis, it might be more useful to design corpora according to joke counts so that there is a more equal representation. Although statistical methods can balance uneven comparison ratios, this methodology could help establish stronger results by simply adding more data.

### **5.3 Implications**

These results seem to point towards a possible reason that puns are commonly regarded as unsophisticated. Although punning wordplay can be used in multiple situations across varying topics, puns are regarded negatively by many people. The many puns and riddles found in the child schemata corpora may point to a reason why puns are seen as low-brow or unsophisticated. If puns are continually paired with simple NS structure and content geared towards children, then the qualms that many people have with puns may be due to association. Puns may be regarded negatively among adults because so many pun jokes use material and joke structures that are basic and irrelevant to adult lives and minds. The prevalence of puns in scripts that appeal to children may create an association that puns are childish in and of themselves.

There also seemed to be a pattern in the types of negative humor with adult topics. It may not be fair to say that there is a direct correlation between negative humor and growing maturity. However, it could be said that the incongruities of life are more apparent as people mature and

encounter greater challenges and responsibilities. In order to find a more conclusive connection between polarity and maturity, such research would require a more thorough analysis of adult and child schemata. This thesis did not quantitatively assess polarity of jokes according to schema. However, this would be an excellent platform for future research.

#### **5.4 Future Research**

One additional feature that could be explored through a computational sentiment analysis is joke polarity. Although this thesis touched on the polarity of humor in the qualitative section of the results, there is much that could be done in quantifying positive versus negative humor with this corpus. Using machine learning techniques, certain statistical models could be used to track positive and negative polarity words within jokes. The model would subsequently assign polarity to the jokes themselves. This sentiment analysis would provide another layer of understanding of how puns are used in real world situations

This thesis compared pun strategy and topic, but did not investigate other types of wordplay according to schemata. Puns may have a prevalence among child schemata, but are all types of wordplay childish in their own right? Are there perhaps other types of wordplays that appeal to more adult schemata? These are questions that should be explored. However, this would also require greater exploration into standardized taxonomies.

No clear consensus exists as to what constitutes certain wordplay categories. Wordplay taxonomies abound much like humor taxonomies. However, there is little consensus or standardization. Any future research that tracks all types of wordplay in a given corpus will require a standard taxonomy. This is an area that will need more attention before more conclusive empirical data can be gathered.

The methodology of this thesis employed some technical skills to organize and sort the data, but the majority of labeling and analysis was conducted manually. Such time-consuming work makes it difficult to measure multiple features and triangulate data. However, with advances in machine learning techniques and greater feature engineering, humor analysis (and text analysis as a whole) will become more precise and quicker. Sentiment analysis as well as other deep learning techniques may help extract more objective data from large corpora. This will help us understand more humor strategies across schemata as well as culture.

## 6. References

- Apresjan, J. D. (2009). Regular polysemy. *Linguistics*, 12(142), 5-32.
- Attardo, S. (1994). Linguistic Theories of Humor. In V. Raskin, & M. Apte (Eds.), *Survey of Literature* (pp. 14-18). New York: Mouton de Gruyter.
- Attardo, S., & Raskin, V. (1991). Script Theory Revis(it)ed: Joke Similarity and Joke Representation Model. *Humor*, 4(3), 293-348.
- Attardo, S., Attardo, D. H., Baltes, P., & Petray, M. J. (1994). The linear organization of jokes: analysis of two thousand texts. *Humor*, 7(1), 27-54.
- Attardo, S., Raskin, V., & Ruch, W. (1993). Toward an Empirical Verification of the General Theory of Verbal Humor. *Humor*, 6(2), 123-136.
- Augarde, T. (1986). *The Oxford Guide to Word Games*. Oxford: Oxford University Press.
- Bates, C. (1999). The Point of Puns. *Modern Philology*, 96(4), 421-438.
- Beard, F. K. (2008). *Humor in the Advertising Business Theory, Practice, and Wit*. Lanham, Maryland: Rowland & Littlefield Publishers, Inc.
- Bell, N. D., Crossley, S., & Hempelmann, C. F. (2011). Wordplay in church marquees. *Humor*, 24(2), 187-202.
- Bergler, E. (1956). *Laughter and the sense of humor*. New York: Intercontinental Medical Book Corp.
- Biber, D., & Reppen, R. (Eds.). (2015). *The Cambridge Handbook of Corpus Linguistics*. Cambridge, UK: Cambridge University Press.
- Bush Family History*. (2018, January 10). Retrieved from Ancestry:  
<https://www.ancestry.com/name-origin?surname=bush>



- Cekaite, A., & Aronsson, K. (2005). Language Play, a Collaborative Resource in Children's L2 Learning. *Applied Linguistics*, 26(2), 169-191.
- Chrysostom, J. (1889). *On the Priesthood: Ascetic Treatises; Select Homilies and Letters; Homilies on the Statues*. New York: Christian Literature Co.
- Compound Words*. (2018, April 3). Retrieved from Cambridge Dictionary Online: <https://dictionary.cambridge.org/us/grammar/british-grammar/word-formation/compounds>
- Cook, G. (1994). *Discourse and literature : the interplay of form and mind*. New York: Oxford University Press.
- Dienhart, J. M. (1998). A Linguistic Look at Riddles. *Journal of Pragmatics*, 31, 95-125.
- Espy, W. (1971). *The Game of Words*. Devon, NY: Reader's Union Group.
- Freud, S. (1901). *The Psychopathology of Everyday Life*. London: Hogarth.
- Giora, R. (2002). *Optimal Innovation and Pleasure*. Tel Aviv: Tel Aviv University.
- Guidi, A. (2012). Are pun mechanisms universal? A comparative analysis across language families. *Humor*, 25(3), 339-366.
- Gulas, C. S., & Weinberger, M. G. (2006). *Humor in Advertising A Comprehensive Analysis*. Armonk, New York: M. E. Sharpe, Inc.
- Hempelmann, C. F. (2004). Script opposition and logical mechanism in punning. *Humor*, 17(4), 381-392.
- Hempelmann, C. F., & Ruch, W. (2005). 3 WD meets GTVH: Breaking the ground for. *Humor*, 18(4), 353-387.
- Kao, J. T., Levy, R., & Goodman, N. D. (2015). A Computational Model of Linguistic Humor in Puns. *Cognitive Science*, 40(5), 1270–1285.

- Knight, N. K. (2013). Evaluating Experience in Funny Ways: How Friends Bond through Conversational Hum. *Text & Talk*, 33(4-5), 553–574.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Last Name Meanings and Origins*. (n.d.). Retrieved from Ancestry:  
<https://www.ancestry.com/name-origin>
- Lems, K. (2011). Pun Work Helps English Learners Get the Joke. *The Reading Teacher*, 65(3), 197-202.
- Lillo, A. (2006). Cut-down puns. *English Today*, 22(1), 36-44.
- Lillo, A. (2008). Covert Puns As a Source of Slang Words in English. *English Studies*, 89(3), 319-338.
- Lippman, L. G., Sucharski, I. L., & Bennington, K. (2001). Contextual Connections to Puns in Fables: Perceived Humor. *The Journal of General Psychology*, 128(2), 157-169.
- Morreall, J. (2009). *Comic Relief: A Comprehensive Philosophy of Humor*. Chichester, UK: John Wiley & Sons.
- Oaks, D. D. (1995). Structural Ambiguities and Written Advertisements: An Inventory of Tools for More Resourceful Advertisements in English. *Journal of Technical Writing and Communication*, 25(4), 371-392.
- Oaks, D. D. (2010). *Structural Ambiguity in English: An Applied Grammatical Inventory*. Bloomsbury Academic.
- OED: Oxford English Dictionary*. (2017-2018). Retrieved from Oxford English Dictionary:  
<http://www.oed.com.erl.lib.byu.edu/>
- Online Etymology Dictionary*. (2017). Retrieved from Etymonline: <https://www.etymonline.com/>

- Otake, T., & Cutler, A. (2013). Lexical Selection in Action: Evidence from Spontaneous Punning. *Language and Speech*, 56(4), 555-573.
- Partington, A. S. (2009). A linguistic account of wordplay: The lexical grammar of punning. *Journal of Pragmatics*, 41(9), 1794-1809.
- Pepicello, W. J., & Green, T. A. (1984). *The Language of Riddles: New Perspectives*. Columbus: Ohio State University Press.
- Piddington, R. (1933). *The Psychology of Laughter: A Study in Social Adaptation*. London, UK: Figurehead.
- Raskin, V. (1984). *Semantic Mechanisms of Humor*. Dordrecht: Springer.
- Ravin, Y., & Leacock, C. (Eds.). (2000). *Polysemy: Theoretical and Computational Approaches*. Oxford: Oxford University Press.
- Samson, A. C., & Gross, J. J. (2012). Humour as emotion regulation: The differential. *Cognition and Emotion*, 26(2), 375-384.
- Sutton-Smith, B. (1976). A Developmental Structural Account of Riddles. *Society for Research in Child Development*, (pp. 1-7). Philadelphia.
- Taylor, C. S. (2013). *Validity and Validation*. New York: Oxford University Press.
- The Primer of Humor Research. (2008). In V. Raskin (Ed.). New York: Mouton de Gruyter.
- Valitutti, A., Doucet, A., Toivanen, J. M., & Toivonen, H. (2016). Computational generation and dissection of lexical replacement humor. *Natural Language Engineering*, 22(5), 727-749.
- Vande Kopple, W. J. (1995). Pun and Games. *English Journal*, 84(1), 50-54.

## 7. Appendix

### 7.1 Data Sources for Corpus and In-Text Examples

*10 Of The Corniest Food Jokes On The Planet.* (2017, October 25). Retrieved from Huffington

Post: [https://www.huffingtonpost.com/2014/07/02/food-jokes\\_n\\_5549424.html](https://www.huffingtonpost.com/2014/07/02/food-jokes_n_5549424.html)

*170 LOL-Worthy Jokes About Marriage.* (2018, March 12). Retrieved from Reader's Digest:

<http://www.readersdigest.ca/health/relationships/jokes-about-marriage-to-tell-at-a-wedding/8/>

*Animal Jokes.* (2017, October 25). Retrieved from Funology: [http://www.funology.com/animal-](http://www.funology.com/animal-jokes/)

[jokes/](http://www.funology.com/animal-jokes/)

*Animal Jokes.* (2017, October 25). Retrieved from Ducksters Education Site:

<https://www.ducksters.com/jokes/animals.php>

*Animal Jokes.* (2017, October 25). Retrieved from Reader's Digest:

<https://www.rd.com/jokes/animal/>

*Artichoke Jokes.* (2017, October 25). Retrieved from Jokes 4 Us :

<http://www.jokes4us.com/miscellaneousjokes/foodjokes/artichokejokes.html>

*Bagpipe Jokes.* (2018, March 1). Retrieved from Jokes 4 Us:

<http://www.jokes4us.com/miscellaneousjokes/musicjokes/bagpipejokes.html>

*Bell Pepper Jokes.* (2017, October 25). Retrieved from Jokes 4 Us:

<http://www.jokes4us.com/miscellaneousjokes/foodjokes/bellpeppersjokes.html>

*Donald Trump Jokes.* (2018, March 12). Retrieved from Jokes 4 Us:

<http://www.jokes4us.com/celebrityjokes/donaldtrumpjokes.html>

*Food Jokes.* (2017, October 25). Retrieved from Jokes 4 Us:

<http://www.jokes4us.com/miscellaneousjokes/foodjokes/index.html>

*Funny One-liner #1251.* (2018, March 12). Retrieved from One Line Fun:

<https://onelinefun.com/oneliner-1251/>

*Funny Puns and One Liners.* (2018, March 12). Retrieved from Pun.me: <http://pun.me/id/282>

*Husband and Wife Jokes.* (2017, October 25). Retrieved from Short-Funny: [http://www.short-](http://www.short-funny.com/marriage-wife-husband-jokes.php)

[funny.com/marriage-wife-husband-jokes.php](http://www.short-funny.com/marriage-wife-husband-jokes.php)

*Joke #11654.* (2018, March 12). Retrieved from Unijokes: <https://unijokes.com/joke-11654/>

*Joke #4925.* (2017, October 25). Retrieved from Unijokes: <https://unijokes.com/joke-4925/>

*Joke Bank - Animal Jokes.* (2017, October 25). Retrieved from The Laugh Factory:

<http://www.laughfactory.com/jokes/animal-jokes>

*Joke Bank - Political Joke.* (2017, October 25). Retrieved from The Laugh Factory:

<http://www.laughfactory.com/jokes/political-jokes>

*Joke Bank- Animal Jokes.* (2017, October 25). Retrieved from Laugh Factory:

<http://www.laughfactory.com/jokes/animal-jokes>

*Jokes for Happy Harry.* (2018, March 12). Retrieved from Happy Harry Jokes:

<https://happyharryjokes.wordpress.com/2009/07/02/did-you-hear-about-the-man-who-fell-into-an-upholstery-machine/>

*Kids Jokes - Animal Jokes.* (2018, March 12). Retrieved from Kid's Jokes of the Day:

[http://www.kidsjokesoftheday.com/htmls/kids-jokes/Animal-Jokes-for-Kids/20150413180711\\_254.html](http://www.kidsjokesoftheday.com/htmls/kids-jokes/Animal-Jokes-for-Kids/20150413180711_254.html)

*Marriage Jokes.* (2017, October 25). Retrieved from Reader's Digest:

<https://www.rd.com/jokes/married-life/>

*Peanut Jokes.* (2018, March 12). Retrieved from Jokes 4 Us:

<http://www.jokes4us.com/miscellaneousjokes/foodjokes/peanutjokes.html>

*Political Jokes.* (2017, October 25). Retrieved from Jokes 4 Us:

<http://www.jokes4us.com/politicaljokes/index.html>

*The Best Marriage Jokes.* (2017, October 25). Retrieved from Unijokes:

<https://unijokes.com/marriage-jokes/>

*Tomato Jokes.* (2018, March 12). Retrieved from Jokes 4 Us:

<http://www.jokes4us.com/miscellaneousjokes/foodjokes/tomatojokes.html>

## 7.2 Survey of Children's Joke Books

Brown, P. (2017). *Silly Jokes for Kids*. Lexdo Publications Ltd.

Dahl, M. (2001). *The everything Kids' Joke Book*. Avon: Simon and Schuster Inc.

Elliot, R. (2010). *Laugh Out Loud Jokes for Kids*. Grand Rapids: Baker Publishing group.

Grace, C. (2017). *Best Jokes Ever: Jokes for Kids*. Racine, Wisconsin: Broadstreet Publishing.

Hogan, J. (2015). *Jokes For Kids: Kids Jokes: 300 Funny Jokes*. Wiq Media.

MacDonald, P. (2012). *Best Joke Book for Kids*. CreateSpace Independent Publishing Platform.

Rothman, J. (1990). *1,000 Howlers for Kids*. Manhattan: Random House Publishing Group.

Superstars, A. (2012). *Archie's Giant Kids' Joke Book*. Mamaroneck: Archie Comics.

Voeg, C. (2015). *365 Jokes For Kids: A Joke A Day Book +5 Bonus Magic Tricks*. Philadelphia:  
The Kruse Group.

Willy, S. (2014). *Silly Jokes for Silly Kids*. CreateSpace Independent Publishing Platform.