5-31-2022

Integrating Cultural Knowledge into Artificially Intelligent Systems: Human Experiments and Computational Implementations

Anurag Acharya
Florida International University, aacha007@fiu.edu

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

INTEGRATING CULTURAL KNOWLEDGE INTO ARTIFICIALLY INTELLIGENT SYSTEMS: HUMAN EXPERIMENTS AND COMPUTATIONAL IMPLEMENTATIONS

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTER SCIENCE

by

Anurag Acharya

2022

To: Dean John L. Volakis
College of Engineering and Computing

This dissertation, written by Anurag Acharya, and entitled Integrating Cultural Knowledge into Artificially Intelligent Systems: Human Experiments and Computational Implementations, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Monique Ross

Fahad Saeed

Leonardo Bobadilla

Ronald Fisher

Mark A. Finlayson, Major Professor

Date of Defense: May 30, 2022

The dissertation of Anurag Acharya is approved.

Dean John L. Volakis
College of Engineering and Computing

Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2022

DEDICATION

To my mother Bijaya and sister Alaka.

They say Saraswati is the goddess of knowledge and wisdom. So the woman who despite not being allowed to continue her education raised both her children to be PhDs, and the woman who taught me everything I know, you two are my Saraswati.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor Dr. Mark Finlayson for being my teacher, mentor, and friend these five years. I would also like to thank Victor, my friend and colleague, without whom this would not be possible. I also want to acknowledge my undergrduate researcher Diego for all numerous help he has provided. I am also grateful to my forever partners-in-crime Mustafa and Labiba, as well as all my other labmates and friends at Cognac lab for all their help and support. I would also like to thank my brother Puskar, and my friend and brother Jose and the entire Lopez family for making me feel at home so far away from home. I also want to thank Anshu, Anushuiya, Paranjay, Soph and all my friends for their support over the years. And finally, but most importantly, I want to thank Manushi for being my rock through everything; and my parents, my sister, my brother-in-law, and my entire family for their support.

ABSTRACT OF THE DISSERTATION

INTEGRATING CULTURAL KNOWLEDGE INTO ARTIFICIALLY INTELLIGENT SYSTEMS: HUMAN EXPERIMENTS AND COMPUTATIONAL IMPLEMENTATIONS

by

Anurag Acharya

Florida International University, 2022

Miami, Florida

Professor Mark A. Finlayson, Major Professor

With the advancement of Artificial Intelligence, it seems as if every aspect of our lives is impacted by AI in one way or the other. As AI is used for everything from driving vehicles to criminal justice, it becomes crucial that it overcome any biases that might hinder its fair application. We are constantly trying to make AI be more like humans. But most AI systems so far fail to address one of the main aspects of humanity: our culture and the differences between cultures. We cannot truly consider AI to have understood human reasoning without understanding culture. So it is important for cultural information to be embedded into AI systems in some way, as well as for the AI systems to understand the differences across these cultures.

The main way I have chosen to do this are using two cultural markers: motifs and rituals. This is because they are both so inherently part of any culture. Motifs are things that are repeated often and are grounded in well-known stories, and tend to be very specific to individual cultures. Rituals are something that are part of every culture in some way, and while there are some that are constant across all cultures, some are very specific to individual ones. This makes them great to compare and to contrast.

The first two parts of this dissertation talk about a couple of cognitive psychology studies I conducted. The first is to see how people understood motifs. Is is true that in-

culture people identify motifs better than out-culture people? We see that my study shows this to indeed be the case. The second study attempts to test if motifs are recognizable in texts, regardless of whether or not people might understand their meaning. Our results confirm our hypothesis that motifs are recognizable.

The third part of my work discusses the survey and data collection effort around rituals. I collected data about rituals from people from various national groups, and observed the differences in their responses. The main results from this was twofold: first, that cultural differences across groups are quantifiable, and that they are prevalent and observable with proper effort; and second, to collect and curate a substantial culturally sensitive dataset that can have a wide variety of use across various AI systems.

The fourth part of the dissertation focuses on a system I built, called the motif association miner, which provides information about motifs present in input text, like associations, sources of motifs, connotations, etc. This information will be highly useful as this will enable future systems to use my output as input for their systems, and have a better understanding of motifs, especially as this shows an approach of bringing out meaning of motifs specific to certain culture to wider usage.

As the final contribution, this thesis details my efforts to use the curated ritual data to improve existing Question Answering system, and show that this method helps systems perform better in situations which vary by culture. This data and approach, which will be made publicly available, will enable others in the field to take advantage of the information contained within to try and combat some bias in their systems.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation

You are walking down the road, when you come across a woman dressed in white. Can you guess whether she is likely happy or sad?

If you were brought up in the Western hemisphere, your first instinct was likely that she is happy – it is probably her wedding day. If you were brought up in the Eastern part of the world, the Indian subcontinent in particular, however, your understanding is that she must be quite upset – whites are for death and mourning, and on a woman likely indicating a very close family member: perhaps a parent or spouse.

While we humans are inherently similar, we also have quite a bit of differences. These differences, while not defining who we are, are definitely part of us. And a big chunk of those differences come from the difference in culture.

As the field of Artificial Intelligence (AI) and Natural Language Processing (NLP) has evolved, there has been a significant amount of focus on how to make them more universal, intelligent, and overall more *human-like*. But the majority of the work in the field seems to miss out on one key thing that is inherently part of us humans: culture. This is one glaring omission.

Our cognition is shaped quite a bit by culture, because what we see or do growing up and almost all our lives are more or less defined by our culture. (DiMaggio, 1997; Hong et al., 2000; Berry and Dasen, 2019) Most of our perspective and language is shaped by culture as well. (Sapir, 1985; Freire and Macedo, 1995; Lazear, 1999; Kramsch, 2014) Without this supplemental knowledge that is culture-specific, we are missing out on a wealth of information. It is hard to imagine being able to construct a human-like system when one of the factors that shape human cognition is not incorporated into the process. As we move forward. it is vital that we embed cultural information into AI systems.

But in order to be able to use cultural knowledge for AI systems, we need to understand more about the cultural differences. It is my belief that if one is to look to break ground into culturally sensitive AI, it is vital that we begin from the actual people whose nuances we are tying to integrate into the systems. This is why I think before getting into any computational methods, we need to first understand how the people themselves understand and perceive cultural similarities and differences. With this idea, I focused on learning about cultural phenomenon from humans, and then in turn using the knowledge from them to then build computational systems to incorporate those learning.

However, this is not straightforward, since cultural knowledge can be fairly arbitrary and complex across groups. (Asad, 1993; Turner, 1973) Prior literature from the study of cultural groups seems to suggest that it is hard enough just to define "culture" given the array of complexities and nuances; Schein (1991); Spencer-Oatey and Franklin (2012) however, it is even harder to identify the knowledge that comes along with it and to differentiate across cultural groups. (Bell, 1992, 1997) Furthermore, as cultural norms and practices are so varied across various groups, it is hard to find knowledge structures that can be clearly compared across different groups. The first task, therefore, is to find topics that are relatively common across different groups, and that can be compared in a like-for-like fashion, thus demonstrating the relevance of culture-dependent representation of commonsense knowledge. Therefore the process of injecting cultural information must be done by selecting the appropriate means. I propose that this begin using two different types of cultural knowledge: *rituals* and *motifs*.

1.2 Background

1.2.1 Motifs

Motifs are distinct, recurring narrative elements found in folklore and, more generally, cultural materials. They are specific narrative elements that are repeated across artifacts found within the same cultural group. Motifs occur everywhere—in social media posts, everyday conversation, news stories, books, and so forth. A popular definition of motif is given by Thompson, who defines it as "a motif is the smallest element in a tale having a power to persist in tradition. In order to have this power it must have something unusual and striking about it. Most motifs fall into three classes. First are the actors in a tale [...] Second come certain items in the background of the action [...] In the third place there are single incidents [...] " (Thompson, 1977, pp. 415–416).

The idea of the motif first originated in folklore, and it is in the study of folklore where the classic example of motifs are found, such as a hero fighting a dragon (event), a troll under a bridge (character), or a glass slipper (prop). Examples from the modern day include, for example, the 3 a.m. phone call (an event used to indicate the start of a trying crisis for a political leader), a good Samaritan (a character who selflessly helps someone, originating from Judaeo-Christian mythology when a Samaritan helps a needy traveler), and a magic wand (a prop that can be used to suddenly fix big problems magically).

Let us look at one of them that is particularly common in the west - "troll under a bridge." To members of many Western cultures, this combination entails a number of related ideas that are by no means directly communicated by the surface meaning of the words: the bridge is along the critical path of the hero, and he must cross to achieve his goal; the troll often lives under the bridge, crawling out to waylay innocent passers-by; the troll charges a toll or exacts some other payment for crossing the bridge; the troll is a squatter, not the 'officially' sanctioned master of the bridge; the troll enforces his illegitimate claim through threat of physical violence; and the hero often ends up battling (and defeating) the troll instead of paying the toll. Because of this density of information, motifs are often retained as a tale is passed between cultures and down generations, and folklorists have observed that a tale's specific composition of motifs can be used to trace the tale's lineage (Thompson, 1977, Part 4, Chapter V).

1.2.2 Rituals

There are several definitions of ***ritual***, and the concept often comes entangled with religion and rites (Braun and McCutcheon, 2000, p. 259-262). For the purposes of this work, I consider two definitions. Broadly defined, ritual is simply a "culturally defined set of behavior" (Leach, 1968). More specifically, a widely accepted definition is by (Turner, 1973), who defines ritual as: "…a stereotyped sequence of activities involving gestures, words, and objects, performed in a sequestered place, and designed to influence preternatural entities or forces on behalf of the actors' goals and interests".

There are several reasons for focusing on rituals as indicators of cultures. First, rituals are well-studied and quite a bit of prior work exists for us to build on. Prior work has identified not just specific rituals, but also the genres and types of rituals and their variance across cultures, etc. (Durkheim and Swain, 2008; Turner, 1973; Bell, 1992, 1997). There is also a rich body of literature on the analysis of cultural practices for various rituals across several cultures, clearly indicating a strong relation between rituals and cultures (Cliford, 1973; Gray, 1979; Ulrich, 1984; Smith, 1986; Dean, 1997; Cantú, 1999; Underhill, 2000).

Over time, it has been observed that people ritualize all sorts of activities to varying degrees. The question around rituals as knowledge, therefore, is less about whether or not a specific ritual is observed in a culture but rather the *degree* or *extent* to which it is

Figure 1.1: The then UK Prime Minister Theresa May (first from left) shares the stage with co-candidate ”Lord Buckethead” (second from right). Photo by BBC News.

observed (Bell, 1997). Rituals can be thought of as a spectrum: on one end are rituals that are codified by tradition or text, and presided over by experts, while on the other are ritual-like activities such as sports events (e.g. Superbowl) or cocktail parties—including everything from social etiquette to sports events and political spectacles.

While a lot of rituals seem to have connections with religion, they do not have to have that. Let us look at an example of a particularly popular ritual from the United Kingdom: it is customary that after each election, all the candidates from that particular constituency stand together on stage for the announcement of the votes. And this means *all* candidates, including joke candidates that simply stood the election as a form of protest. So this sometimes creates a hilarious situation as it did in June 2017, where the then incumbent Prime Minister Theresa May had to be in a stage as an equal competitor to *Lord Buckethead* (See Figure 1.1).

While there is great variety in the conceptualization of ritual, the most important taxonomies of ritual show basic agreement on their core categories. Rituals can be classified in several ways: one of the more widely accepted categorizations of rituals by Bell (1992, 1997) presents a compromise of six categories that are not necessarily mutually exclusive. These categories are:

1. *rites of passage*, a.k.a. *life-cycle rites*;
2. *calendrical and commemorative rites*;
3. *rites of exchange and communion*;
4. *rites of affliction*;
5. *rites of feasting, fasting, and festivals*;
6. *political rituals*.

1.3 Research Components

For my dissertation, I propose five main thrusts of work across two dimensions – cognitive psychological studies and computational systems – with the ultimate aim of integrating cultural knowledge into artificially intelligent systems. These five components are listed below:

Component I: Motif Understanding Experiment – This step involves conducting a cognitive psychological experiment to see if motifs are understood more by people within the culture from which the motif originates, as compared to the people outside the culture. To do this I designed an experiment to test the understanding of people when they read various texts with or without motifs present in them.

Component II: Motif Recognizability Study – Similar to the first component, this is a cognitive psychological study to determine if motifs are recognizable in texts, i.e., can people detect that there is some extra information present in the text that they are missing, regardless of whether or not they understand the meaning of the said motif. This was also conducted in the form of a survey which involved participants reading a prompt text and answering questions based on them.

Component III: Ritual Survey / Data Collection – This task involves two main tasks. The first is to conduct a survey for ritual data to determine if ritual data is actually recordable and distinguishable across cultures. After I was able to collect data to show that ritual data can be properly collected showing similarities and differences, I go on to scale the data collection to a larger scale, creating enough datapoints to create a culturally sensitive dataset based on rituals. This component also serves as a direct basis for the final piece, which uses this dataset.

Component IV: Motif Association Miner – This computational system I built takes texts that contain motifs in them, and generates a report that lays out the various associations of that particular motif. This system is expected to be useful for any system that might need motif information, but also to human operators who might need information about said motifs. This systems helps make the motif meaning clearer than any system has done before.

Component V: Cultural Commonsense QA – This system take the curated cultural dataset from the third component and uses it to improve upon an existing Question Answering system. In addition to improving the performance of QA systems on culturally sensitive questions, this also achieves the larger goal of demonstrating how we can use cultural dataset to make a more enhanced and more inclusive system.

1.4 Dissertation Contributions

My work presented in this dissertation has the following major contributions, roughly aligning to each of the components listed above:

1. I tested the relationship of motif understanding with group membership and determined the relationship that exists between the two. In addition, I also looked at how these understandings differ for different categories of motifs: character, event, and prop.

2. I demonstrated that motifs are inherently recognizable in texts, irrespective of whether or not people might understand the meaning of these motifs. This in turn also demonstrates that motifs carry a markedly significant amount of information in any given text.

3. I demonstrated that cultural similarities and differences across groups can be recorded and collected using rituals, and layed out an approach for how it can be used for large-scale data collection.

4. I collected, cleaned, and curated a large dataset for culturally sensitive information, allowing future work to use this dataset and leverage the wealth of information present in them.

5. I built a system that can take as input some text containing motifs and generate a report with the associations for the motifs. This also has the additional significance of demonstrating how we can use NLP systems to enhance our ability to learn and know about specific cultural knowledge that might otherwise not be available to people out-of-culture.

6. I used the cultural dataset to improve the performance of Question Answering (QA) systems on culturally sensitive questions. Again, this work also pioneered the way for other systems in the future to be enhanced and upgraded so that they become less oblivious to cultural differences.

1.5 Outline

This dissertation is organized as follows: First in chapter 2, I lay out some of the other work done in the field related to my work described here. The rest of the dissertation is organized so that it closely follows the components mentioned in §1.3, in order. Secondly in chapter 3, I talk about the experiment I performed for Motif Understanding. I explain in detail the experiment instruments I designed and why I designed them that way, and the way I went about conducting the experiment. I also detail the results of the study and discuss the implications. In chapter 4, I talk about the Motif Recognizability study I conducted, going into similar details as the first one. Then in chapter 5, I detail the way I conducted the ritual survey, including my choice of rituals and my choice of the questionnaire. I also go into the details of the data I gathered. In chapter 6, I describe the Motif Association Miner (MAM) system I built. I talk both about the larger picture I envision for the MAM to be used in the future, as well as the components of the system I built and the output obtained. Then in chapter 7, I explain how I use the cultural dataset collected as described in chapter 5 to retrain and improve an existing QA system. Finally, I finish in chapter 8 with a conclusion that briefly re-assesses the contributions and results of each of the preceding chapters, and the overall implication of this work. The specific motivation for each individual work are also included in their respective chapters.

CHAPTER 2

RELATED WORKS

In this chapter, I mention several related works in the fields that this dissertation deals with. Where possible, I will try to relate the work I have done to the existing work. For work that I have done that don't quite have a parallel in existing literature, I take note of similar work or similar approach albeit used for a different end result. I have divided this into sections depending on the areas of research.

2.1 Motifs

Folklorists have constructed motif indices that identify motifs and note their presence in specific folktales. The most well-known motif index is Stith Thompson's motif index (TMI) (Thompson, 1960), which references folktales from over 600 collections, indexed to 46,248 motifs and submotifs. In addition to this, Thompson provides substantial discussion on motifs and the compilation of motif indices in his book The Folktale (Thompson, 1977). Additionally, there are many motif indices that target specific cultures and periods, for example, early Irish literature (Cross, 1952), traditional Polynesian narratives (Kirtley, 1971), Japanese folk-literature (Ikeda, 1971), or early Icelandic literature (Boberg, 1966). These works show that motifs are prevalent across cultures while individual motifs also being unique to the particular cultures.

There is limited work in the field of computational linguistics on motifs, but there have been some. Darányi (2010) has called attention to the need for research into the automation of extraction and annotation of motifs in folklore, and suggested that motifs have application in storing, indexing, and retrieving documents based on the motifs contained within. Work has also been done examining the shortcomings and potential applications of motifs. Darányi and Forró (2012) have determined, based on cluster analysis, that motifs may not be the highest level of abstraction in narrative, echoing criticisms that many motifs are interdependent (Dundes, 1997). Darányi et al. (2012) have made substantial headway towards using motifs as sequences of "narrative DNA", and Ofek et al. (2013) have demonstrated learning tale types based on these sequences. Declerck et al. (2012) have worked on converting electronic representations of TMI and ATU (Uther, 2004) to a format that enables multilingual, content-level indexing of folktale texts, building upon past work (Declerck and Lendvai, 2011). Currently, this work appears to be focused on the descriptions of motifs and tale types, without reference to the stories.

More recently, Yarlott et al. (2021) did work on automatic detection of some Irish, Jewish, and Puerto Rican motifs. This system, utilizing a pipeline of NLP components and a Machine Learning model, is able to detect the motif in text and its motif type. This work is especially relevant to me as it was done in conjunction with my work explained here, and thus uses the same list of motifs as I use for several of my tasks. Despite being one-of-its-kind when it comes to automatic work on motifs, the system reports a modest performance, and raises issues about how the difficulty in generalization and lack of sufficient data make any computational work in motifs unusually difficult.

2.2 Rituals

Rituals are well-studied and quite a bit of prior work exists for us to build on. One of the most foundational work on rituals is by Bell (1997). In this work, she examines the history of interpretation of rituals as well as lays out the spectrum of ritualistic activities. It is also in this work that she laid out the widely accepted six categories of rituals: (i) rites of passage, (ii) calendrical rite, (iii) rites of exchange and communion, (iv) rites of affliction, (v) rites of feasting, fasting, and festivals, and (vi) political rites. Another major work is by Durkheim and Swain (2008) in which they study various religious practices, beliefs and rites. Moreover, a crucial contribution in the field is the study of the genealogy of rituals is by Asad (1993), who examines rituals by focusing on Shamanic Performances and examining how these relate to wider processes of change, adaptation, or conflict within the society at large.

I am mainly using rituals as an analog of culture for portions of this work. To do this, I look to prior work in the field to indicate that there is a strong concomitance between rituals and cultures. Gray (1979) and Smith (1986) look at rituals that involve sacrificial rites as a key feature in South Asian cultures. Dean (1997) looks at rituals as one of the markers for Chinese society. Likewise, Cantú (1999) studies life-cycle rituals in Hispanic and latin American cultures. Another important thing to look at is to see if there exists studies that look at the different genres and types of rituals, as well as their variations across cultures. Durkheim and Swain (2008) examines what she calls the *elementary form* of religious life across different cultures, and Turner (1973) looks at symbols used in rituals across several cultures in Africa. Bell (1992, 1997) examines rituals in a much more detailed manner, studying their various types and how they resemble and differ across cultures. These works indicate that rituals can be considered the markers for cultures, and that they can be used to compare and contrast across cultures.

There have also been work done on the specific rituals that I am using. Beckman (1983) studies the birth rituals in the Hittite Empire, while Matthews et al. (2005) studies the birth rituals in rural South India. Dessing (2001) studies the birth, wedding, and funeral rituals in The Netherlands. Roche and Hohmann (2013), Chesser (1980), and Nelson and Deshpande (2004) all study wedding rituals across various cultures, while Cantú (1999) and Gray (1979) talk about coming-of-age rituals. Some of the rituals I focus on – like New Year's and Birthday celebrations – are non-religious in nature. There have been some prior work in studying non-religious rituals. Sadomskaya and Dragadze (1990) looks at the social, non-religious rituals of the (then) Soviet Union, while Ulrich (1984) examines the rituals and culture of organizations treating them like individuals.

While we see that cultures and rituals are a well-studied field, the computational study of the same has been slightly less popular. But there are still some work done in the field. Nielbo et al. (2012), Nielbo and Sørensen (2013) apply computational techniques to study religions and rituals. Frank and Reiter (2010) use frame semantics to study and analyze ritual description in texts. There have also been a handful of works that leverage AI to study rituals. Reiter et al. (2010, 2011) utilizes NLP tools to process linguistic data in order to study Indian rituals. However, there has yet been any work on incorporating rituals *into* AI systems. I take this first step to leverage this crucial information to better inform AI systems and inject them with cultural sensitivity.

2.3 Survey-based Research Studies

Needless to say, surveys have always been popular technique for a long time now, and especially in language comprehension studies. To name a couple, Daneman and Carpenter (1980) looked at the difference between people in terms of their reading comprehension by assigning them texts to read and asking questions based on them. Similarly, Carretti et al. (2013) studied language comprehension in older healthy adults with the help of text surveys. For several components of this dissertation, I have utilized text-based surveys to conduct studies as well as collect data from participants.

While there does not seem to have been any study done specifically on motif understanding or recognizability, there have been similar work done in related topics in the past. For instance, there have been several works done on other types of non-literal linguistic components. Jamrozik et al. (2013) and Wolff and Gentner (2011) conducted studies on metaphor comprehension, asking participants to read a text and answer questions based on them. Another couple of studies done in the field were by Thibodeau and Durgin (2011), who asked people to read texts containing metaphors and then had them judge those text for aptness and conventionality, and by Gokcesu (2009), who asked people to read texts containing metaphors and asked them to rate the aptness using a Likert scale. As metaphors are probably the closest linguistic thing to a motif, these studies show that text based surveys are appropriate to study motifs and their comprehension.

Another main component of my work here is to examine the cultural differences using motifs and rituals. There have been several studies conducted to examine cultural differences across a variety of domains using text-based surveys. Li and Kirkup (2007) examines the cultural differences in Internet use, while Zwikael et al. (2005) studies how cultural differences affect project management capabilities. Similarly, Fuchs et al. (2004) conducted a study on the cultural difference in how people perceive risk in tourist destinations. Another work done using this technique is by Yi and Park (2003), who surveyed college students across five different countries to understand the cultural differences in decision-making styles. Finally, Ayabe-Kanamura et al. (1998) studies how people from different cultures perceive everyday odors differently. The fact that text-based surveys can be used across these wide variety of domains to study cultural difference indicates it is also a good procedure to study difference across cultures for motifs and rituals.

In addition to these works, I also take guidance from various prior work that provide insight and guidance into how survey-based research should be conducted in linguistics (Hatch and Lazaraton, 1991; Dörnyei, 2007; Litosseliti, 2018), psychology (Breakwell et al., 2006; Willig, 2013), social sciences (May, 2011; Sarantakos, 2012; Lune and Berg, 2017; Mason, 2017), and in general overall (Oppenheim, 1992; Rossi et al., 2013; Brace, 2018); as well as how to best utilize Likert scales. (Jamieson, 2004)

2.4 Relation Detection and Information Extraction Systems

The Motif Association Miner work described in chapter 6 is best thought of as a mix of targeted information extraction, relationship detection, and template filling tasks. There have been several works done in the general field of relation detection and information extraction in the past.

Soares et al. (2019) uses distributional similarity to build a general relation extraction system using BERT. Devlin et al. (2018) Similarly, Wu and He (2019) also uses BERT to perform relation classification while also incorporating information about large entities. Meanwhile Ye and Ling (2019) and Wang et al. (2016) implement CNN-based few-shot relation classifiers. There are several other systems that perform well when it comes to relation detection task, like Bastos et al. (2021), a neural network based model; Kim et al. (2019), an RNN based model; and Cai et al. (2016), a BRCNN based model. While all these systems differ quite a bit from my approach to Association Miner – they all use deep learning and I do not, for starters – they all aim to solve a similar task to mine.

There has also been a lot of work on Information Extraction systems, both as a whole and focused on individual problems inside information extraction. One important work is by Stanovsky et al. (2018) which re-frames open information extraction as a sequence tagging problem. Similarly, Cetto et al. (2018), Gashteovski et al. (2017), and Bhutani et al. (2016) are more recent information extraction systems that perform close to the state of the art (SOTA).

In terms of work that focuses on specific subtask of OpenIE, Pal et al. (2016) focuses on nominal OpenIE, which finds an efficient way to extract open relations for compound noun phrases. Similarly, Saha et al. (2017) focuses specifically on numerical relations to extract OpenIE tuples, and Saha et al. (2018) addresses the issue of extracting relation tuples for conjunctive sentences. As with the relation detection systems, while these work resemble mine in terms of the end goal, they differ in terms of the technique used to achieve these goals.

Finally, there are a few template filling work that are at least distantly related to my work in Association Miner, for which I perform a pseudo-template filling task of creating a fixed set of columns to be filled by the processed data. One of the work by Jean-Louis et al. (2011) which combines text segmentation and graph techniques to perform template filling. Another similar work is by Miliani et al. (2019) which splits text into frames in order to accomplish slot filling. But perhaps the work that most closely relates to my work is work done by Chambers and Jurafsky (2011) which combines the tasks of information extraction and template filling, but is able to do so without having to know a fixed template for the output in advance. For the task I perform, I draw on these for the top level approach, even though they differ significantly in implementation.

2.5 Commonsense Reasoning and Question Answering

The ability to incorporate commonsense knowledge into various NLP tasks (Clark et al., 2018; Boratko et al., 2018) can vastly improve the quality of the returned responses, as well as the accuracy of the work done in the field (Marcus, 2018). While the concept of incorporating cultural knowledge into commonsense is fairly unique, there have been several previous works that have laid the groundwork for it by building commonsense reasoning systems. I build on the ATOMIC (Sap et al., 2019b) system and knowledge repository, where crowdsourced commonsense information was used to build an atlas for if-then reasoning. The dataset is then used on the social question answering system SocialIQA (Sap et al., 2019c), which shows an increase in performance using the commonsense knowledge from ATOMIC.

In addition to these aforementioned works, there have been several other efforts in the field of commonsense reasoning. Eventent (Espinosa and Lieberman, 2005) deals with inferring temporal relations between commonsense events. Related work includes a system called Event2mind (Rashkin et al., 2018) – a commonsense inference system on events, intents, and reactions; an updated version of ConceptNet (Liu and Singh, 2004), a practical commonsense reasoning toolkit, which is now a multilingual graph of general knowledge (Speer et al., 2016); and Webchild 2.0, a fine-grained commonsense knowledge distillation (Tandon et al., 2017). Several of the other important pieces of work done in the field of commonsense have been reviewed by Davis and Marcus (2015), which lays out the uses, successes, challenges, approaches and possible future work in the field of commonsense reasoning.

Another important one is the Stanford Question Answering Dataset (SQuAD) (Rajpurkar et al., 2016), consisting of 100,000+ questions and a reading comprehension dataset. They contrast three types of tasks: reading comprehension (RC; read a passage, select a span that answers); Open-domain QA (answer a question from a large set of documents); and Cloze datasets (predict a missing word in a passage).

Apart from these applications, (Gordon and Hobbs, 2017) lays out a formal theory of commonsense psychology and how people assume others think, while (Lake et al., 2017) put forward their argument as to how we can go about building machines that learn and think like people. There are several older noteworthy works like Cyc (Lenat, 1995) and Ordinal common-sense inference (Zhang et al., 2017).

2.6 Data Collection Efforts

Collecting useful data is vital to the success of any AI or computational system and the collection and curation of a dataset by itself is considered a significant contribution in the field. So unsurprisingly there have also been several efforts to facilitate Question Answering (QA) systems and commonsense knowledge by building datasets. The seminal work in this was by (Sap et al., 2019c), which introduces a benchmark dataset for social and emotional commonsense reasoning consisting of almost 45k multiple-choice items. The MCTest dataset (Richardson et al., 2013) consists of a total of 500 stories and 2000 multiple-choice reading comprehension questions that were targeted at 7 year olds.

Similarly, Sap et al. (2019c) introduces a benchmark dataset for social and emotional commonsense reasoning and the MCTest dataset (Richardson et al., 2013), which consists of multiple-choice reading comprehension questions comprised of short (150–300 words) fictional stories which were targeted at seven-year olds.

Another prominent line of work is on building and analyzing (Boratko et al., 2018) the AI2 Reasoning Challenge (ARC) (Clark et al., 2018). ARC consisted of a dataset of almost $8,000$ science questions in English, and consisted in part a set of questions that neither a retrieval-based algorithm nor a word co-occurrence algorithm were able to answer correctly. The AI2 Reasoning Challenge (ARC) (Clark et al., 2018) is another major work in this field. ARC consisted of a dataset of almost 8,000 science questions in English. This dataset was split into the *Easy* set and the *Challenge* set; the *Challenge* set consisted of questions that neither a retrieval-based algorithm nor a word co-occurrence algorithm were able to answer correctly. Later, Boratko et al. (2018) more precisely analyzed the ARC knowledge, defining seven knowledge types and nine reasoning types, as well as triple annotating 192 ARC questions.

There are several other commonsense challenge datasets that are useful, like COPA: 84.4% - 1000 items (Roemmele et al., 2011), original Winograd: 72.9% - 150 items (Levesque et al., 2012), extended Winograd: 86.1% - 943 items (Rahman and Ng, 2012) and CommonsenseQA - 12k multiple choice questions (Talmor et al., 2018). Several other datasets of note are the 3rd & 6th grade reading comprehension (Hirschman et al., 1999), NewsQA: 10k news articles (Trischler et al., 2016), Search QA: 140K QA pairs (Dunn et al., 2017), TriviaQA: 650K QA pairs with evidence (Joshi et al., 2017), AI2 Science Questions: 1k multiple choice questions (Clark, 2015), SciQ Dataset: 13,679 multiple choice science questions (Welbl et al., 2017), and CNN/DailyMail dataset (Hermann et al., 2015). The volume of work in this area indicates that it is a major area of interest in the field.

CHAPTER 3

MOTIF UNDERSTANDING AS A FUNCTION OF GROUP MEMBERSHIP

3.1 Motivation and Background

I have established from previous chapters that motifs are well-grounded in folktales or other such sources, contain a wealth of information, and seem to be heavily culture-specific. But we do not yet know if the meanings behind motifs are understood differently by people from the cultural group that spawned the motifs versus the out-group people, so to speak. If we are able to determine this, it would mean that understanding motifs would be a great step towards unlocking the meaning of these highly dense information sources for a much wider audience than would ever have been possible.

The main motivation behind this study, therefore, was to determine how motifs are understood in real life. To do this, I designed an experiment to look at motif understanding as a function of group membership. The goal was to examine how recognizable motifs are within cultural groups? (How similar are inferences drawn by members of the same cultural group with knowledge of the motif vs. dissimilar to those from a different group without knowledge and/or of different interpretation of the motif) Simply put, do people who are *in-culture* understand motifs better than people who are *out-culture*?

One of the approaches I use in this experiment is called the Situation Awareness Global Assessment Technique (SAGAT) system. Endsley and Garland (2000) It was originally developed for cockpit UI assessment. It involves pausing a simulation, hiding screens and interjecting questions about current awareness and predicted future states (of world, of ship, etc.) There are three levels of situation awareness (SA):

- **Level 1 SA: Perception of Data**

  (i) What's going on and what the agent is trying to achieve?

  (ii) Purpose, Process, Performance

- **Level 2 SA: Comprehension of Meaning**

  (i) Why does the agent do it?

  (ii) Belief and Purpose

- **Level 3 SA: Projection of the Near Future**

  (i) What should the operator expect to happen?

  (ii) Projection to End State

  (iii) Potential Limitations

The more detailed explanation of how SAGAT was adopted for this experiment is described later in the chapter, in §3.2.4.

3.2 Experiment Design

For the actual experiment, I start with the claim that the following research question mentioned above can be answered using the Personality Attribute paradigm and the SAGAT paradigm modified for motif understanding. By comparing the attribute rating and SAGAT responses from motif-culture versus non-motif culture(s) participants, we can determine if there is a significant difference between motif understanding for those in-culture and out-culture.

3.2.1 Choice of target cultures

There were two main factors to consider while choosing the target cultures: (i) easy access to target population to conduct the study, and (ii) having a readily available digital motif index to work with. As with most work in motifs, the main base for the work was Thompson's Motif Index (Thompson, 1960), which in turn linked to the individual motif indices of other cultures.

For (i), we [1] focused on large population groups either in Miami, where FIU is located, or in Minneapolis or Boston, where our external collaborator's offices were located. We explored a wide range of options, and decided that we would choose Puerto Rican, Irish, and Jewish as the three target cultures. (This eventually ended up being a moot point as Covid-19 forces us to conduct the whole study online. The switch from in-person to online had no other impact on the study.) All three also have their own motif indices listed by Thompson, which fulfilled our criteria for (ii).

3.2.2 Choice of motifs

To choose the motifs, we first looked at the individual motif indices for each of the three target cultures. For Irish culture, we used T.P. Cross' "Motif-Index of Early Irish Literature" (Cross, 1952) as a main source. For Puerto Rican, we drew motifs from S.R. Lamarche's "The Mythology and Religion of the Tainos" (Hurley et al., 2021), R.E. Alegría's "The Three Wishes: A Collection of Puerto Rican Folktales" (Alegria et al., 1969), and J. Ramírez-Rivera's "Puerto Rican Tales: Legends of Spanish Colonial Times" (Ramírez-Rivera et al., 1977). For Jewish motifs, we drew on D.N. Noy's "Motif-index of Talmudic-Midrashic literature" (Noy, 1954).

I will not go into the details of the process of choosing individual motifs, as it was joint work with my colleague as mentioned earlier and therefore not part of this dissertation. The three main factors considered were: (i) commonly used, (ii) having a source within the cultural group, and (iii) having a high potential strength rather than just being used as a reference. Overall, a total of 38 motifs were chosen. The list of all the motifs used is shown in Table 3.1

[1] This choice of target cultures and motifs were done for the overarching project and not just these experiments I conducted. So there were several other people involved in this decision, primarily my fellow PhD candidate W Victor H Yarlott.

IRISH	The Salmon of Wisdom, Finn McCool, leprechaun, King Conchobar, Aos si, Banshee, Cu Chulainn, The wren, The magic harp, Tir na nog, Shamrock, Fairy fort, and The Children of Lir.
JEWISH	Haman, Golem, Amalek, Tower of Babel, Leviathan/Behemoth, 70 languages, Name in vain, Milk with meat, The ark of the Covenant, and Kiddush.
PUERTO RICAN	Reyes Magos/Three Kings, Agueybana, Atabey, Roberto Cofresi, Divina Providencia, Guanina, Juan Bobo, Yocahu, The coqui, Hormigueros, Jibarito, Guaraguao, Pitirre, Chupacabra, and Pava

Table 3.1: List of motifs from the different cultures used for this study

3.2.3 Self identification questions

In order to be able to test the hypothesis that people in-culture and out-culture perceive motifs differently, I need to be able to assess who are the ”in-culture” participants and who are the ”out-culture” participants for each of our three target cultures. To do this, I came up with a self-identification questionnaire to precede the main questionnaire of the survey. I wanted to know about a person's cultural background because I wanted to to see if there is a correlation between how strongly a person identifies with a culture (Jewish, Irish, Puerto Rican), and how well they know the motifs from said culture.

For this, I consulted psychologists, anthropologists, and ethno-linguists, as well as a review of relevant fields (Oppenheim, 1992; Bourhis et al., 1981; Bourhis and Sachdev, 1984; Clark and Barrows, 1981; Yagmur et al., 1999), such as (i) Ethno-linguistic Vitality, (ii) Communities of Practice, and (iii) Social Network Theory. I developed a survey based heavily on *ethno-linguistic vitality* with the help of an existing survey (Clark and Barrows, 1981) that focuses on the health of specific languages within communities and how those languages persist or languish. Ethno-linguistic vitality has many previously developed questionnaires that align with information we want to know. *Communities of practice* was not used due to its focus on teaching rather than groups with a cultural aspect. *Social Network Theory* was not used due to the degree to which it was abstracted from the problem: while potentially useful, I needed examples of what to ask and how to ask, which is readily available in ethno-linguistic vitality literature. The survey focuses on five main areas:

1. **Ethnicity, Race, and Cultural Group** Standard self-reporting questions on cultural group identity and country of origin and birth.
2. **Caregiver** Questions about the cultural identity and country of origin of the participant's primary caretaker.
   Included due to our discovery that parents of friends who are part of a cultural group are often from the country in which motifs originate
3. **Native language** Questions about the participant's and caretakers' native language, as well as the language used with friends, family, and neighbors.
   Included as languages are sometimes associated with cultural groups
4. **Religion** Questions about the participant's, caretakers', and participant's friends' religion. Like languages, religions are sometimes associated with cultural groups.
5. **Social interactions and media** Questions about interaction with others in-/out-group as well as consumption of culture-related media.

Please see Appendix A for the complete set of self-identification questions used.

3.2.4 Main questionnaire

The main questionnaire had two components: attribute questions and Situation Awareness (SA) questions. The attribute questions involved providing the participant with a text prompt – either with motif, no motif, or a combination of both – and have them answer

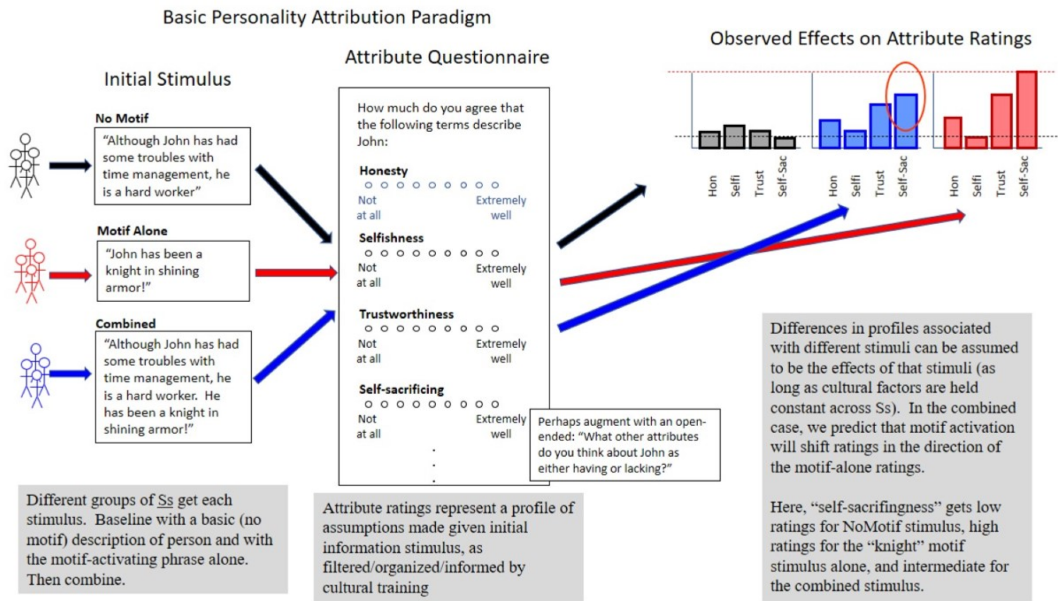


Figure 3.1: Demonstration of how the Basic Personality Attribute Paradigm works, with example questions and predictions shown.

how likely they think an attribute is aligned to the said motif. This concept is illustrated in Figure 3.1.

The second part of the main questionnaire was designed based on a modified SAGAT system, i.e. a SAGAT-adapted paradigm for Motif situational awareness. For this, we[2] created a story-based scenario invoking or not invoking a motif. We show that story to the participants with sufficient time to read and think a bit about it. We then ask questions about the story from SA levels, with an emphasis on motif-pertinent and motif–informed elements. Let us consider the following hypothetical (brief, but maybe not overly so) example:

No Motif: "John is very poor and his wife is ill and may die if she doesn't get an expensive medicine. John sees no way to get the money, so he asks a man he knows,

[2]For this specific portion of the design, our collaborator Dr. Christopher Miller provided with guidance and ideas, and I designed the questions together with my colleagues Victor and Diana Gomez.

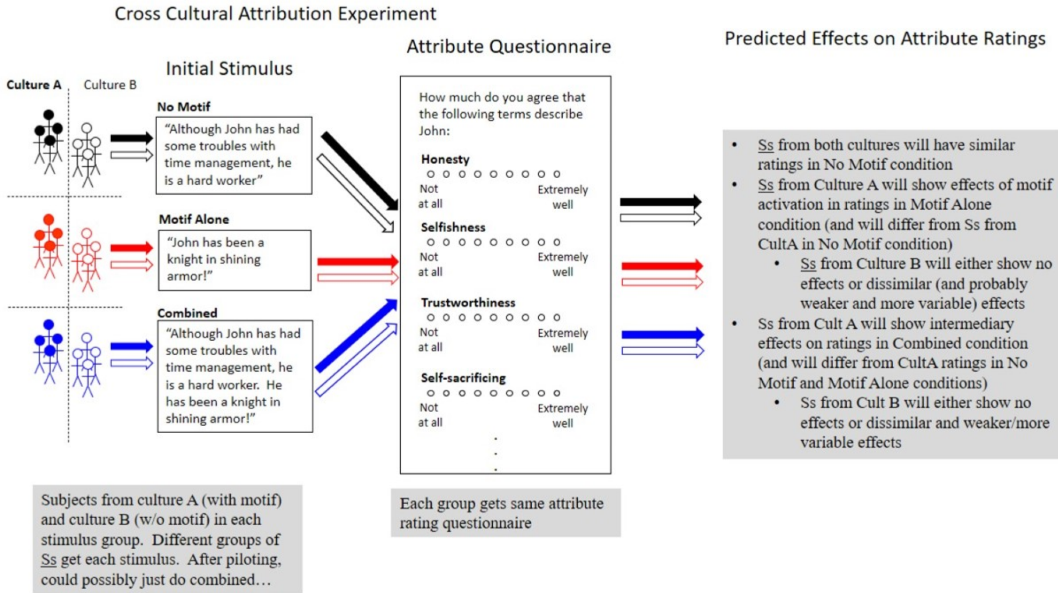


Figure 3.2: Demonstration of Basic Personality Attribute Experiment

George, to help him. George lives in his neighborhood, but is quiet and not very sociable."

Motif: "Everyone says George is a knight in shining armor."

Combined: Combine the above two prompts

The prompt text is followed up with questions (in the form of yes/no and a Likert scale to aid in quantitative analysis) from three SA levels, as such:

- Level 1 SA: Perception of Data (factual questions about the story)
  - Do George and John live near each other?
  - Is George friendly with many people?
- Level 2 SA: Comprehension of Meaning (deductions from story)
  - Does John know George well?
  - What does John want George to do?
- Level 3 SA: Projection of the Near Future

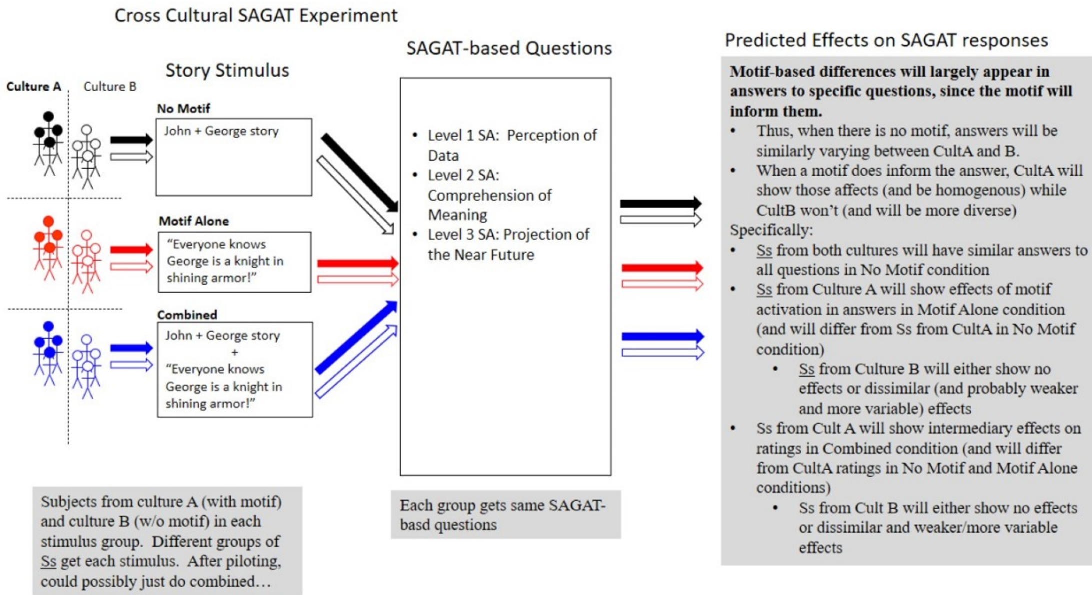


Figure 3.3: Cross Cultural SAGAT Experiment

– What do you think George will do?

Following these techniques, we created questionnaire for each of the target motifs. For creating the prompt text for each motif, we consulted external in-culture informants at every step of the process, and extensively followed their guidance in creating both the prompt texts as well as candidate answers provided to the participants. Let us look at an example story, with some questions that might be asked of them:

Example of a story with the motif: The principal of a school walks into a classroom to pick out students for a news interview. The principal wants the best, good-looking students to represent the school well. When the principal goes to pick out Jimmy for the interview, another student proclaims "No! Jimmy is the Juan Bobo of the classroom."

Example of an attribute question: Based on the passage above, how much would you think each of the following terms could be used to describe Jimmy: Intelligent [Likert Scale 1-9 from "Not at all" to "Extremely well"]

Example of an inference question: Based on the passage above, how would you answer the following questions:

- Do you think Jimmy is a comedic character? [Likert Scale 1-9 from "Not at all" to "Definitely"
- How confident are you in your answer? [Likert Scale 1-9 from "Not at all" to "Extremely"]

Based on the answers for these questions (the actual survey contained many more questions, all of which are submitted as part of the other deliverable), we are able to see several effects. The technique used to analyze these results are shown in Figures 3.2 and 3.3. In addition to this, I designed the survey such that each participant gets their own in-culture questions, and those of one other culture. It is balanced out so we have an equal number of responses for all cultures, both in and out culture ones. I also estimated the time required to take the surveys by having several colleagues and friends take the survey.

Please see Appendix B for all the questionnaires used for the main section of the survey.

3.2.5 Sample Size

To ensure we had sufficient number of participants for the experiment, I calculated the required sample size. Israel (1992) To do this, I generated sample responses from the sample experimental design questions from the different types of participants I anticipated. I accounted for the fact that there would be differences in values I got from the sample responses between In-Culture and Out of Culture due to out of culture participants not being familiar with the motifs. With this process, I was able to determine a sample size of 21, and I chose to keep some margin for myself and chose 30 as the participant size for our pilot. For the full experiment, however, I decided to have a larger scale and made it 40 participants per culture for a total of 120 participants.

3.3 Methodology

3.3.1 IRB Approvals

Because this experiment involved human subjects, before I conducted any part of it, I got approval from FIU IRB (Institutional Review Board). After that I got approval from HRPO of US Air Force Research Laboratory (AFRL), which was required as they were the source for the funding for this project. The experiment was classified as exempt research by both regulatory bodies. No personally identifiable information (PII) was collected from the participants for the experiment. I and everyone else in the team also went through the Citi Training for Human Subject Research as required by the FIU IRB.

The FIU IRB approval was obtained on November 1st 2019. The IRB Protocol Exemption number was IRB-19-0381. I got the subsequent HRPO approval on April 23rd 2020, FWR2020065Xe. With each change in the instruments, language of questionnaire, and other stuff, I got amendment approval from both FIU and HRPO. Following the guidelines of both FIU and AFRL, some amendments only required amendment approval from the FIU side. In total, I got four amendment approvals from FIU (03/13/2020, 04/22/2020, 10/23/2020 & 01/12/2021) and two from AFRL (06/30/2022 & 10/30/2021) for this main phase experiment.

3.3.2 Participant Recruitment

Recruitment was done by flyers and emails, targeting adults (18 years or older) of the three target cultures (Puerto Rican, Jewish, and Irish). The primary location for recruitment was

FIU premises, although eventually other participants were recruited via personal contacts of the researchers who identified as part of the target culture, which is a valid method used by other surveys before, called Snowball Sampling. (Goodman, 1961) For the full experiment, extensive recruitment was done by contacting various clubs, organizations, and social groups, both in-person and online.

3.3.3 Compensation

The participants were compensated for their time via Amazon® gift cards. While generic Visa® or MasterCard® gift cards that can be used on any platform were preferred, FIU regulations did not allow for this. This decision was based solely for convenience and there was no benefit to us for choosing Amazon as the method of payment, nor is it an endorsement of Amazon or any of its services.

3.4 Pilot Study

The pilot study was planned to be conducted at FIU premises. Before the actual pilot study, we (me alongside other FIU students) performed a dry-run of the pilot experiment in the Cognac Lab at FIU, and overseen by Dr. Miller. The dry-run followed a draft script. Other members of the Cognac lab working under Dr. Finlayson helped by acting as participants with the dry-run. This dry-run went fairly smoothly, and there was a round of discussions following the dry-run. A final script was then created and agreed upon to be used for the actual survey.

However, like the entire rest of the world, we also had to adapt to the pandemic. So we switched the study to be conducted online instead. We set up an online survey platform, hosted on Dr Finlayson's machine located at Florida International University premises so we did not have our data pass through third-party servers. We used a localized copy of an open source online survey tool, called Limesurvey. [3] The surveys given to people were done pseudo-randomly, [CITE] and were designed to ensure a similar number of participants get each of the non-motific, motific, and combined example. Real randomization was not done intentionally to ensure the balance in responses across different types of questions. (Hatch and Lazaraton, 1991; Rossi et al., 2013) For the pilot, we also planned to interview the participants and note their feedback to improve the survey. This was conducted through the Zoom video calling platform, which made it convenient to conduct the study with participants that were not in the same geographical location as I.

3.4.1 Second Pilot

I also discovered a few issues while analyzing the pilot data, such as issues with wording of some questions and some bad performing motifs. I also discovered an issue with Limesurvey that causes data to be imbalanced in some cases, resulting in questions that cannot be analyzed. In order to check the effects of the experiment, I ignored the data with errors and analyzed the remaining data. I saw that there appeared to be sizable effects for 'motif' and 'combined' conditions, compared to 'no motifs'. It was not as clear in all cases, but there were still errors to be fixed.

I then fixed the issue with survey balancing, and also changed the wording of some questions based on feedback from the participants. I also removed some problematic motifs based on feedback from in-culture participants, and added new motifs for Puerto Rican culture. I then created and tested the surveys internally within the ACUMEN and Cognac teams [4] before making them live. Finally, I ran the second pilot for 30 participants.

[3] My undergraduate research assistant Diego Castro Estrada helped with setting up the survey on the server.

[4] These are teams within and outside FIU that were involved in the overarching project that included this experiment.

3.5 Main Phase

The initial results indicated there might still be some minor errors in the process. Upon analysis and based on user feedback, I discovered that the problem was that some of the prompts and/or attributes used for the survey were not the most suitable, and that this problem was localized to three specific motifs: two from Jewish and one from Irish. I removed the problematic motif from Irish, and changed an attribute each for the two Jewish motifs. I also appended some additional feedback question to the end of the survey to elicit motif associations from the participants. This was not directly related to the objective of this experiment, but rather was done with a hope to get extra information from in-culture participants. I also forewent the Zoom live interviews for the main phase that I did for the pilot, as that was mainly done to collect feedback on the survey design and mechanism.

After all the corrections, I ran the main phase with the goal of 40 participants each for the three surveys for a total of 120 participants. I paused the survey once we reached 10% to make sure we were on the right track and were getting expected responses. The initial analysis verified that the survey was going smoothly as expected, with no discernible errors or flaws, and discovered no significant problems. So I continued on with the participant recruitment.

3.6 Results

To analyze the effect, I used Cohen's d (Cohen, 1988) to measure the effect size. Cohen's d is a measure of effect size: the strength of the difference between groups. I observed that there are several results that indicate there is a significant effect (as defined by Cohen, effect size > 0.2).

The main phase of this study was conducted for all 120 participants as planned. In the analysis, I focus on the effect size (measured by Cohen's d) of the attribute and inference questions for the motif, no motif, and combined group between in-culture and out-culture groups.

3.6.1 Results by Culture

First let's look at the results based on the three cultures. The results of this analysis are displayed in Figure 3.4. We can see that for Jewish and Irish cultures, for both attribute and inference questions, in-culture response clearly differed from out-culture ones for motif-based prompts than no-motif ones, indicating that in-culture people understand motifs in text more clearly than out-culture people. We do also observe some effect for the *combined* prompt which tends to differ for different motifs, which I had not expected. While I can not know for certain why this is without further experiments, one possible explanation is that more context and explanation makes the meaning of some of the motifs clearer, while in other cases our chosen context gave a little too much away, causing the out-culture people to also comprehend the motifs more clearly. However, these results do validate our initial hypothesis that motifs are recognizable for participants that are part of the target culture.

For the Puerto Rican motifs, however, we do not see the effects we hypothesized before the study. I suspect this might be because of two main reasons. Firstly. the Puerto Rican motifs were found in Spanish texts, which had to be translated to English for the study. Since motifs are so nuanced, it is possible that a lot of meaning was lost in this process.

Secondly, Puerto Rican motifs seemed to have blended in to the larger Hispanic and American culture and language, indicating that they might be more recognizable than the

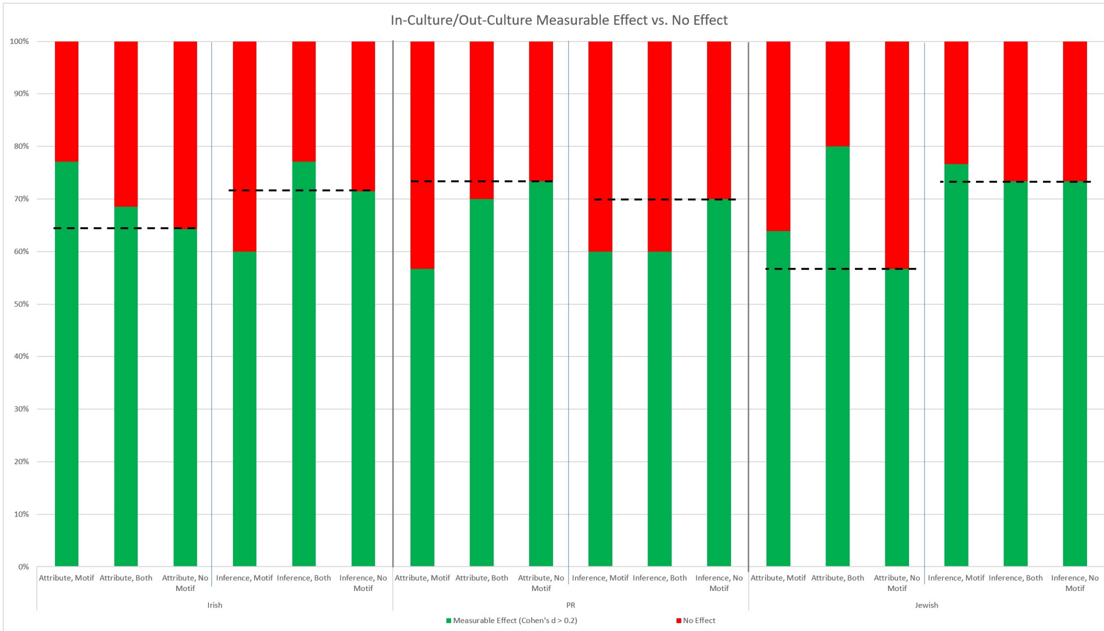


Figure 3.4: % of questions with at least a small effect vs. no effect, split by culture and prompt type.

other motifs. Indeed, my results from the other study (in §4.5) shows it to be true that Puerto Rican motifs are in fact more recognizable than Irish or Jewish ones.

In addition to the Cohen's d test, I performed the t test on each of the categories shown in **??**, comparing the responses of in-culture participants and out-culture participants for each of these categories. The results for the test are shown in table 3.2. The questions were considered to have statistically significant difference ($p<0.05$).

3.6.2 Results by Motif Type

Now let us look at the results split by the three types of motifs: Event, Character, and Prop. These results are displayed in Figure 3.5.

We see that for all cases except one – Inference questions for event motifs – either motif-only or combined prompts have a greater effect than no-motif prompts. This is consistent with our finding that motifs do tend to be more recognized by in-culture group

		Statistically Significant ($p < 0.05$)	Not Significant
Irish	Attribute, Motif	1	34
	Attribute, Both	2	34
	Attribute, No Motif	2	34
	Inference, Motif	3	32
	Inference, Both	2	33
	Inference, No Motif	4	31
PR	Attribute, Motif	3	27
	Attribute, Both	3	27
	Attribute, No Motif	3	27
	Inference, Motif	3	27
	Inference, Both	6	24
	Inference, No Motif	1	29
Jewish	Attribute, Motif	6	24
	Attribute, Both	2	28
	Attribute, No Motif	2	28
	Inference, Motif	9	21
	Inference, Motif	1	29
	Inference, Both	1	29

Table 3.2: % of questions with statistically significant difference, split by culture and prompt type.

more than out-culture. As I mentioned in the previous analysis, it is again unclear why the combined prompt behaves differently based on different motif types. But the behaviour is consistent even for motif-type analysis, so it makes it even more plausible that the post-hoc explanation I gave is true.

3.6.3 Discussion

Based on the results of the experiment, we can conclude that motifs do seem to carry noticeably more implications for in-culture people. However, as shown by the Puerto Rican motifs, while we can say that it's true for *most* motifs, it's not true for *all* motifs, especially ones that are more recognizable by out-culture people.

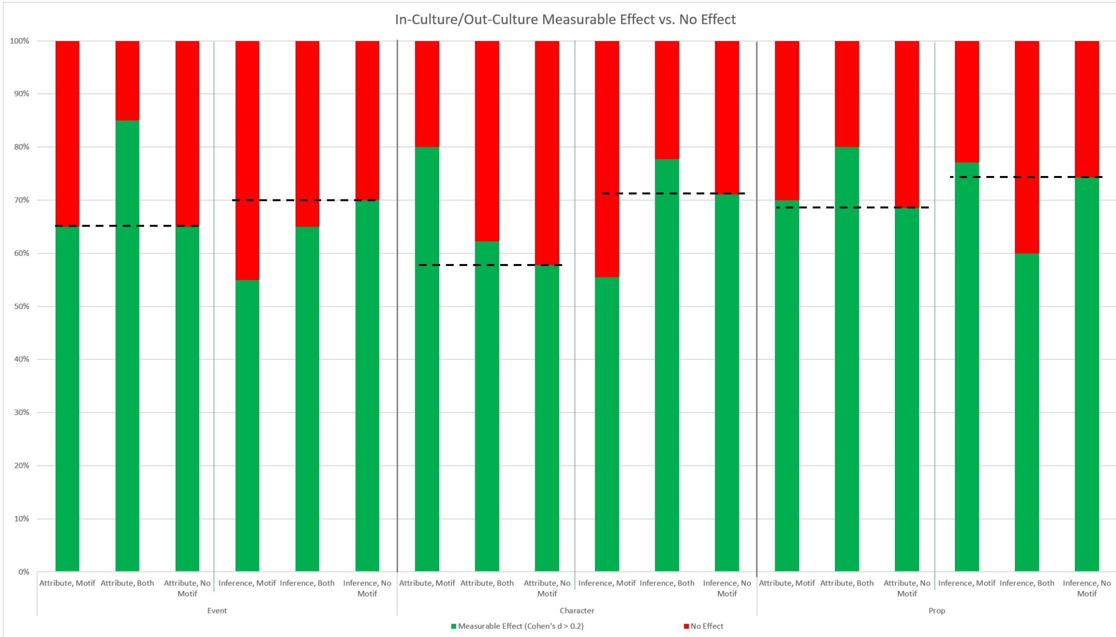


Figure 3.5: % of questions with at least a small effect vs. no effect, split by motif type and prompt type.

We also see that most of the questions do not meet the threshold of statistical significance in this case, as shown by the results of the t-test. My plan is conduct further studies in future works to try and remedy this.

We also see that the task of studying motifs is incredibly difficult, and great care should be taken in the task of choosing the appropriate motif to ensure that the motifs are not too obscure for the target culture, while also ensuring they're not well-known outside the said culture. This is compounded by the fact that I am not an in-culture person for either of the target cultures, making my own insight impartial but limited. Another challenge I faced during this study is for recruiting participants. Recruiting participants from a specific cultural group online proved quite a challenge, especially for an experiment like this one that needs high quality text responses. This is another thing to be mindful of for future studies of similar nature.

CHAPTER 4

MOTIF RECOGNIZABILITY IN TEXTS

4.1 Motivation and Background

With this study, I am primarily trying to determine if motifs are recognizable in texts. Secondarily, I am trying to see what associations people perceive the motifs to have when they have / do not have understanding of the motifs.

In addition to the study I conducted, I also considered several other alternatives, all of which had merit, but were not chosen due to various reasons. These alternatives considered are listed below:

1. **Elicit associations for selected motifs from in-culture participants** For this study, we could use either free text questions ("what comes to mind when you hear...") or a more structured questionnaire, similar to my ritual work style questionnaire. (See (Acharya et al., 2021) and §5.2.4). The trade-off for this is that for the former, we risk not getting useful answers, while for the latter we might be asking too many leading questions, which is shown to be bad when conducting surveys. (Oppenheim, 1992; Brace, 2018)
2. **For in-culture participants, ask for motif "synonyms" either whole or partial** This is similar to the first one, but even more narrow, where we try to get other words that are semantically similar to the motifs. Again, this could be either generative or recognized in multiple choice questions.
3. **Use a multi-step process with different groups re-tested at varying time lengths** For this idea, we train on one or more known-to-be novel motifs in session 1 (either because they are very rare in this in-culture, or because we invent them). We also test in session 1 to establish baseline for associations. Then, we re-Test in session

2 after various different inter-test intervals (e.g., 1 week, 2 weeks, 1 month) to establish durability of associations. We also have the option of exploring different training methods (e.g., explicit vs. embedded vs. no context). While this proposed study is very promising, it would require significant amount of time and effort. However, this is an interesting idea I hope to explore in the future.

4.2 Methodology

The experiment design was very similar to the Motif Understanding experiment, differing only when needed; please see §3.2 for details.

4.2.1 Choice of Motifs

From the larger list of motifs I had from the three target cultures, 21 of them were chosen. The main factor for this was to choose motifs for I had examples in real life text that I could use as prompt for the study. I refrained from creating my own prompts for the study to ensure participants were given text where motifs were naturally used to avoid any accidental error or bias that might arise from me – an out-culture person for all three target groups – using the motifs in text. The list of all motifs used for this study are shown in table 4.1.

IRISH	Finn McCool, Leprechaun, Cu Chulainn, Tir na nog, Shamrock, and The Children of Lir.
JEWISH	Haman, Golem, Amalek, Tower of Babel, Leviathan, Behemoth, The Ark of the Covenant, and Kiddush.
PUERTO RICAN	Reyes Magos/Three Kings, Atabey, Divina Providencia, Juan Bobo, Yocahu, Coqui, and Chupacabra

Table 4.1: List of motifs from the different cultures used for this study

4.2.2 Self identification questions

The Self ID section of this study's questionnaire is very similar to the one in the Motif Understanding experiment (see chapter 3). The most important changes were the removal of the religion and caregiver questions, and the addition of a question to record a Prolific ID for payment purposes. The reason for removing the religion and caregiver questions is that, because the survey was given to participants of cultures other than the one's whose motifs were present in the survey, such a distinction would not make for very useful data.

4.2.3 Main questionnaire

For the main part of this study, I designed a questionnaire to elicit information that can answer our question about whether motifs are recognizable in texts, even to those which do not understand the meaning of the text. Firstly I offer the participants a piece of text containing a motif. Unlike the last study where we created the text prompt by ourselves for the experiment, for this study I used texts that contain motif found in real life, with an aim to test our hypothesis on real text rather than the ones we wrote. I then ask the participants to answer several questions about the text, as such:

- Ask about meaning of text
- Ask if any unknown section - is there something you don't understand?
- Ask for unknown word speculations – what do you think this could mean?
- Ask for what information they would like to know to determine meaning (e.g., used in this kind of setting, used by this kind of person, etc.)
- Ask specifically about motif meaning - does this word mean anything to you?
- Ask for motif associations

Using these ideas, I created a questionnaire for each prompt containing a motif. Unlike the previous study, where each motif only had one set of prompts, in this study I included anywhere from 1 to 5 prompts for each motif based on data availability. The reasoning behind this was to make sure I minimized the chance of people learning the meaning of motifs just by the context of the sentence alone; even if one of the sentence happens to give the meaning away by context, it is less likely that this will happen for all the sentences. After the questionnaires were completed, we [1] got input from the respective in-culture informants and modified the survey based on the feedback. We were particularly sensitive to the wording of the questions to ensure we were not giving any meaning away or leading the participants in any way.

4.2.4 Baseline

In order to be able to properly assess the results of the study, it was important to have a proper baseline for this task. To do this, I asked the same questions as above, but with literal instances of motifs not used in a motific way, i.e. motifs used in such a way that they do not invoke the meaning of the motifs. The idea behind choosing this as a baseline rather than just simple English comprehension was to make sure the actual words themselves that form the motif being new to the participants do not affect our study. By making sure that baseline prompt also included motifs, just in a non motific way, the only thing different between the prompts is that in the main questions the motifs invoke their meaning. This way I can truly test if motifs are indeed recognizable for *being motifs*, rather than just because they're new words unknown to the participants.

Table 4.2 shows an example of prompts used for the study. To see all the prompts used for this study, see Appendix C.

[1] I had help from colleague Victor in reaching out to cultural informants.

Motific Prompt	Whether the Nazis were **Amalek**, and I believe they were in the last generation, or Iran is **Amalek** today, and I believe they are, is not the point. The Jews have had many enemies over the millennia. Our enemies all wanted to destroy every one of us.
Baseline Prompt	"Defeat Of **Amalek**" from the album "Not Dead Yet" has one of my favorite metal riffs.
Motific Prompt	I do enjoy the odd holiday abroad and I 'm not ashamed to admit I have a sports car that drinks petrol like Shane McGowan would down a pint of Guinness. I have to be honest that my carbon footprint would be the size of **Finn McCool**'s right foot."
Baseline Prompt	Find out what happened when Gordon Ramsay visited the **Finn McCool** pub and read about when and why **Finn McCool**'s closed.

Table 4.2: Examples of the motif prompt and baseline prompt used in the study, shown here for the motifs Amalek and Fin McCool.

4.2.5 Sample Size

Using the same process as in the previous study (§3.2.5) , I was able to determine a sample size of 112 to see an effect, and I chose to keep some margin for myself and chose 120 as the participant size for the study. Please see §3.2.5 for more details.

4.3 Methodology

4.3.1 IRB Approvals

Similar to the previous study, I got approval for this study from both the FIU IRB (Institutional Review Board) as well as the HRPO of the AFRL. For this study, approval was requested as an amendment of the original research. No personally identifiable information (PII) was collected from the participants in this study either. The study was also classified as exempt research – like the other experiment – by both the regulatory bodies. Everybody in the team already had their mandatory human subject trainings valid. The FIU IRB amendment approval for this study was obtained on 23 September 2021. I got the subsequent AFRL approval on 5 October 2021. The approval numbers for both were same as that for the previous experiment.

4.3.2 Participant Recruitment

For this study, I decided to use a crowdsourcing platform to recruit participants to get an easier access to the target population. After conducting some research, we [2] came to the conclusion that Prolific is the optimal choice. Prolific has higher naivety, lower dishonesty and cheating rate, and a more diverse population (Peer et al., 2017) than MTurk, all of which increase the chance of collecting better, more sincere responses. Although CrowdFlower has the highest diversity and education in comparison to MTurk and Prolific, their population's English reading levels were lower (which may be significant in a survey relying on figurative language) (Peer et al., 2017). Furthermore, Prolific is cheaper for us to use than MTurk and has a more flexible pre-screening process. Other alternatives other than MTurk, Prolific, and CrowdFlower were considered, but Prolific still seemed like the strongest choice also due to their active population size.

4.3.3 Compensation

Since the participants were recruited on Prolific, they were compensated using the platform's native payment system.

4.4 Main Phase

The actual survey was once again hosted on Limesurvey, and the participants were directed to the platform from Prolific to take the survey. As in the previous study, this

[2] Diana Gomez assisted with evaluating the different platforms.

study was paused after collecting 10% of the responses to see if everything was running smoothly. After verifying that it was, the survey was resumed and eventually completed.

4.5 Results

I collected responses from 120 participants, with participants answering questions about 67 different prompts of 21 different motifs. The results, grouped by the target cultures, are shown in Table 4.3, while the full results for each motif are shown in Table 4.4.

Overall, we can see that 53% of participants said they didn't understand the motific text, compared to 38% for the literal, non-motific use (baseline). Similarly, 32% of the participants said they would be able to understand the text better if they knew the meaning of the motif, compared to just 20% for the baseline for the same. I performed the Binomial Test and found that the difference in the responses were statistically significant for the overall study, as well as for the Jewish and Irish cultures, but not for the PR culture.

We can also see from Table 4.3 that for all three cultures, participants selected motific usage more frequently as the span of the text they didn't understand the most and needed more information about than the baseline. We do also see that Puerto Rican motifs, while still surpassing the baseline, performed the poorest of the three. This is even more evident when we look at the motif-by-motif breakdown, where we can see that half of the Puerto Rican motifs perform below the baseline.

These results prove our original hypothesis that motif are recognizable: i.e., even if people might not understand the meaning of motifs, they are able to recognize that there is something they do not recognize about the motifs, and that they are missing some information about them.

In addition to this, this study also shows that different motifs seem to have different levels of recognizability. We see in the results that different motifs were selected at different rates. Even though most motifs were selected more frequently when they were used motifically, some motifs did not exhibit such a clear effect or, indeed, any effect at all. In particular, Puerto Rican motifs exhibited the least pronounced difference between the selection rate for the motifs being used literally and the motifs being used motifically.

Culture	Unrecognized	Baseline	Info Needed	Baseline
IR	62.49	46.84	39.25	23.89
JW	59.59	30.87	33.91	17.21
PR	38.21	36.07	24.05	19.90
All	**53.12**	**38.28**	**32.33**	**20.49**

Table 4.3: Motif Recognizability study results grouped by culture.

4.5.1 Error Analysis & Discussion

The results of this study are consistent with my original hypothesis that most motifs, while not understandable by everyone, especially outside the origin cultural group, are nonetheless recognizable in text as carrying additional information. It is also evident by the results that this is not merely due to motifs being unique words or phrases, but rather due to their usage and the meaning they evoke.

In terms of an explanation of why some motifs performed poorly, I believe this is because these motifs were already reasonably well known outside of their origin cultures because their literal form consists of words that are very familiar to an average speaker. If we look at most the weakest performing motifs in this study – Coqui, Tower of Babel, Our Lady of Providence – they are all motifs that are made up of words that are used in everyday life. So it is possible that participants did not select these motifs as the unknown span of text because they were so familiar with the words. This, of course, is very hard to test to know for sure.

Culture	Motif	Unrecognized	Baseline	Info Needed	Baseline
IR	CUC	63.79	49.18	43.10	21.31
	FIN	70.69	57.38	39.66	14.75
	GOL	53.23	34.43	56.45	19.67
	LEP	54.84	57.38	35.48	42.62
	SHA	33.87	39.34	27.42	18.03
	SWA	70.69	62.30	25.86	45.90
	TIR	90.32	27.87	46.77	4.92
JW	AMA	70.69	24.59	43.1	8.2
	BEH	44.83	4.92	22.41	3.28
	HAM	66.13	31.15	24.19	8.2
	KID	88.71	70.49	53.23	44.26
	LEV	61.29	14.75	46.77	4.92
	TOW	25.86	39.34	13.79	34.43
PR	ATA	25.86	36.07	13.79	22.95
	CHU	18.97	34.43	18.97	13.11
	COQ	55.17	65.57	41.38	31.15
	DIV	20.69	39.34	10.34	13.11
	JUA	45.16	21.31	32.26	13.11
	KIN	53.23	45.90	30.65	40.98
	YOH	48.39	9.84	20.97	4.92

Table 4.4: Full results of the Motif Recognizability study for each motif, showing the percent of participants that selected motif as the unknown part of the text and the part that they would need more information about to understand, with their corresponding baselines.

Another possible explanation for the poor performance of the Puerto Rican motifs might be that because these had to be translated from the original Spanish, they might have lost their nuanced meaning to some extent. These minor poor results notwithstanding, however, the bigger point still stands that motifs are recognizable in texts.

4.6 Limitations and Future Work

Most of the limitations come from the inherent nature of the motifs we are studying: as they are highly specific within cultures and not well known to out-groups, the information generally available about them is also limited. We are always limited by the information we have, or lack thereof. Because I also do not have social scientists or cultural anthropologists working with me or in the team, I can only rely on existing lists of curated motifs. Furthermore, I am also limited in the amount of information available about these known motifs themselves.

One future plan is to scale this work. We had the idea to do this experiment in two phases, with time to educate a portion of people in between the phases. But this would be a years-long study. We were short on both time and funds to do that. But if that sort of experiment were done, the output we get could have a wealth of potential in understanding motifs better.

CHAPTER 5

RITUAL SURVEY

5.1 Motivation and Background

There has been a lot of work to try and inject commonsense knowledge in AI systems (See §2.5). A large body of recent work has focused on the creation, curation, and use of large-scale commonsense knowledge bases and knowledge graphs (Sap et al., 2019b; Bosselut et al., 2019). Importantly, these types of knowledge acquisition efforts have a long history in and have been of great use to a wide variety of AI systems (Shi et al., 2017; Olteanu et al., 2017; Sap et al., 2019a; Liu et al., 2020).

One glaring omission in all of this prior work has been the lack of focus on context-contingent aspects of commonsense knowledge; that is, most prior work views commonsense as a universal monolith. While some events included in prior work are not variable across groups—-like reading a book or breaking a window, for instance—many events are variable, and here we focus on one highly relevant type of context-specific commonsense knowledge, namely cultural commonsense. Consisting of ritualistic, geographical, and social knowledge, cultural commonsense plays a large but hidden role in humans' day-to-day social interactions. For example, let us consider a very simple social setting: *You are invited to a wedding. How long do you expect to be gone for, and how many people do you think will be there?* For most people in the United States or the wider Western world, the answer would probably be a few hours; probably half a day, starting in the early afternoon; and somewhere around a 100 people. However, for many people in India, the obvious answer is that you will probably have to lay aside several days for the whole event, and anywhere between several hundred to over a thousand people will attend. Such socially-conditioned knowledge is inherently obvious to people from the respective cultures, and hints at the differences in commonsense knowledge across cultural and social settings, particularly when it comes to ritualistic practices.

So it is essential to have a way to encode this cultural knowledge. To do this, I use *rituals* as the markers that are commonly represented across cultures. There is more than sufficient evidence in literature to justify selecting rituals as a marker of culture that can be used to compare and contrast across cultural groups, as detailed in §2.2. The primary research question this work is trying answer is if the information about the variations in ritual practices that exist across cultures can be crowd-sourced, especially in a format that might be useful for computational systems.

I introduce an approach that extends prior work on crowdsourcing commonsense knowledge by incorporating differences in knowledge that are attributable to cultural or national groups. Specifically, I start by surveying the extensive prior literature on cultural knowledge and ritual practices, and select a short list of six rituals to focus on for our study. I demonstrate the technique by collecting commonsense knowledge that surrounds these six fairly universal rituals—birth, coming-of-age, marriage, funerals, new year, and birthdays—across two national groups: the United States and India. The study expands the different types of relationships identified by existing work in the field of commonsense reasoning for commonplace events, and uses these new types to gather information that distinguish the identity of the groups providing the knowledge. It also moves us a step closer towards building a machine that does not assume a rigid framework of universal (and likely Western-biased) commonsense knowledge, but rather has the ability to reason in a contextually and culturally sensitive way. The hope is that cultural knowledge of this sort will lead to more human-like performance in NLP tasks such as question answering (QA) and text understanding and generation. I begin this by improving upon an existing QA system, the details of which work is layed out in chapter 7.

5.2 Experiment Design

In this section, I outline the design of the survey for gathering the ritual-based cultural knowledge. I first describe the target cultures, followed by the details of our pilot experiment; I explain our data collection method on Amazon Mechanical Turk (MTurk); finally I describe the survey questionnaire that I used.

5.2.1 Selecting Rituals

Given the vast numbers of rituals that can fall into the six categories previously outlined, and the variance in the extents of their observance across cultures, another crucial decision is in selecting specific rituals as markers of cultural knowledge. There were several factors that we had to consider while making this decision. First and foremost, we needed activities whose identifying names are used in multiple cultures. For example, a ritual like *Passover* cannot be used since it is highly specific to Judaism and Jewish cultures. It would make no sense to ask *"What are your cultural practices with regards to Passover?"* of a Hindu or a Muslim. Furthermore, the more specific the names, the easier it is for a QA system to associate its knowledge to that activity. As an example, *"Catholic Christmas Mass"* is a highly denomination- and group-specific ritual, and will exhibit very little variation across cultures. Second, in order to ease the data collection process, the selected rituals needed to have fairly concise and telegraphic names. For instance, it is confusing to probe a participant in a study about *"the kinds of things you do before a sports game popular in your culture"*; even though this is an activity that is fairly widespread and at the same time variable across cultures. Instead, we are seeking rituals that can be described in just a few words and bring a very specific activity or event to mind. My choice of rituals is intended to primarily ease the collection of crowd-sourced data—I thus pick activities that may have different practices across different cultural groups but are likely to be found in all of them. I take guidance from the analysis of Bell (1997) and use the following six rituals as our target rituals in this work, as shown in table 5.1.

Ritual	Alternate Names	Category
Wedding	wed, marriage, marry, matrimony, nuptials, wedlock, union, hymeneals	*Rite of passage*
Funeral	funerary, burial, cremation, interment, entombment, obsequy	*Rite of passage*
Coming of Age	becoming a man, becoming a woman, manhood, womanhood, adulthood	*Rite of passage*
Birth	childbirth, delivery, birthing, childbearing, parturition, nativity	*Rite of passage*
Birthday	name day, natal day	*Rite of passage*
New Year	N/A	*Calendrical rite*

Table 5.1: List of rituals used for the task, with their alternate names and categories.

These rituals all have the advantage that across cultural groups there are limited number of ways of naming or expressing them, and the meaning is evident to most subjects answering our survey.

5.2.2 Selecting Target Cultures

For this study, I focused on two specific target groups: Americans (people from the United States of America) and Indians (people from India). I chose these two groups for a variety of reasons. First, both the countries use English as one of their major languages, either officially or unofficially, and since our study was to be conducted in English, this was a key requirement. Second, Amazon Mechanical Turk has a high presence of workers from both these countries (Difallah et al., 2018). Since the bulk of the data collection was to be done via Amazon's Mechanical Turk platform, it was essential that I considered the demographics of the crowd-workers. Moreover, these groups allowed us to set up a unique contrast and high degree of cultural variation between the two groups. Apart from providing contrasts against each other, the fact that the United States and India are large and diverse countries consisting of various cultures allows us to capture a varied amount of data within the groups themselves.

The data was collected for three groups: US, India, and all others. For the preliminary data collection, I did it for just the first two groups, and for the large scale data collection, I also collected data for *others* category. While the initial plan was to use Philippines as the third cultural group, the idea was then discarded after the pilot collection as I was not able to recruit enough participants to be able to scale up the study. This is why I will only discuss the study for India and the United States in this chapter. However, I was able to use the data for the Philippines that I collected during the pilot work for improving a QA system (as described in chapter 7).

For the large scale data collection, despite the initial plan to have a distinct third cultural group, the demographics of Amazon Mechanical Turk proved insufficient in any individual culture to provide the volume of data that was needed. This is why I went for the *other* category.

5.2.3 Self identification questions

The survey questionnaire was divided into two sections: the self-identification questionnaire, and the main questionnaire.

The self-identification questions were similar to those used in the previous two studies, so please see §3.2.3 for more details. Portions of the questionnaire relevant to this current study were modified to suit my purposes, and the wording changed to be consistent with contemporary terminology, all the while ensuring these modifications did not alter the validity of the questionnaire. This self-identification questionnaire was further modified to prioritize the targeted countries with regard to language and religion. The options for these two fields were based on the most likely answers given the demographics of those populations. While for this work I have considered only nationality as the marker for culture, moving forward I would love to use a broader set of factors to define the group membership.

5.2.4 Main questionnaire

Event-specific questions

For the main questionnaire, a prompt specifying the ritual event under consideration was first shown, and the participants were asked a set of questions pertaining to the specific event; see the questions asked and the layout of the form in Figure 5.1 for details. The questions asked were the following:

1. Where does this event typically happen?
2. When does this event typically happen?
3. How long does this event typically last?
4. How many people typically participate in an event like this?
5. Who are the important people involved in this event? (Maximum 5)
6. Is one or more of the important people the focus of this event?
7. Who are the people who are the focus of this event? (check all that apply)

Question 7 only appeared if participants answered "Yes" to Question 6, and the list was automatically populated from the answers of Question 5.

Before the event
Does this person typically have an intent in causing the event? What is this person's typical intent in causing the event? Does this person typically need to do anything before this event? What does this person typically need to do before this event?
During the event
Does this person typically use something during the event? What things does this person typically use during this event?
After the event
How would this person be described as a consequence of the event? Does this person typically want to do something after this event? What does this person typically do after this event? What is the typical effect of the event on this person? What does this person typically feel after this event?

Table 5.2: Questions asked of the survey participants for each important person associated to an event in their responses.

Person-specific questions

After the event-specific questions, the participants were asked to answer questions that involved the specific persons mentioned in the events (as part of their responses). This part of the survey was adapted from ATOMIC (Sap et al., 2019b). Questions were divided into three temporal categories: before the event, during the event, and after the event. We considered four types of questions: (1) intent & reaction; (2) need & want; (3) effects; and (4) attributes. These four types evolved into 11 questions on the survey form, as shown in table 5.2. In terms of the presentation of the questions, the terms `PersonX` or `PersonY` were replaced by the actual names that participants provided (in the Person fields) in order to make the questions feel more natural to the survey participants. These person-specific questions were repeated for each person that the survey participant deemed “important” to a given event.

For the entire survey questionnaire, please see Appendix D.

5.3 IRB Approvals

As with the last two studies, because this survey also involved human subjects, before I conducted any part of it, I got approval from FIU IRB (Institutional Review Board). This study was classified as exempt research by the IRB. No personally identifiable information (PII) was collected from the participants for the survey. The FIU IRB approval was obtained on 4 March 4 2020, and the reference number was 108783. In order to scale the survey up for the large data collection, I got an amendment approval from the IRB. This approval was obtained on 21 January 2022.

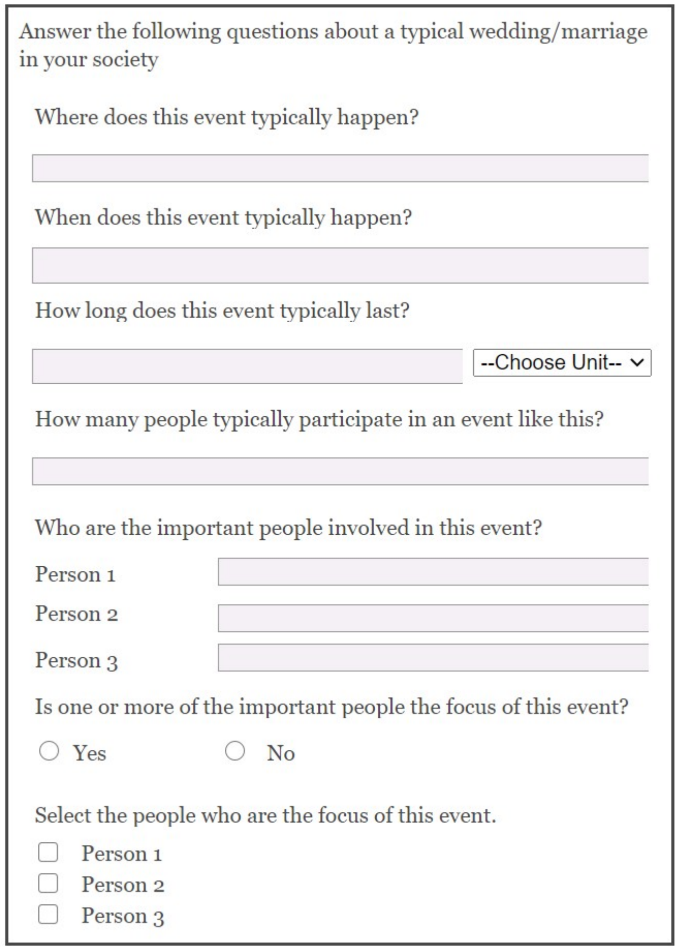
5.4 Preliminary Data Collection

The preliminary data collection efforts for this was split in two ways: I first performed a pilot study with a handful of participants to verify the correctness of approach, Van Teijlingen and Hundley (2001); Hazzi and Maldaon (2015) and then a small scale data collection on Amazon Mechanical Turk before moving on to the larger effort. In this section, I explain both the procedure of these efforts as well as the results.

5.4.1 First Pilot

For the pilot experiment, I collected data from a small number of participants for three rituals: `coming of age`, `wedding`, and `death rites/funeral`. I collected two unique sets of responses per ritual per culture, for a total of 12 unique responses. The identification of cultural group membership was done via self-identification by the participants, based on a demographic questionnaire that preceded the main survey. The participants took the survey using an online form. The actual survey consisted of a series of questions that were modifications of the ATOMIC (Sap et al., 2019b) question set (see

Section 5.2.4 for details). The questions remained constant across rituals and cultures with only the initial prompt changing to keep the method as consistent as possible. The survey was conducted asynchronously and participants were compensated for their time.



Answer the following questions about a typical wedding/marriage in your society

Where does this event typically happen?

When does this event typically happen?

How long does this event typically last? --Choose Unit--

How many people typically participate in an event like this?

Who are the important people involved in this event?

Person 1

Person 2

Person 3

Is one or more of the important people the focus of this event?

Yes No

Select the people who are the focus of this event.

Person 1

Person 2

Person 3

Figure 5.1: Layout of the main survey form as seen by the survey participants.

5.4.2 Pilot Crowdsourcing

Once the small pilot study affirmed my idea, I moved ahead with collecting data on a slightly larger scale. I used Amazon Mechanical Turk (MTurk) to collect the second round of data. I set up separate tasks on MTurk for United States (US) participants and Indian (IN) participants, and geo-restricted the tasks to workers from the respective countries. I also restricted the tasks to *Master* workers—workers with a work approval rate of 90% or more. I set up a system of auto-generated survey codes that linked responses to MTurk workers without having to collect any Personally Identifiable Information (PII); these codes were used to filter out spam entries. The workers that provided spam responses were banned from the tasks. Overall, I collected a total of 32 useful data points for Indian participants, and 33 for US participants; these were divided roughly equally across rituals, with at least 5 responses per ritual.

5.4.3 Preliminary Results

In this section, I detail the findings of our study. I first report on the demographics of the survey participants, particularly with an eye towards cultural background, and then recount and discuss the responses to the survey.

I obtained a total of 77 unique survey responses with each participant limited to one response per ritual per culture. Only information that was deemed necessary for the purposes of the experiment was collected from participants, with an aim to avoid collecting PII as far as possible. Out of the 77 responses, 38 were from India and 39 from the US.

Religion

Since many rituals have a basis in religion—to the extent that they are often intertwined (Goody, 1961; Geertz and Banton, 1966; Bell, 1992)—it is important to ensure a diversity of religious practice among the respondents. Of the participants in the study, 29 said they practiced Hinduism; 17 said they practiced some form of Christianity; four Islam; four other religions; and 23 said they did not practice any religion and/or were atheist ($N = 77$). Among participants who identified as Indian, there was significantly less variation in religion compared to the participants who identified as Americans. There was also significantly more change in religion practiced over a lifetime for US participants as compared to IN.

Language

Among the $N = 77$ participants, 39 identified their native language as English; 17 Tamil; seven Hindi; five Telugu; four Urdu; and one each as five other languages.

5.4.4 Discussion

In this preliminary work, I have come up with a valid setup – as proven by the pilot studies – of a method for collecting cultural information from diverse groups about different life rituals. This work has shown that the results confirm my hypothesis that variations in cultures can be observed using studies like this one, and if done in a grander scale, can yield a host of useful information to be used in Question Answering and other NLP systems. While the data I have collected thus far is too small to be used to directly improve the performance of QA or other NLP tasks, the approach here allows for scale-up into a full dataset. With this conclusion I moved on to a larger scale of data collection.

5.5 Crowdsourced Data Collection - Large

The large scale data collection on Amazon Mechanical Turk was done following almost exactly the same process as the preliminary effort, with only minor changes. The first was obviously the number of participants involved in the study, which I scaled up massively. The second was to change the worker qualification for the data collection, as I found out that the volume in which I wanted the data would not be possible with the default *Masters Qualification* restriction, which only allows the top workers on the platform across a variety of domains. The restriction of having worked in a wide variety of domain was unnecessary for my task as it simply dealt with their native cultural practices. Instead, I opted for people who have had high quality work (over 90% approval rate) regardless of whether they worked in one domain or multiple. This allowed me to reach a much wider audience.

5.5.1 Results

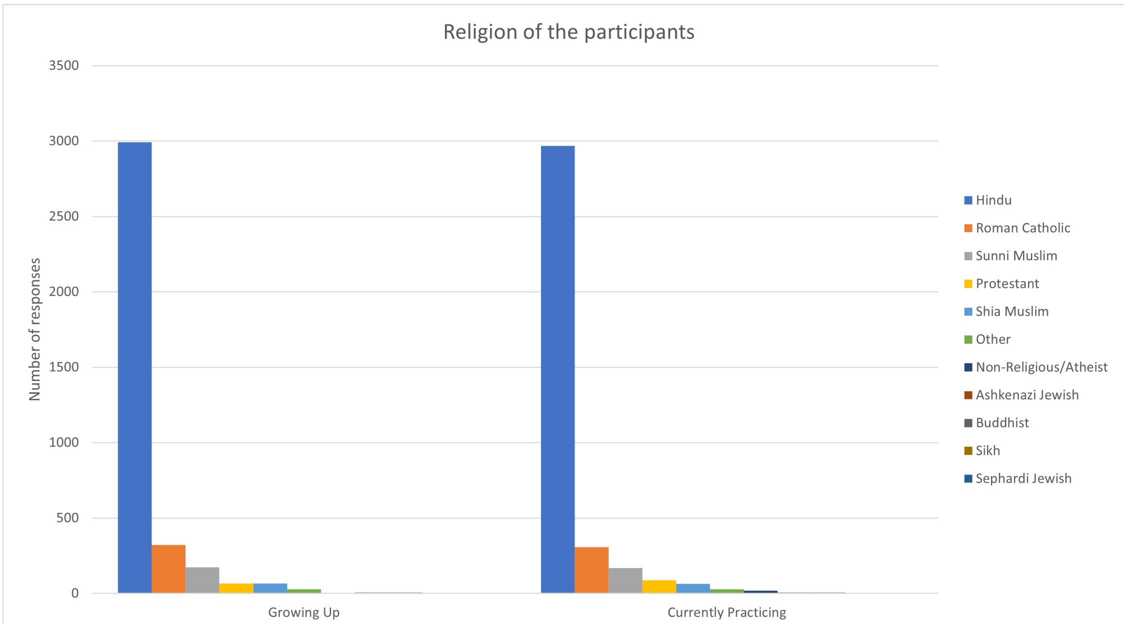
There are a total of 10,284 unique survey responses with each participant limited to one response per ritual per culture as before. Out of the 10,284 responses, 3,658 were from India, 4,271 from the US, and the remaining 2,355 from other countries. In terms of rituals, although the number of responses per ritual varies slightly, all rituals have at least 1,500 unique responses each. The detailed statistics about the data collected can be seen in table 5.3.

Ritual	US	India	Others	Total
Birth	633	604	319	1556
Birthday	770	630	409	1809
Coming of Age	667	612	414	1693
New Year	674	617	386	1677
Wedding	761	588	421	1770
Funeral	766	607	406	1779
Total	4271	3658	2355	10284

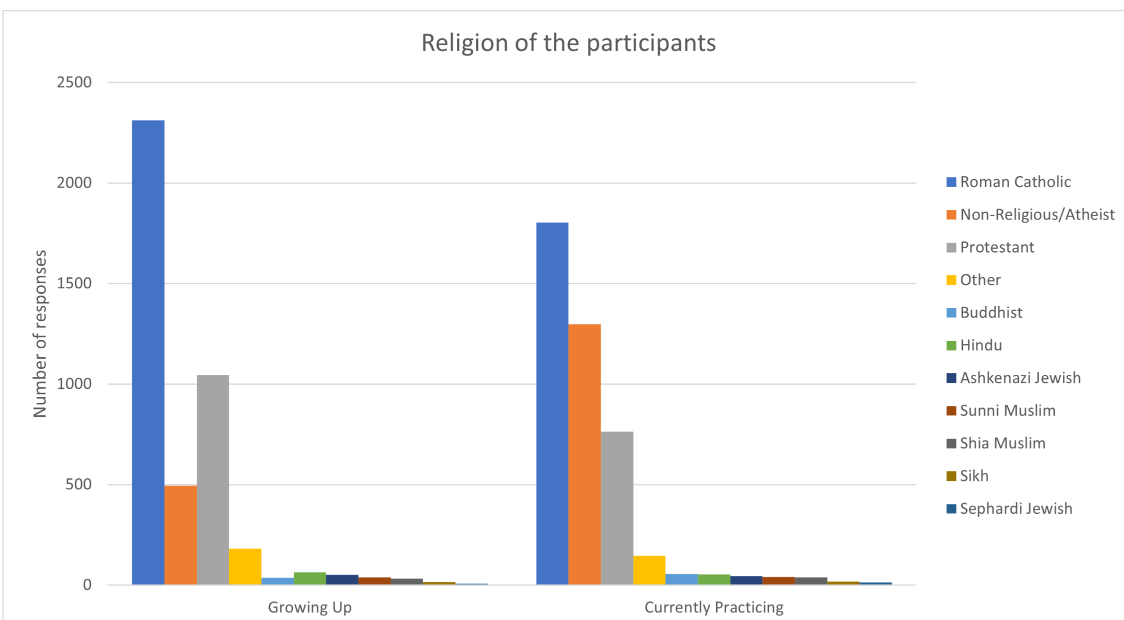
Table 5.3: Full statistics for the cultural data collected using MTurk during the main phase data collection effort. The numbers are unique responses for each category.

Religion

The major religion among the participants was Christianity, with 4,146 responses saying they practiced some form of Christianity. Hinduism was the second most common with 3,045 responses. 2,210 responses mentioned they practices no religion and/or were atheists. There were also 434 responses that said they practiced Islam, 89 Judaism, 78 Buddhism, and 25 Sikhism. There were 317 responses for other religions except these ones.



(a) Religion Data for the participants from India.



(b) Religion Data for the participants from the US.

Figure 5.2: Variation in the religion practiced by participants growing up versus currently for both cultures.

Like in the small data collection, there was again much more variation in religion in the participants from the US compared to India, while the respondents from India seemed to change their religion practices much less than their US counterparts. This is illustrated in fig. 5.2.

Language

Among the participants, 5,235 responses identified their native language as English; 1,190 Tamil; 1,089 Portuguese; 1,078 Hindi; 318 Italian; 262 Malayalam; 146 Spanish; 101 Bengali; and other languages made up the remainder.

5.6 Qualitative Analysis of Data

The results show us that while rituals have some common features across cultures, they can also have significant variations to the point where the *common knowledge* would be noticeably different. Let us look at one significant difference seen in the responses: for the *wedding* ritual, participants from the US said the bride would focus on the wedding planning part of the event, like dresses and so forth; while the Indian participants focused on the cultural aspects of the wedding, as well as the fact that the bride might have to get to know the groom's family, and possibly the groom himself, as illustrated in Figure 5.3. It would be extremely unlikely for a bride to not know the groom's family, let alone the groom himself, in a typical US wedding; while this is still reasonably prevalent in Indian society. This is an excellent example of the type of knowledge that is collected by our work, where a machine can now leverage this information as commonsense knowledge that is culturally sensitive and correct.

Figure 5.3: The differences in expectations of a bride in the US versus in India.

While the data showed that the person-roles involved in rituals were more or less similar across cultures, the order of importance indicated was noteworthy, and shown in Table 5.4. For birth rituals, participants in the US considered the birth itself the main event, and said the Doctor (physician) was an important person (after the parents); while Indian participants focused more on the family. It is also worth noting that US participants put Mother before Father, while Indian participants reversed the priority order. All participants gave similar responses in terms of the important person-roles for a wedding: the bride, the groom, and their parents. An important difference was that Americans identified the Bride's mother as the more important person, while Indians picked the Groom's mother for the same. Likewise, for a funeral, more priority was given to the pastor or priest by US participants; unlike Indian participants, who considered the family to be the more important participants.

Another set of major differences were seen in terms of the durations of the rituals. While the duration of the other four rituals were more or less comparable, weddings and funerals had significant differences across cultures. For the *wedding* ritual, participants from the US said that the ritual typically lasted a few hours, while Indian participants responded by saying that weddings lasted multiple days. The difference was even more striking for funerals, which US participants said lasted only a few hours; while Indian participants reported funeral rites lasted up tp 13 days.

US	*India*	*US*	*India*
Birth		**Birthday**	
Mother	Father	Self	Parents
Father	Mother	Partner	Family
Doctor	***Family***	Family	Relatives
New Year		**Coming-of-age**	
Spouse	Friends	Self	Parents
Parents	Spouse	Parents	Family
Friends	Family	Siblings	Relatives
Wedding		**Funeral**	
Bride	Bride	Spouse	Father
Groom	Groom	***Pastor/Priest***	Son
Bride's Mother	***Groom's Mother***	Parents	Family

Table 5.4: Important people for each ritual by culture.

One other key difference observed was the number of people that participated in each ritual. Across all rituals, those in India seemed to involve a lot more people than in the US. While the difference in numbers was already pronounced for other rituals (Indian funerals seem to have *twice* as many people as US ones), the striking difference is weddings, where Indian ones had several hundred guests, while US ones averaged about 100.

These fin dings validate our expectation that rituals can give us a peek into cultures and how they vary, and that commonsense knowledge cannot truly be complete without including cultural nuances.

5.7 Discussion and Future Work

With this work, I sought to create a repository of cultural commonsense knowledge, and have succeeded in doing so. I envision that such knowledge can greatly improve the ability of AI systems to exhibit human-like performance by addressing gaps in their current knowledge. The task of injecting cultural sensitivity into commonsense reasoning, while being crucial to developing a true human-like AI, has not been previously explored in the field. I exposed this gap, and in order to bridge it, performed the difficult task of choosing suitable cultural markers that would work within existing frameworks of commonsense knowledge, and then collecting large-scale data for it. The obvious next step is to use this dataset to enhance an existing system, the details of which I mention in chapter 7.

Another direction for future work is to explore the various knowledge representation techniques to find the best way to represent what will likely be multi-dimensional data. While I did not have time to explore this path, it is an important future step that could make the field of cultural knowledge even better.

5.7.1 Knowledge Graph

One main future thrust of this work is to construct a large scale knowledge graph using this data. Past efforts at systematizing commonsense knowledge (Sap et al., 2019b; Bosselut et al., 2019; Speer et al., 2016; Liu and Singh, 2004) for the NLP community's use have taken a similar path, starting from data collection through analysis and summarization of the data, then eventually the construction of knowledge graphs from those summaries. For my work, these summaries look like the examples shown in Figure 5.4. Specifically, I consider the *wedding* ritual for both US and IN cultures. A quick examination of Figure 5.4(a) shows that while for IN, there seems to be a significant emphasis on the wedding & ”arrangement” of the wedding [1] as well as the family of the participants; whereas for US, shown in Figure 5.4(b), the emphasis is much more on the celebration as a *party*, and the social aspect of the event, noted by words like ”honeymoon”, ”dress”, ”ring”, ”reception”, etc.

The next step in our process is to represent this abstract space of words in the form of a directed graph composed of entities and relations, akin to most existing knowledge

[1] In context of Indian weddings, *arrangement* typically refers not to the programmatic of the event itself, but rather the arragement of the match itself.

(a) India

(b) USA

Figure 5.4: Most common words in the responses for *wedding* ritual for both cultures represented in word clouds, showing us what concepts appear more often in each case.

graphs. I performed this on the small scale dataset as a demonstration of what can be done on the larger dataset.

In order to do this, I need to performed several stages of cleaning and data filtering of the data to extract the relevant sections, followed by aggregating the details present in the data for each *event* and *person* together. Case-folding was done on the nodes and edges to avoid duplicates. After this, I extracted all potential nodes from each instance of data, with the relevant *event* or *person* being the origin node; the prompt text as seen

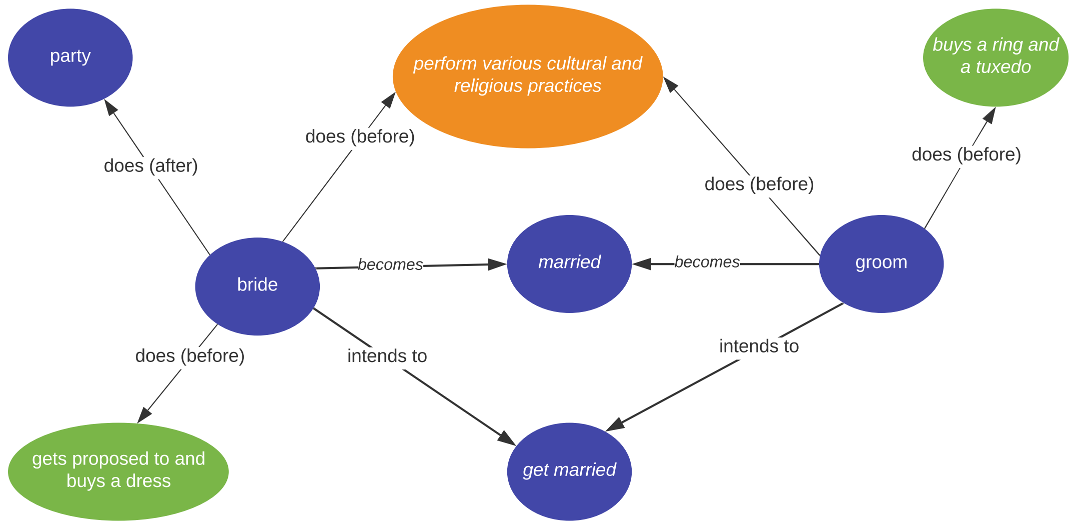


Figure 5.5: A graph for the subset of the data collected for the *wedding* ritual. The nodes in orange show responses of Indian participants; those in green show responses of US participants; nodes in purple are common to both. The figure has been manually edited for grammar, and is intended for illustration.

in Table 5.2 and Figure 5.1 as the (relation) edge, while the answer for each question was the corresponding destination node. This data was then visualized as a network using Python's GraphViz package (Ellson et al., 2004), merging nodes with the exact same labels into a single node, with nodes corresponding to the two target countries being assigned different colors. An illustrative example of the end-result of this process is shown in Figure 5.5. This sub-graph is a glimpse of what the entire knowledge base will look like after we run an entire dataset through a rigorous NLP pipeline containing semantic role labeling, (Palmer et al., 2010) word sense disambiguation, (Ide and Véronis, 1998; Agirre and Edmonds, 2007) syntactic and semantic parsing, (Earley, 1970; Klein and Manning, 2003) among others. This sub-graph and the larger graph that it is a part of—which can also be represented in the traditional adjacency graph format for consumption by NLP systems—are the ultimate goal of our work on creating a cultural commonsense knowledge graph.

CHAPTER 6

MOTIF ASSOCIATION MINER

6.1 Motivation

As I've discussed in earlier chapters, motifs are unique cultural markers that carry a wealth of information. They typically arise from scripture and folktales, and have significant meaning within a cultural group. I have shown in chapter 3 that out-culture people understand meaning of motifs in text much less than those in-culture. However, as proven in chapter 4, these people can still recognize that motifs carry special meaning and are aware that they're missing out on that meaning. The natural next step, then, is the need to build a system that can take this meaning of motif and make it available to people who are not part of the *in-group*.

But as it turns out, a simple dictionary of motifs is not sufficient for out-culture people to understand the deeper meaning and context of a motif. Let us take an example of a Hindu motif, Saraswati. If one unfamiliar with the motifs looks up Saraswati on Google, they will know that she is the Goddess of knowledge, wisdom, music and art, among other things. But it is unlikely that they will know the meaning if they hear a sentence like "Kartik has Saraswati on his tongue." It might sound like that means Kartik is highly intelligent or intellectual, but what it actually means is Kartik never tells a lie. So we see that with just a simple glossary style resource, it is not possible to bring forth the nuance and the gravity of meaning that motifs typically carry. It requires several different information about the motifs included together, alongside example usage, to provide the larger context of what those motifs really mean. Obviously no approach will be truly perfect when it comes to conveying the whole backstory that is contained in a motif, but with proper effort and carefully curated data, we can get a reasonably good approximation.

This is why I aim to build this system, the Motif Association Miner, that can bring forth these various information about motifs to a wider audience for better understanding.

6.2 Approach

For this work, I planned on developing a prototype for the Motif Association Miner (MAM) that can provide associated concepts and other information pertaining to a given motif. The MAM is a system that can take as input chunks of texts that contain known motifs in them, and use information from various sources like online discourse and source narratives to provide various associations and meanings of the said motifs. This system is meant to work with the Motif Detector from the main phase, and this MAM assumes an existing Motif Detector capable of detecting motifs in texts. In the long term, this MAM is envisioned as the primary component of a bigger Motif Messaging System (MMS), as shown in Figure 6.1. I should note that the MMS is an imagined entity, and does not yet exist, nor do I try to work on it now. I present it only as a vision of how motifs could be used in a much larger way in the future.

So the immediate goal, then, of this work is to see if I can build a system that can automatically find motif associations. This system exploits motifs to better understand and construct influence narratives. It is built on top of an existing motif detector and motif list. Since a generalized motif detector does not exist, I use use motif detector built by a team including myself with my colleague Yarlott (2021) – which is not part of this dissertation – that can detect select motifs.

It is worth noting that I did try to see if the task can be achieved by using an existing off-the-shelf Information Extraction System. However, upon testing several SOTA Information Extraction systems like OpenIE (built based on Christensen et al. (2011) and Pal et al. (2016)), CALMIE, Saha et al. (2018) and BONIE Saha et al. (2017) and examining

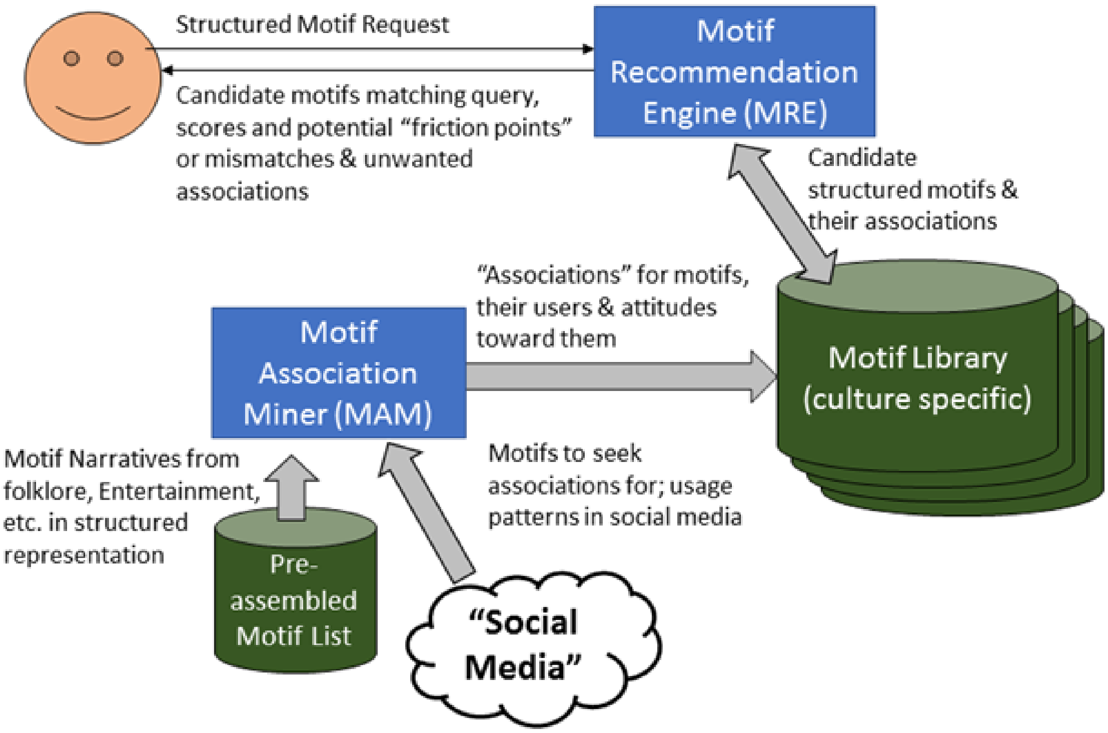


Figure 6.1: The core, motif-centric elements of ACUMEN-MMS. The Motif Association Miner (MAM) shown here is the entity we're working on building during the option phase of ACUMEN.

the output, I found that this was woefully inadequate and did not add any value to my task. One main reason for this was that most Information Extraction systems that currently exist focus on each sentence in isolation, i.e. they consider sentences to be isolated portions of text and do not consider the rest of the text. So I came to the conclusion that motif relations and associations can not be solved by existing information extraction tools, and continued with the work of creating a custom Motif Association Miner.

The system takes as input the motif list, as well as the raw text from various sources. The sources for this raw text include social media and news data, as well as thesauri, dictionaries, encyclopaedias, or motif indices that often contain direct descriptions of motifs and their associations. I then use the motif list to find texts excerpts from the raw text that has motifs. This text excerpt with motif then goes through an NLP pipeline consisting of various standard NLP tools and components, like spaCy (Honnibal and Montani, 2017), Stanford CoreNLP (Manning et al., 2014), Semantic Role Labeler, Word Sense Disambiguation, and so on. This results in the motif text with NLP annotations. This motif text with NLP annotations is then used to extract several associations for the motifs within it. I pass them through a pipeline of various NLP tools that mine several other associations. This includes using Semantic Role Labeler to get semantic actors and Discourse Parser to detect other properties associated with the motifs.

6.3 System Architecture

The Motif Association Miner system is composed of various component pieces of existing NLP tools and modules that are combined to fit my need. I used Python as the primary language for this system for the code I wrote myself, and also used the Python version of the various existing NLP tools for ease of access and seamless integration. Figure 6.2 shows the system architecture of the Motif Association Miner (MAM). This system is designed to take a piece of text containing motif as input, and produce a report that contains information about the motif present in the input text. To do this, the input text is provided as input to the various components. As most of these component processes take a significant amount of time to run over large batches of text, I save the output of each individual component of the system to disk. This ensures the speed of some of the components do not become a bottleneck when it comes to running the entire system.

The Association Miner takes raw input text that contains motifs within it as the main input to the system. This text then goes through the Motif Detector, which finds motif present in the system. This text tagged with motif then goes into the information synthesizer as well as various NLP processes: Semantic Role Labeler, Named Entity

Recognizer, Sentiment Analyzer, Discourse Parser, and Part-of-Speech Tagger, where the respective NLP tasks are performed on the text.

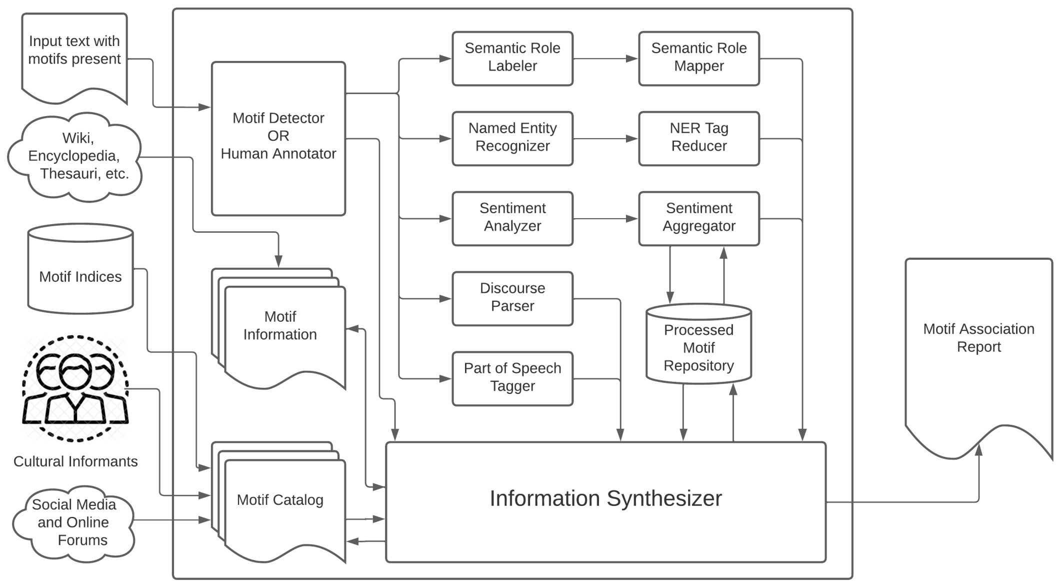


Figure 6.2: The main architecture design of the Motif Association Miner, including the external components that feed into the system.

The text processed through the SRL goes to a Semantic Role Mapper, which maps the semantic role of the motif to a simpler output form. Similarly, the text processed through the NER goes to the NER Tag reducer, which again changes the NER tag to a simpler output format. The output from the Sentiment Analyzer goes into the Sentiment Aggregator. This Sentiment Aggregator also takes input from the Processed Motif Repository, which contains outputs from the NLP processes for past motific texts, and calculates the overall sentiment of the motif. It also categorizes the sentiment into one of the four classes. The output from all these components, as well as the Discourse Parser and the Part-of-Speech Tagger, goes into the Information Synthesizer.

In addition to the output from these components, the Information Synthesizer also takes input from Motif Information module and the Motif Catalog. It saves the NLP information of the motific text to Processed Motif Repository for future use. It then uses all the information to generate the Motif Association Report. A more detailed explanation of the Information Synthesizer, as well as all other individual components of the system, is provided in the following section.

6.4 Component Processes

6.4.1 Motif Detector

The Motif Detector is one of the most crucial components of the Motif Association Miner. The MAM system relies on the motif detector to be able to recognize motif present in the input text. Traditionally, this work always had to be done by human annotators, since there was no automatic motif detector system in place. However, recent work from my colleague Victor Yarlott et al. (2021) means I can use that motif detector for this system. During the development of the motif association miner, however, I substituted the output of the motif detector with the human annotated data, as the motif detector was not yet ready at the time of start of development of this system. This is an acceptable choice because the motif detector that is part of the system is meant to be the automated method for this exact human annotation process.

The Motif Detector (automatic or human) takes the input text and detects whether any motif is present in the text, and if so which motif it is. This input is then passed onto the Information Synthesizer, Semantic Role Labeler, Named Entity Recognizer, Sentiment Analyzer, Discourse Parser, and Part-of-Speech Tagger.

6.4.2 Semantic Role Labeler

An important piece of linguistic information about any word or phrase in a text is to know it is semantic role (Babko-Malaya, 2005) in the said text. Using this component, I am looking to see what sort of semantic role the provided motif plays in a sentence. This information can tell us how to use the motif in a sentence, and potentially detect incorrect usages. For instance, if a motif is typically used as the semantic subject of a sentence, but we encounter a text where the same motif is used as an instrument, it might mean that the motif is improperly used, implying the origin of the motific text isn't from in-culture sources.

To implement this component, I used AllenNLP's SRL. This SRL is a slightly modified version of a BERT based model (Shi and Lin, 2019) that is currently the single-model SOTA in English news wire sentences. It reports an F1 of 86.49 on the Ontonotes 5.0 (Weischedel et al., 2013) dataset.

The Semantic Role labeler takes the output of the motif detector as input, and performs SRL on the text containing the motif. This SRL output is then sent to the Semantic Role Mapper.

6.4.3 Semantic Role Mapper

There are numerous different semantic roles words can have in any sentence. Appropriately, the AllenNLP SRL that I utilize is able to categorize the sentence into all the available semantic roles based on the ProppBank Annotation Guide. (Babko-Malaya, 2005). However, these details are not required for the sake of the project, and would add unnecessary complexity to the final output that is supposed to be understood by someone with no expert knowledge in linguistics. So I map the various SRL tags into just four main roles, while ignoring all other roles as *"non-specific."* The full mapping is shown in table 6.1.

The mapping is done on the output of the SRL, which this system takes as input. The text with the new SRL labels are sent to the Information Synthesizer.

Old Semantic Role	Simplified Semantic Role
ARG0	The subject
ARG1	The object
ARG2	The instrument
ARG3 *ARG4* *ARGM-TMP* *ARGM-EXT* *ARGM-REC* *ARGM-PRD* *ARGM-PNC* *ARGM-CAU* *ARGM-DIS* *ARGM-ADV* *ARGM-MOD* *ARGM-NEG*	No specific major role

Table 6.1: Mapping of SRL categories by removing some linguistic details to simplify the output

6.4.4 Named entity recognizer

Named Entity Recognition – as the name suggests – refers to the task in NLP in which we detect the named entities present in a given text. This kind of information is quite useful for a task like ours, because with this information we can tell if our motifs are used to refer to a person or a location, for example.

To perform this task of Named Entity Recognition, I again use AllenNLP's implementation. Specifically, I use AllenNLP's ElMo-Based NER, which is an implementation of the baseline model from (Peters et al., 2017). This model achieves a reported score of 96% F1 score on the CoNLL-2003 validation set. (Tjong Kim Sang and De Meulder, 2003)

6.4.5 NER Tag Reducer

The AllenNLP NER tagger uses a twenty-one class output format using the Inside-Outside-Beginning tokenization and tagging style (Peters et al., 2017). While this is extremely useful information for tasks requiring deeper linguistic tasks, it is unnecessary for the purpose of this work. So I reduce the number of possible NER tags for the motifs by consolidating the named entities tags of the same type to single hypernym class. This approach resulted in just five NER classes, making it more appropriate for audience with no linguistic background ensuring a larger target audience for the output report.

This component takes as input the NER tagged text, and the text with the new tags is sent to the Information Synthesizer.

Old Tag	**New Tag**	**Meaning**
[PER], I-[PER], O-[PER], B-[PER], L-[PER], or U-[PER]	[PER]	Person
[ORG], I-[ORG], O-[ORG], B-[ORG], L-[ORG], or U-[ORG]	[ORG]	Organization
[LOC], I-[LOC], O-[LOC], B-[LOC], L-[LOC], or U-[LOC]	[LOC]	Location
[MIS], I-[MIS], O-[MIS], B-[MIS], L-[MIS], or U-[MIS]	[MIS]	Miscellaneous
[O]	[O]	Not a named entity

Table 6.2: Reduction of NER categories by removing tokenized tagging to simplify the output. The different sub-types of the same named entity are shown together. Each word can only have one tag.

6.4.6 Sentiment analyzer

Another main information needed for our final output is the connotation in which the motif is used, i.e., whether the motif is used generally with a positive, negative, or neutral connotation. Understanding how a motif is used can be hugely important in understanding the meaning of the said motif. For instance, in a sentence like “That person is such an Amalek.”, just the connotation alone can give us significant insight into the meaning. It can be the difference between a piece of friendly text versus a hostile one.

For this, I utilize an existing sentiment analyzer. Because the work of sentiment analysis is still a field of active research, I considered quite a few different systems (Bhatia et al., 2015; Heerschop et al., 2011; Chenlo et al., 2014; Gao et al., 2019; Xu et al., 2019; Rao et al., 2018) before deciding on implementing one. However, I ultimately decided to again use AllenNLP’s implementation of sentiment analysis. This system used a LSTM classifier with GloVe embeddings (Pennington et al., 2014) and performs close enough to SOTA performance, with a reported performance of 87% accuracy on the Stanford Sentiment Treebank corpus (Socher et al., 2013). It also has the added advantage of running out of the box.

The Sentiment Analyzer takes as input the output from motif detector, and computes the sentiment value for the text with motif. It then passes the output to the Sentiment Aggregator.

6.4.7 Sentiment Aggregator

The Sentiment Aggregator takes two inputs: first, the output from the Sentiment Analyzer the sentiment value; and second, all the sentiment values of the same motif recorded in the Processed Motif Repository. It then averages the sentiment value across all instances of the motifs. This is done to make sure that the output report does not rely on the single example of the motif.

The original sentiment output from the Sentiment Analyzer is binary: 0 for negative and 1 for positive. This is too coarse-grained for the purposes of motifs. So once the sentiment is averaged over all values, it is then mapped into a sentiment type. I use four sentiment categories for this system, as shown in table 6.3. This final sentiment class output is then sent to the Information Synthesizer.

Sentiment Label	Sentiment Value
Negative	0 – 0.25
Slightly Negative to Neutral	0.25 – 0.5
Neutral to Slightly Positive	0.5 – 0.75
Positive	0.75 – 1

Table 6.3: New sentiment categories for the motifs

6.4.8 Discourse parser

The discourse relation of the motif in a text with other spans in the same text is another useful information for us to have for the Association Miner. While the discourse relation is not directly useful for the final output report, it can provide us good insight into how the motif interacts with other parts of the discourse.

For the discourse parser, I studied several existing systems (Soricut and Marcu, 2003; Fisher and Roark, 2007; Lin et al., 2009; Hernault et al., 2010; Feng and Hirst, 2012; Li et al., 2014; Ji and Eisenstein, 2014; Joty et al., 2015; Nguyen et al., 2021) to see which one would be the best fit. While several systems had very similar performance to SOTA and I could have chosen any of them, I eventually decided to use the RST parser (Nguyen et al., 2021) which, in addition to being SOTA, had the added benefit of being implemented in Python. This system reports a full F1 score of 50.2 on the gold data, which is better than all reported major existing systems, and is comparable to human agreement of 55.0.

When used, the Discourse Parser takes as input the output from the Motif Detector. It then tags the text for discourse relations, and then the output is sent to the Information Synthesizer.

Unfortunately, the annotated data currently available for the motifs did not have enough narrative / discourse structure for me to be able to utilize this information for the current version of the output. Nevertheless, I include this component here because it could be a vital piece of information provided we have enough good data.

6.4.9 Part-of-speech tagger

The role of part-of-speech of a word in any text is fairly common knowledge. This information helps us realize how a word is used and what other words it could be used with, among other things.

The part of speech tagging is done using spaCy's (Honnibal et al., 2020) medium-sized CPU-optimized pipeline. Since part of speech tagging is practically a solved problem in NLP, there isn't a lot of difference between using any of the major packages. I chose the spaCy one for ease of usage because we were familiar with it.

Again, this Part of Speech Tagger takes as input the output from Motif Detector, and the output – the text with the part of speech tags – goes to the Information Synthesizer.

6.4.10 Processed Motif Repository

The Processed Motif Repository is an extremely crucial component of the system. This is where all the output from the NLP processing of input texts with motifs are saved. This is done so that the output report is not reliant on just a single instance of input text as the example use of motif. This also ensures that the system continues to learn as it processes more and more data.

To seed the system, I utilize a gold-annotated motif data.[1] This gold annotated dataset consists of texts that contain motifs that have been labelled by in-culture human annotators. Overall, this dataset contains a total of 1747 instances of motifs in text across three cultures: Irish, Jewish , and Puerto Rican. The dataset was double-blind annotated and adjudicated. I run the all the NLP components that are part of the Association Miner system on all the instances of text with motifs found in the dataset, and cache the results of all intermediate components. These results combined then becomes the Processed Motif Repository, which is now a component in the Association Miner. This process means that the Association Miner can be adapted to work for any motif as long as there is enough data for the said motif to generate the repository data.

In regular use, the Processed Motif interacts mainly with the Information Synthesizer. It provides old results to the Synthesizer for the generation of output, while accepting the new results to be saved for future use.

6.4.11 Motif Catalog

The Motif Catalog is another essential dependency of the Motif Association Miner. Although I present it as a component of the system, the Motif Catalog has the potential to be a large and public resource on motifs, not dissimilar to how WordNet University (2010) is currently used for word senses and other tasks in NLP.

The Motif Catalog used here was created by me together with my colleagues in the overaching project, so it is not included as part of the dissertation. Hence, I do not describe the process of creating it in detail here. I will explain briefly, the current version that the Association Miner system utilizes.

[1]This dataset comes from annotation work done by my colleague Victor Yarlott, with some help from me. This is unpublished as of the writing of this dissertation.

To create the motif catalog, we first selected specific target cultural groups, which in this case was Irish, Jewish, and Puerto Rican. For these cultures, we curated a list of motifs to use for this first version. This list of motifs can be found in §3.2.2, table 3.1. We then collected relevant information about the motif from various sources, including the original motif indices, the cultural informants for each culture, as well as social media and other online forums. For future use, if the Association Miner is to be expanded for other motifs, this component can be easily automated with the exception of human cultural informants. In its current form, the Motif Catalog contains basic information about the motif, like the origin culture, definition, associations, and example usage.

During runtime, the Motif Catalog shares this information with the Information Synthesizer for the motif that is present in the input text.

6.4.12 Motif Information

The Motif Information component contains, as the name suggests, various information about the motifs. I gathered this information by downloading a Wikipedia page per motif in a readable .txt format. I then parsed through the data and cleaned it up by removing text such as donation advertisements. Afterwards, I extracted relevant information from the data by separating the text into its 'Contents' categories and keeping important categories (such as background information, summary, traditions) and discarding others (such as references). While the current version of the Motif Information only contains information from Wiki sources, it is possible to expand this by adding information from other sources like encyclopedias and thesauri.

The Motif Information module again interacts with the Information Synthesizer, providing the relevant information for the motif present in the input text. Because there is such a large volume of information about each motif in this module, typically only a subsection of it sent, based on the request sent by the Information Synthesizer.

6.4.13 Information Synthesizer

The Information Synthesizer is the central component of the Association Miner that deals with all the final data processing and output generation. After the input data has been processed by the aforementioned components, the output from these components are passed on to the Information Synthesizer. Here, the output from these components are then aggregated, and the relevant information from these outputs is extracted. Finally, it is re-formatted to then produce the final report.

Using the output of Semantic Role Mapper and the information from the Processed Motif Repository, the overall semantic role of the motif is computed. Similarly, with the output of the NER Tag Reducer and the information from the Repository, the typical named entity role for the motif is determined. These two, combined with the output from the POS Tagger, determines how motif is used in a sentence. Likewise, the information from Motif Information module and the Motif Catalog are also extracted for the relevant motif.

In order to present the output in the report, it is first important to have a proper structure of the output defined. To this end, I treat this as a specialized template filling task. I create a template for the output Motif Association Miner Report comprising of fields that I think are necessary to be present in the report, and also are achievable with the information available. To match the fields in the template, all the information in the synthesizer goes through the process of converting tags and stubs into phrases and/or sentences and merging information together from different modules to put relevant information together. One possible source of error in this process is if the information from the external sources or the catalog is too long. To avoid this, I truncate the information to a fixed length. Another possible issue could be that a motif is used in multiple semantic roles across the examples found in the repository. In fact, while performing the initial run of the data I found that this was indeed the case. In such cases where there is no obvious most common usage, I report it as “*no specific role*”.

The final result of this process is a filled out template, which is the final Association Miner Report, as shown in fig. 6.3.

6.5 Results and Discussion

Since this task is a version of targeted Information Extraction and the output is a human-readable report, we cannot use traditional evaluation metrics to assess the success of the work. So I do not present a quantitative evaluation of the output. A qualitative analysis of the output, however, shows that most of the information produced by this system were mostly accurate to the motifs. The report is also easy to read and can be understood by anyone who is looking to get more context and understanding about any motifs they might run into in texts. Overall, we see that the system works as intended for motifs as long as there is sufficient data present about the motif. A report generated by the system is shown in fig. 6.3.

One main feature of the system, of course, is that there is still one component of the system that requires manual processing: the creation of motif catalogs. So as of now the system cannot be considered fully automatic. However, this is only because the current state of study of motifs is inadequate. There is no fundamental difference, for example, between my system using a Motif Catalog as a resource compared to any major SOTA NLP system today using WordNet (Miller, 1995) or VerbNet (Schuler, 2005) as resources, which also required large-scale human effort to create. So this system has the ability to be fully automatic, provided motif catalogs become as common as sense inventories or dictionaries in the coming days.

6.6 Limitations and Future Work

One key future step is to look at the discourse in which motifs are used, rather than look for meaning of motifs in isolation. This will allow us to understand meaning of motifs from larger context. For instance, we not only have sentiment values and linguistic properties for the sentences with motifs themselves, but also for other sentences that form part of the discourse that involves the motif. Using this discourse relation information, we will be able to more accurately predict the associations for the motifs. While I planned to do this for the current system, a lack of sufficient examples of motifs used in narrative style text meant this could not be properly executed. With more data of real-world usage of motifs, this can be implemented in the future version of the system.

Another future direction is to be able to compute the difference between the perception of motifs for in-group people vs out-group ones. Again, this is not currently possible due to lack of sufficient data. As the information about motifs become more abundant, we can envision a small modification in our MAM being enough to generate this information. Alternatively, a large scale data collection study on motifs and how they are perceived by people across cultural groups would also help solve this problem. But that is another large study by itself and unfortunately beyond the scope of my work in this dissertation.

And finally, the ultimate goal for the system is generalizability: to be able to work for any motif, not just known ones. To do this, two principle components this system is based on would need to also be generalized: the Motif Detector and the Motif Catalog.

Title	**Content**
Motif	Leprechaun
Found in	I previously talked about the franchise in my usual breakdown format four years ago, but I've never ranked them. We live in a world where there are seven Leprechaun movies. Seven. I just can't comprehend that. They're not even particularly good movies. Of course, I say this and I own the entire series on blu-ray. Hey, they're still fun and Warwick Davis is always entertaining. So let's look at all seven of these movies and see which is the best!
Motif Type	Character
Origin Culture	Irish
Usually referred to	Mostly used to refer to a person. Generally used as no specific role (eg. subject/object) in a sentence.
Major associations	Tricky, grumpy fairies who steal treasure and horde it. Associated with (1) greed, (2) trickery, and (3) being short.
General Usage Connotation	Neutral to Slightly Negative
Motific Examples	(1) That old miser is a real leprechaun. (2) That leprechaun at the used car lot really got the better of me. (3) I swear I could jump over Ethan, he's a real leprechaun.
Referential Example	A leprechaun is a type of fairy of the aos si in Irish folklore.
Unrelated Example	N/A—most uses are going to be at least somewhat culturally related.
Eponym Example	The photo you see for Leprechaun, Inc. is a 5,000 year old Dolman, or Portal Tomb, built during the Neolithic Period.
Background	A leprechaun (Irish : /leipreachán/luchorpán/) is a diminutive supernatural being in Irish folklore, classed by some as a type of solitary fairy. They are usually depicted as little bearded men, wearing a coat and hat, who partake in mischief. In later times, they have been depicted as shoemakers who have a hidden pot of gold at the end of the rainbow. Leprechaun-like creatures rarely appear in Irish mythology and only became prominent in later folklore. They are usually depicted as little bearded men, wearing a coat and hat, who partake in mischief. In later times, they have been depicted as shoe-makers who have a hidden pot of gold at the end of the rainbow. Leprechaun-like creatures rarely appear in Irish mythology and only became prominent in later folklore

Figure 6.3: A Motif Association Miner report generated by the system, for the motif *"Leprechaun"*. The report has been manually formatted for the sake of clarity, but the contents remain unchanged.

CHAPTER 7

CULTURAL COMMONSENSE AND QA SYSTEM

7.1 Motivation

The main motivation behind this study is to distill the knowledge gathered as mentioned in chapter 5 into a form useful to current NLP systems. Specifically, my hope is to transform the cultural knowledge thus collected into a resource that can be used to produce more human-like performance in NLP tasks. In the past few years, there have been major advancements in the field of question answering (QA) systems (Gan and Ng, 2019; Fan and Ferrucci, 2019; Qu et al., 2019; Zafar et al., 2020), in which researchers have looked at different ways in which these systems can be made more accurate and human-like in both their responses as well as their methodology. Incorporating commonsense knowledge and reasoning into NLP systems is one such area of recent focus (Tandon et al., 2018a,b; Trinh and Le, 2018; Merkhofer et al., 2018). Majority of the current work in the field seems to be focused on improving the performance of the existing system in various ways, and they are doing well at achieving those goals. With this, researchers in the field have looked at ways to make these systems better, and incorporating commonsense knowledge into these systems is one important step towards making better systems that are able to feel more ”human.”

The importance of commonsense knowledge bases and repositories is clear from the volume of recent work that makes use of resources such as ConceptNet (Speer, Robyn, 2020; Speer et al., 2016; Speer and Havasi, 2012) to imbue NLP systems with worldly knowledge obtained from humans. A key recent contribution along these lines was ATOMIC (Sap et al., 2019b), which tackles the task of incorporating commonsense reasoning into NLP tasks by generating an atlas of “if-then” rules that taken together produce behavior akin to commonsense reasoning. Work such as ATOMIC, and similar work like

COMET (Bosselut et al., 2019) has made commonsense knowledge more accessible to the current generation of NLP systems; the progress and pitfalls of this work have been cataloged recently (Sap et al., 2020).

However, like I mentioned before (See §5.1), most previous work, even those that try to inject some commonsense into machines like ATOMIC Sap et al. (2019b) or COMET, Bosselut et al. (2019) have a monolithic view of commonsense, which is simply not the case. We know that what we consider commonsense differs from culture to culture. While the eventual task would be to build new systems like ATOMIC for *cultural* commonsense, the first step before that is to see if the concept of cultural commonsense actually helps systems perform better. This is why for this task, I try to enhance an existing QA system using cultural data. While the number of questions we have in the data collection efforts mentioned in chapter 5 is not large enough to provide a definitive working model of a Cultural QA system, it is sufficient for a proof-of-concept to demonstrate that this approach is valid and would pay dividends if performed at a larger scale in the future.

7.2 Methodology

To develop this cultural commonsense QA system, I use the cultural data collected from the ritual survey (see chapter 5). I use the questions of the survey and the responses from the participants of the survey to hand-craft a set of questions and answers. I then use these questions to train and test an existing Question Answering system. I test the questions both on the original version of the QA system, and also on the model retrained on the cultural dataset, and compare the performance of the two models. I discuss more details of the methodology in the subsequent subsections.

7.2.1 Selecting cultural groups

The first major decision to make is to pick which cultural groups should be the focus of this initial work. Unlike the data cultural collection efforts, where I focus only on two main cultural groups, for this study, I focused on three specific target groups—Americans (people from the United States of America), Indians (people from India), and Filipinos (people from the Philippines)—based on a number of factors. Firstly, all three countries use English as one of their major languages, either *de jure* (in case of India and The Philippines) or *de facto* (in case of the US); since our study was to be conducted in English, this was a key requirement.

Secondly, Amazon Mechanical Turk has a high presence of workers from these three countries Difallah et al. (2018). While this does not affect the actual object of improving a QA system, this was vital to be able to collect data to build a dataset. [1]

Moreover, this trio of groups also allows us to have a unique contrast and high degree of cultural variation. The US-Philippines data is a potential source for contrast in terms of variations of Christianity-based cultures; while the India-Philippines data can be a contrast between the differences in Asian cultures. In addition to this, the fact that the United States and India are large and diverse countries consisting of various cultures allows us to capture a varied amount of data in terms of rituals.

Thus I believe this choice of cultural group is appropriate for the first attempt at a CulturalQA system.

[1] As it turned out, I would eventually discard The Philippines for the larger data collection study due to difficulty in finding large number of participants, but I use it here as I had enough data for this task. See §5.2.2.

7.2.2 Selecting cultural markers

For this task, I am using rituals as the markers for culture. I have already explained in detail why I chose rituals as an indicator of culture in §5.2.1. So I omit explaining it again in this section to avoid repetition.

7.2.3 Selecting Rituals

There were a number of different rituals across various categories to choose from as the markers of culture for the purposes of my study. For the reason detailed in §2.2, I picked the following six rituals for this task:

1. **Wedding** (*rite of passage*)
2. **Funeral** (*rite of passage*)
3. **Coming of Age** (*rite of passage*)
4. **Birth** (*rite of passage*)
5. **New Year** (*calendrical rite*)
6. **Birthday** (*rite of passage*)

7.2.4 Collecting cultural data

The data I used for this task is the data I collected using the survey ritual, as explained in chapter 5. Specifically, I use the preliminary data collected during the study, with the only difference being that I also use the data from the Filipino participants for this task. There were a total of 115 unique responses, with each participant limited to one response per ritual per culture. Out of the 115 responses, 38 were from India, another 38 from the Philippines, and 39 from the US. For details on the data collection process, see chapter 5.

7.2.5 Choice of QA System

The approach for this task is to enhance an existing Question Answering system so that it can be enriched with cultural commonsense. So the choice of a QA system becomes very crucial. It is important to choose a system that is close to the current state-of-the-art as much as possible to ensure that any failing of the system on the cultural dataset is due to the *cultural* aspect of it, rather than due to any limitations of the system itself.

There are a lot of different of QA systems to choose from as the field has made a lot of progress over the last few years. However, the UnifiedQA system (Khashabi et al., 2020b) emerged as the clear choice for my task. This is because UnifiedQA outperforms most of the existing question answering systems on several popular QA datasets. The UnifiedQA reports a better performance than the SOTA on the NewQA, (Trischler et al., 2016) ROPES, (Lin et al., 2017) QASC, (Khot et al., 2020) NP-BoolQ, (Khashabi et al., 2020a) and BoolQ-CS (Clark et al., 2019) datasets. It also outperforms the popular systems RetroReader Zhang et al. (2021) and ROBERTa Liu et al. (2019) on multiple tasks.

Moreover, unlike other question answering systems that typically focus on a single type of question answering, this system works on four common different types of question answering: (i) extractive (derive answers as substrings from a context text provided alongside questions); (ii) abstractive (using some model to *understand* context text and provide an answer that is more than just a substring of the text); (iii) multiple choice (picking one answer out of multiple possible options); and (iv) Boolean (answering yes or no to a question, which might or might not include some context text). Picking a system like UnifiedQA also ensures that I don't have to limit myself to a rigid style of question answer set, as the system can process any type of questions.

7.2.6 Creating QA dataset

In order to be able to train the system on the cultural data, I first needed to create a CulturalQA dataset. To do this, I converted the cultural data I collected (see chapter 5 and §7.2.4) into the proper format of question-answers. I detail the process of creating the dataset below. In sum, the CulturalQA dataset I created contained a total of 100 question-answer choice entries.

Generally, one of the issues with question answering systems is that one has to follow a rigid format for the system to work effectively on the data. However, because I am using UnifiedQA as my base model, I had the opportunity to have a slightly more flexible question style, without worrying about the system not being able to perform well on my questions.

Dataset Format

I created the dataset such that it was structured like a multiple choice question. Every entry consisted of a question, and five possible answer choices. There was no additional context provided. However, unlike traditional multiple choice questions, there were more than one possible answer for most of the questions. Primarily, for questions that dealt with numerical answers (eg. number of people or duration of events), there were a range of possible answers based on the user responses of the survey.

To confirm to the UnifiedQA training dataset format, the questions were organized such that each entry contained the question, followed by '\n', followed by the answer choices labelled as (a), (b), (c), etc. An example of the entry is shown in table 7.1.

"What does the QA dataset format look like? \n (a) Something like this (b) Something like that (c) A little like this (d) A little like that"

Table 7.1: A fictitious example of a question-answer entry in the dataset.

Creating Questions

To create the questions for the dataset, I used a combination of automatic techniques and manual curation. First, I created questions templates with the help of the questions that were asked in the main section of the ritual survey (see §5.2.4). I then programmatically reproduced the questions for all three cultural groups as well as the generic version, (e.g. *wedding in India* v/s *wedding in the USA* v/s *wedding in the Philippines* v/s *wedding*). I did this for all six rituals mentioned in §7.2.3.

Table 7.2 shows an example of how the question generation process worked. The '(empty)' indicates appending an empty string to the end, which corresponds to the generic questions. For instance, taking the first rows of each column, the question "How many people typically attend wedding?" would be constructed, while taking the last rows of each column would generate "How many people typically attend birthday in the Philippines?"

As we can see, the questions generated need some work. So after automatic generation of questions, I reviewed them manually, editing for grammar and readability. I also removed the questions for which there was not an agreed-upon answer in the survey.

<table>
<tr><td rowspan="12">How many people typically attend</td><td rowspan="2">wedding</td><td rowspan="3">(empty)</td></tr>
<tr></tr>
<tr><td rowspan="2">birth ceremony</td></tr>
<tr><td rowspan="3">in the US</td></tr>
<tr><td rowspan="2">funeral</td></tr>
<tr></tr>
<tr><td rowspan="2">coming-of-age</td><td rowspan="3">in India</td></tr>
<tr></tr>
<tr><td rowspan="2">new year's celebration</td></tr>
<tr><td rowspan="3">in the Philippines</td></tr>
<tr><td rowspan="2">birthday</td></tr>
<tr></tr>
</table>

Table 7.2: An example of how the combination process works for question creation for the Cultural QA dataset.

Creating Answer Choices

For the answer choices, I first picked the correct answers based on the responses of the participants. For questions like the duration of an event or the number of people involved, I took the median of the responses from the participants, and selected a range around it as the correct answer. For some questions that required subjective analysis of the responses, I studied the responses manually to find the answer that represented most of the responses.

I also added the correct answers for other cultures as other possible answer choices. For instance, the correct *number of people that typically attend an Indian wedding* was listed as an answer choice for *the number of people that typically attend an American one*. Because some of the rituals are similar across cultures, it was possible that there were multiple choices that could be correct for some questions. The remaining answer choices were created such that they were similar but distinct to the other choices. Table 7.3 shows an example question with the answer choices listed.

Question:	Who is the important person involved in the funeral service in India?
Answer Choices:	(a) The priest (b) The parents of the deceased (c) Family (d) The Son of the deceased (e) The Spouse of the deceased
Question:	How long does a coming-of-age ceremony last in the USA?
Answer Choices:	(a) 2 hours (b) 4 hours (c) One day (d) 1 hour (e) 1 week
Question:	How many people attend the coming-of-age ceremony in the Philippines?
Answer Choices:	(a) 200 (b) 5 (c) 10 (d) 50 (e) 20

Table 7.3: Examples of the question-answer pair created for this system. The correct answer choice(s) are shown in green. As explained above, some questions may have more than one correct answer.

7.2.7 Improving QA system

In order to improve the UnifiedQA system by injecting cultural commonsense, I needed to retrain their model on my data. UnifiedQA lists has two different models that are available for training: the one based on the T5 model, (Raffel et al., 2019) and the one based on the BART (Lewis et al., 2020) model. For this task, I used the BART model of the UnifiedQA system. Both T5 and BART are highly popular and good-performing pre-trained models used for various NLP tasks including Question Answering. They have shown to be highly effective in NLP tasks, producing similar or comparable results to the SOTA on popular test datasets in the field like SQuAD (Rajpurkar et al., 2016) and GLUE (Wang et al., 2018). So it came to the matter of preference as to which model to choose as the base model for my task. As these models are fairly comparable, for the purposes of prototyping, I chose BART due to its additional advantage of ease of adaptation.

I retrained the BART model using the cultural QA dataset I created. I used Python version 3.6.12, and PyTorch Paszke et al. (2019) and Transformers Wolf et al. (2019) for this task, and used UnifiedQA's best performing model checkpoint as provided by their authors as the starting point. I used a 60/15/25 train/dev/test split for the data. It was trained with a batch size of 32. [2]

7.3 Results

7.3.1 Evaluation Method

To evaluate the performance of the CulturalQA, I compared it with the existing system. So I also tested the QA dataset questions on the best performing model of UnifiedQA-BART.

[2]These parameters don't affect the results of the training, but rather are chosen based on the GPU availability.

Additionally, I also tested the dataset on UnifiedQA's T5 (Raffel et al., 2019) based large model. For T5 based models, UnifiedQA provides a small model (60 million parameters) and a large model (770 million parameters). Since it reports that the large model is the better performing of the two, I use the large model for comparison.

Unfortunately, the automatic evaluation of these results are still a work in progress. (Liu et al., 2016) It has also recently been shown that existing automatic metrics (e.g. span-based F1) often do not properly assess the performance of QA systems, but rather simply focus on the commonality of n-gram present in the correct answer versus the produced answer, and so human judgement is the best evaluation of these results. (Chen et al., 2019). This is why, also owing to the fact that some of these questions have multiple possible answers, and the small size of the dataset, I was able have human evaluation of all the results to check whether the answer provided by each system were correct or not.

For the human evaluation, two different evaluators separately reviewed the answers to the questions generated by all three models – UnifiedQA-BART, UnifiedQA-Large and CulturalQA (the model trained on the Cultural QA dataset) – and compared it to the answers from the survey. The names of the models were not displayed at the time of evaluation to ensure the evaluators did not know which set of answers came from which model. For answers with partial match, anything within a margin of +/- 20% were accepted as correct (e.g. if an event was said to last 4 hours, 3 or 5 hours were also deemed acceptable answers). The evaluators agreed on all the judgements, which indicates there were no questions in the dataset that had ambiguous answers in the results of the survey.

7.3.2 Results

Here I compare the results of the results of my model explained in this chapter (CulturalQA) with that of the two UnifiedQA's models: UnifiedQA-BART and UnifiedQA-

Large. The results are shown in table 7.4. We can see that my CulturalQA model significantly outperforms the UnifiedQA-BART model on which it was trained, with 80% of the answers correct compared to just 44% of the BART model. It also outperforms the UnifiedQA-Large model, which got 72% of the answers correct. A statistical analysis confirmed that the the improvement from the UnifiedQA-BART model to the CulturalQA model system was statistically significant, but the one betweem UnifiedQA-Large and CulturalQA was not. However, this is not too much of an issue in terms of the performance of the system, for the CulturalQA was not built on top of the UnifiedQA-Large system, and it does not preclude the possibility that a system built top of the *Large* model will surpass it like this current Cultural model did for the system it was built on.

Model	Correct Answers (Count)	Correct Answers (%)
UnifiedQA-BART	11/25	44% ±0.67%
UnifiedQA-Large	18/19	72% ±0.98%
CulturalQA (This system)	20/25	**80%±0.57%**

Table 7.4: Results comparing the performance of UnifiedQA BART model v/s UnifiedQA Large model v/s CulturalQA model on culturally sensitive questions. The error range shown are standard errors.

7.3.3 Discussion

We see that the system retrained on cultural data clearly outperforms the existing SOTA systems. While it would be ideal to have a large dataset, it is nevertheless clear that the existing SOTA question answering systems perform poorly on culturally sensitive questions. This makes it clear that there is much room for improvement for QA systems when it comes to culturally sensitive questions.

On the other hand, the results of the CulturalQA system shows that there seems to be a clear method to improve the performance of these systems. And while the collection and creation of culturally sensitive dataset might not be the easiest task when it comes to creating a QA system, this work shows that it is possible to do so.

Additionally, we see that while the UnifiedQA's Large model outperforms its BART model on the cultural QA dataset significantly, the Cultural model still reports impressive performance despite being trained on the worse-performing BART model. So it would not be unreasonable to hypothesize that using the UnifiedQA Large as the base would make the system even better. In the future, I intend to do just that: explore using the T5-based Large model as a base to determine if it will result in even better performance.

7.4 Limitations and Future Work

This current work is to be taken as a proof-of-concept that existing QA system under-perform on culturally sensitive questions but can be improved with appropriate steps. So naturally there are a lot of improvements and enhancements that can be done in the future.

The obvious, immediate future direction is to test this dataset on different other base models – including the T5 model as I mentioned in the previous section – to see if that improves the performance of the cultural model even more. In addition to UnifiedQA, the dataset can also be used to retrain other question answering system models to see if we can achieve a better performance.

The first major limitation of the work is that the Cultural QA dataset that I created is fairly small. It is important to scale up this work to a dataset of thousands of question-answer pairs. Unfortunately, creating such a large dataset proved to be beyond my means during the course of the project. So this is an obvious first future direction for this work. Once the dataset shows better performance on existing systems, the next step would be to then train a new model on this data to create a new, true Cultural QA system. In fact, a larger dataset might even enable to construct a cultural commonsense knowledge systems by themselves, not unlike a *cultural* ATOMIC. (Sap et al., 2019b)

The other improvement that can be done for this QA system is that we can make it such that the system gives multiple answers for a question instead of just one, each pertaining to a certain cultural group. Similarly, another way the system might be able to be improved is by using other culturally sensitive topics apart from rituals.

In addition to the improvements to the system itself, I have also seen a need for better performance metrics for these systems. The current metrics do a very poor job of assessing the performance of these systems, (Chen et al., 2019) meaning we get numerical progress while not being close to solving the problem in the real world.

Finally, another future direction, of course, is to collect the data for a variety of cultures around the world, which would lead to the ultimate goal of having a QA system that is truly sensitive to all major cultures. But as I discuss across this entire thesis, this is an issue for all AI systems, not just QA systems.

CHAPTER 8

CONCLUSION

We need to incorporate cultural knowledge into AI systems if we want to be on the path to a truly universal AI. But before we get to that stage, the essential question to ask was if we can collect, encode, and use cultural knowledge in such a manner. To be able to use cultural knowledge, we need to first have data grounded in real life, that comes from actual people who hold this knowledge. We need to think about what kind of data we are collecting, and make an active effort to balance it across cultural biases. I have tried, to the best of my abilities, to do just that with my work here. Overall, I have demonstrated the following main things with my dissertation work.

Demonstrate Culture-Specific Knowledge Exists

In chapter 3, I described the details of the cognitive psychological experiment I conducted to test if motifs are truly understood more by people in the cultural group of origin compared to those outside of the said cultural group. Using the psychological experiment that I conducted, I was able to demonstrate that there exists culture-specific knowledge that carry a host of information. Using motifs as my cultural marker, I showed that despite reading the same text, people in-culture v/s out-culture understand the meaning at different levels due to the wealth of culture-specific information present in the text.

As an additional output of this work, I also provide a road-map for cultural psycholinguistic studies to be carried out in the future, having improved the study over multiple iterations.

Demonstrate Cultural Knowledge is Recognizable across cultural boundaries

In chapter 4, I lay out my work that proves that even when people outside a culture might not be aware of the true meaning of cultural motifs, they are nonetheless aware that these carry significant information that they are missing. The results of the study show that for most motifs, people seem to be aware that there is some hidden meaning behind these words, and that this effect is distinct from the fact that the motifs are simply words they haven't encountered before.

Demonstrate that cultural knowledge can be quantified and collected

Using rituals as the cultural marker this time, I test a technique to collect cultural knowledge. As described in chapter 5, I successfully demonstrate that cultural knowledge can be collected such that they can be used for AI and other computational systems.

Collect cultural knowledge that's useful for AI systems

Using this aforementioned technique, I collect large scale cultural knowledge data, as reported in §5.5. I will be releasing the data for public use in the future, which will ensure that the data will be useful not only for myself for future use, but also for any other researchers in the field who might want to take steps towards making their system more balanced.

Make cultural knowledge available across cultural boundaries

I demonstrated that we can build systems that brings forth specific cultural knowledge to the wider audience. I describe in chapter 6 how I built a Motif Association Miner, a system that can collect information about culture-specific motifs, and create a simple, easy-to-read report that is helpful for people out-of-culture to understand these motifs they might come across in text. This makes sure that people who would otherwise have none or partial knowledge about the motifs can now understand them and the meaning conveyed more clearly.

In addition to the main system, the curated Motif Catalog (see §6.4.12) used for this project could be the first step towards a large-scale motif information repository.

Demonstrate that cultural knowledge can improve existing systems

Finally, in chapter 7, I offer a proof-of-concept of how cultural knowledge can be used to enhance existing systems, by using the ritual data I collected to train and improve a state-of-the-art Question Answering system. The results show that the system was indeed significantly improved by my cultural dataset.

To conclude, I would like to go back to the main point that I stated in the beginning of this document, about how current AI systems tend to generally be biased to the way *Western*, and more typically North American society works. I proposed the need to incorporate cultural knowledge into AI systems. Unlike in the past, researchers in the field now do acknowledge that this is a fact. However, more often than not work in our field tends to try and fix these issues at a superficial level, rather than trying to find a way to build systems that are truly sensitive to culture. With the work described in this document, I show that if we start from the ground up, and put an effort to actually carefully collect and understand cultural differences and nuances across various groups, we can truly make strides towards having a more universal AI. Furthermore, I hypothesize that similar approaches might be feasible to combat other biases in AI, like gender, race, etc. It is not quick, and it is certainly not easy, but it needs to be done if we want a *true* AI in the future. And it can be done, as demonstrated by this work.

BIBLIOGRAPHY

Anurag Acharya, Kartik Talamadupula, and Mark A. Finlayson. Toward an atlas of cultural commonsense for machine reasoning. In *Proceedings of the Workshop on Common Sense Knowledge Graphs (CSKGs)*, Online, February 2021. URL `https://usc-isi-i2.github.io/AAAI21workshop/papers/Acharya_CSKGsAAAI-21.pdf`. Held in conjunction with the 35th AAAI Conference on Artificial Intelligence (AAAI 2021).

Eneko Agirre and Philip Edmonds. *Word sense disambiguation: Algorithms and applications*, volume 33. Springer Science & Business Media, 2007.

R.E. Alegria, R.E. Alegría, L. Homar, and E. Culbert. *The Three Wishes: A Collection of Puerto Rican Folktales*. Harcourt, Brace & World, 1969. ISBN 9780152868710. URL `https://books.google.ca/books?id=cq5QAAAAMAAJ`.

Talal Asad. Toward a genealogy of the concept of ritual. *A: Baltimore*, pages 55–79, 1993.

Saho Ayabe-Kanamura, Ina Schicker, Matthias Laska, Robyn Hudson, Hans Distel, Tatsu Kobayakawa, and Sachiko Saito. Differences in perception of everyday odors: a japanese-german cross-cultural study. *Chemical senses*, 23(1):31–38, 1998.

Olga Babko-Malaya. Propbank annotation guidelines. 2005.

Anson Bastos, Abhishek Nadgeri, Kuldeep Singh, Isaiah Onando Mulang, Saeedeh Shekarpour, Johannes Hoffart, and Manohar Kaul. Recon: relation extraction using knowledge graph context in a graph neural network. In *Proceedings of the Web Conference 2021*, pages 1673–1685, 2021.

Gary Beckman. *Hittite birth rituals*. Otto Harrassowitz, 1983.

Catherine Bell. *Ritual theory, ritual practice*. Oxford University Press, Oxford, 1992.

Catherine M Bell. *Ritual: Perspectives and dimensions*. Oxford University Press on Demand, Oxford, 1997.

John W Berry and Pierre R Dasen. *Culture and cognition: Readings in cross-cultural psychology*. Routledge, 2019.

Parminder Bhatia, Yangfeng Ji, and Jacob Eisenstein. Better document-level sentiment analysis from RST discourse parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2212–2218, Lisbon, Portugal, September 2015. Association for Computational Linguistics. doi: 10.18653/v1/D15-1263. URL `https://aclanthology.org/D15-1263`.

Nikita Bhutani, HV Jagadish, and Dragomir Radev. Nested propositions in open information extraction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 55–64, 2016.

Inger Margrethe Boberg. *Motif-index of early Icelandic literature*. Munksgaard, 1966.

Michael Boratko, Harshit Padigela, Divyendra Mikkilineni, Pritish Yuvraj, Rajarshi Das, Andrew McCallum, Maria Chang, Achille Fokoue-Nkoutche, Pavan Kapanipathi, Nicholas Mattei, Ryan Musa, Kartik Talamadupula, and Michael Witbrock. A systematic classification of knowledge, reasoning, and context within the arc dataset. In *Proceedings of the Workshop on Machine Reading for Question Answering (MRQA) at ACL 2018*, pages 60–70, Melbourne, Australia, 2018.

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. Comet: Commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4762–4779, Florence, Italy, 2019.

Richard Y Bourhis and Itesh Sachdev. Vitality perceptions and language attitudes: Some canadian data. *Journal of language and Social Psychology*, 3(2):97–126, 1984.

Richard Yvon Bourhis, Howard Giles, and Doreen Rosenthal. Notes on the construction of a 'subjective vitality questionnaire'for ethnolinguistic groups. *Journal of Multilingual and Multicultural Development*, 1981.

Ian Brace. *Questionnaire design: How to plan, structure and write survey material for effective market research*. Kogan Page Publishers, 2018.

Willi Braun and Russell McCutcheon. *Guide to the Study of Religion*. Bloomsbury Publishing, London, 2000.

Glynis M Breakwell, Sean Ed Hammond, Chris Ed Fife-Schaw, and Jonathan A Smith. *Research methods in psychology*. Sage Publications, Inc, 2006.

Rui Cai, Xiaodong Zhang, and Houfeng Wang. Bidirectional recurrent convolutional neural network for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 756–765, 2016.

Norma E Cantú. La quinceañera: Towards an ethnographic analysis of a life-cycle ritual. *Southern Folklore*, 56(1):73, 1999.

Barbara Carretti, Erika Borella, Michela Zavagnin, and Rossana de Beni. Gains in language comprehension relating to working memory training in healthy older adults. *International journal of geriatric psychiatry*, 28(5):539–546, 2013.

Matthias Cetto, Christina Niklaus, André Freitas, and Siegfried Handschuh. Graphene: Semantically-linked propositions in open information extraction. *arXiv preprint arXiv:1807.11276*, 2018.

Nathanael Chambers and Dan Jurafsky. Template-based information extraction without the templates. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, pages 976–986, 2011.

Anthony Chen, Gabriel Stanovsky, Sameer Singh, and Matt Gardner. Evaluating question answering evaluation. In *Proceedings of the 2nd Workshop on Machine Reading for Question Answering*, pages 119–124, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-5817. URL `https://aclanthology.org/D19-5817`.

José M. Chenlo, Alexander Hogenboom, and David E. Losada. Rhetorical structure theory for polarity estimation: An experimental study. *Data & Knowledge Engineering*, 94:135–147, 2014. ISSN 0169-023X. doi: https://doi.org/10.1016/j.datak.2014.07.009. URL `https://www.sciencedirect.com/science/article/pii/S0169023X14000688`. Special issue following the 18th International Conference on Applications of Natural Language Processing to Information Systems (NLDB'13).

Barbara Jo Chesser. Analysis of wedding rituals: An attempt to make weddings more meaningful. *Family Relations*, pages 204–209, 1980.

Janara Christensen, Stephen Soderland, and Oren Etzioni. An analysis of open information extraction based on semantic role labeling. In *Proceedings of the sixth international conference on Knowledge capture*, pages 113–120, 2011.

Christopher Clark, Kenton Lee, Ming-Wei Chang, Tom Kwiatkowski, Michael Collins, and Kristina Toutanova. Boolq: Exploring the surprising difficulty of natural yes/no questions. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2924–2936, 2019.

J Clark and T Barrows. College students' knowledge and beliefs: A survey of global understanding. Technical report, Educational Testing Service, 1981.

Peter Clark. Elementary school science and math tests as a driver for ai: take the aristo challenge! In *AAAI*, pages 4019–4021. Citeseer, 2015.

Peter Clark, Isaac Cowhey, Oren Etzioni, Tushar Khot, Ashish Sabharwal, Carissa Schoenick, and Oyvind Tafjord. Think you have solved question answering? try arc, the ai2 reasoning challenge. *arXiv preprint arXiv:1803.05457*, 2018.

Geertz Cliford. The interpretation of cultures: Selected essays, 1973.

Jacob Cohen. *Statistical power analysis for the social sciences*. Hillsdale, NJ: Erlbaum, 1988.

Tom Peete Cross. *Motif-index of early Irish literature*. Indiana University, 1952.

Meredyth Daneman and Patricia A Carpenter. Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, 19(4):450–466, 1980.

Sándor Darányi. Examples of Formulaity in Narratives and Scientific Communication. In *Proceedings of the First International AMICUS Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts*, pages 29–35, 2010. URL `http://ilk.uvt.nl/amicus/amicus{_}ws2010{_}proceedings.html`.

Sándor Darányi and László Forró. Detecting Multiple Motif Co-occurrences in the Aarne-Thompson-Uther Tale Type Catalog: A Preliminary Survey. *Anales de Documentación*, 15(1), 2012. URL `http://revistas.um.es/analesdoc/article/view/analesdoc.15.1.134691/131801`.

Sándor Darányi, Peter Wittek, and László Forró. Toward Sequencing "Narrative DNA": Tale Types, Motif Strings and Memetic Pathways. In Mark A. Finlayson, editor, *Third Workshop on Computational Models of Narrative (CMN)*, pages 2–10, Istanbul, Turkey, 2012. European Language Resources Association (ELRA).

Ernest Davis and Gary Marcus. Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the ACM*, 58(9):92–103, 2015.

Kenneth Dean. Ritual and space. *Civil Society in China, Armonk: ME Sharpe*, pages 172–192, 1997.

Thierry Declerck and Piroska Lendvai. Linguistic and semantic representation of the thompson's motif-index of folk-literature. In *Research and Advanced Technology for Digital Libraries*, pages 151–158. Springer, 2011.

Thierry Declerck, Piroska Lendvai, and Sándor Darányi. Multilingual and Semantic Extension of Folk Tale Categories. In *Proceedings of the 2012 Digital Humanities Conference (DH 2012)*, 2012. URL http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/multilingual-and-semantic-extension-of-folk-tale-catalogues/.

Nathal M Dessing. *Rituals of birth, circumcision, marriage, and death among Muslims in the Netherlands*, volume 2. Peeters Publishers, 2001.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Djellel Difallah, Elena Filatova, and Panos Ipeirotis. Demographics and dynamics of mechanical turk workers. In *Proceedings of the eleventh ACM international conference on web search and data mining*, pages 135–143, Los Angeles, CA, 2018.

Paul DiMaggio. Culture and cognition. *Annual review of sociology*, 23, 1997.

Zoltán Dörnyei. *Research methods in applied linguistics*. Oxford university press, 2007.

Alan Dundes. The motif-index and the tale type index: A critique. *Journal of Folklore Research*, pages 195–202, 1997.

Matthew Dunn, Levent Sagun, Mike Higgins, V Ugur Guney, Volkan Cirik, and Kyunghyun Cho. Searchqa: A new q&a dataset augmented with context from a search engine. *arXiv preprint arXiv:1704.05179*, 2017.

Emile Durkheim and Joseph Ward Swain. *The elementary forms of the religious life*. Courier Corporation, London, 2008.

Jay Earley. An efficient context-free parsing algorithm. *Communications of the ACM*, 13 (2):94–102, 1970.

John Ellson, Emden R Gansner, Eleftherios Koutsofios, Stephen C North, and Gordon Woodhull. Graphviz and dynagraph—static and dynamic graph drawing tools. In *Graph drawing software*, pages 127–148. Springer, Berlin, 2004.

Mica R Endsley and Daniel J Garland. *Situation awareness analysis and measurement*. CRC Press, 2000.

Jose Espinosa and Henry Lieberman. Eventnet: Inferring temporal relations between commonsense events. In *Mexican International Conference on Artificial Intelligence*, pages 61–69. Springer, 2005.

James J Fan and David A Ferrucci. Scoring candidates using structural information in semi-structured documents for question answering systems, March 5 2019. US Patent 10,223,441.

Vanessa Wei Feng and Graeme Hirst. Text-level discourse parsing with rich linguistic features. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 60–68, 2012.

Seeger Fisher and Brian Roark. The utility of parse-derived features for automatic discourse segmentation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 488–495, 2007.

Anette Frank and Nils Reiter. Harvesting event chains in ritual descriptions using frame semantics. In *Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts*, page 75, 2010.

Paulo Freire and Donaldo Macedo. A dialogue: Culture, language, and race. *Harvard educational review*, 65(3):377–403, 1995.

Galia Fuchs, Arie Reichel, et al. Cultural differences in tourist destination risk perception: An exploratory study. *Tourism (Zagreb)*, 52(1):21–37, 2004.

Wee Chung Gan and Hwee Tou Ng. Improving the robustness of question answering systems to question paraphrasing. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6065–6075, Florence, Italy, 2019.

Zhengjie Gao, Ao Feng, Xinyu Song, and Xi Wu. Target-dependent sentiment classification with bert. *IEEE Access*, 7:154290–154299, 2019. doi: 10.1109/ACCESS.2019.2946594.

Kiril Gashteovski, Rainer Gemulla, and Luciano del Corro. MinIE: Minimizing facts in open information extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2630–2640, Copenhagen, Denmark, September 2017. Association for Computational Linguistics. doi: 10.18653/v1/D17-1278. URL `https://aclanthology.org/D17-1278`.

Clifford Geertz and Michael Banton. Religion as a cultural system. In Michael Banton, editor, *Anthropological Approaches to the Study of Religion*, chapter 4, pages 87–125. Tavistock London, London, 1966.

Bahriye Selin Gokcesu. Comparison, categorization, and metaphor comprehension. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 31, 2009.

Leo A Goodman. Snowball sampling. *The annals of mathematical statistics*, pages 148–170, 1961.

Jack Goody. Religion and ritual: the definitional problem. *The British Journal of Sociology*, 12(2):142–164, 1961.

Andrew S Gordon and Jerry R Hobbs. *A formal theory of commonsense psychology: How people think people think*. Cambridge University Press, Cambridge, 2017.

John N Gray. Keep the hom fires burning: sacrifice in nepal. *Social Analysis: The International Journal of Social and Cultural Practice*, February(1):81–107, 1979.

Evelyn Marcussen Hatch and Anne Lazaraton. *The research manual: Design and statistics for applied linguistics*. New York, NY: Newbury House Publishers, 1991.

Osama Hazzi and Issa Maldaon. A pilot study: Vital methodological issues. *Business: Theory and Practice*, 16(1):53–62, 2015.

Bas Heerschop, Frank Goossen, Alexander Hogenboom, Flavius Frasincar, Uzay Kaymak, and Franciska de Jong. Polarity analysis of texts using discourse structure. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, page 1061–1070, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450307178. doi: 10.1145/2063576.2063730. URL https://doi.org/10.1145/2063576.2063730.

Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. In *Advances in neural information processing systems*, pages 1693–1701, 2015.

Hugo Hernault, Helmut Prendinger, David A du Verle, and Mitsuru Ishizuka. Hilda: A discourse parser using support vector machine classification. *Dialogue & Discourse*, 1 (3):1–33, 2010.

Lynette Hirschman, Marc Light, Eric Breck, and John D Burger. Deep read: A reading comprehension system. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics*, pages 325–332, 1999.

Ying-yi Hong, Michael W Morris, Chi-yue Chiu, and Veronica Benet-Martinez. Multicultural minds: A dynamic constructivist approach to culture and cognition. *American psychologist*, 55(7):709, 2000.

Matthew Honnibal and Ines Montani. spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. *To appear*, 7(1), 2017.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. spaCy: Industrial-strength Natural Language Processing in Python. *Online code repository*, 2020. doi: 10.5281/zenodo.1212303.

A. Hurley, C.R.R. de Arellano, and S.R. Lamarche. *The Mythology and Religion of the Tainos*. Independently Published, 2021. ISBN 9798747575653. URL `https://books.google.ca/books?id=o02MzgEACAAJ`.

Nancy Ide and Jean Véronis. Introduction to the special issue on word sense disambiguation: the state of the art. *Computational linguistics*, 24(1):1–40, 1998.

Hiroko Ikeda. *A type and motif index of Japanese folk-literature*. Orient Cultural Service, 1971.

Glenn D Israel. Determining sample size. 1992.

Susan Jamieson. Likert scales: How to (ab) use them? *Medical education*, 38(12): 1217–1218, 2004.

Anja Jamrozik, Eyal Sagi, Micah Goldwater, and Dedre Gentner. Relational words have high metaphoric potential. In *Proceedings of the First Workshop on Metaphor in NLP*, pages 21–26, 2013.

Ludovic Jean-Louis, Romaric Besançon, and Olivier Ferret. Text segmentation and graph-based method for template filling in information extraction. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 723–731, 2011.

Yangfeng Ji and Jacob Eisenstein. Representation learning for text-level discourse parsing. In *Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 1: Long papers)*, pages 13–24, 2014.

Mandar Joshi, Eunsol Choi, Daniel S Weld, and Luke Zettlemoyer. Triviaqa: A large scale distantly supervised challenge dataset for reading comprehension. *arXiv preprint arXiv:1705.03551*, 2017.

Shafiq Joty, Giuseppe Carenini, and Raymond T Ng. Codra: A novel discriminative framework for rhetorical analysis. *Computational Linguistics*, 41(3):385–435, 2015.

D. Khashabi, T. Khot, and A. Sabhwaral. Natural perturbation for robust question answering. *arXiv preprint*, 2020a.

Daniel Khashabi, Sewon Min, Tushar Khot, Ashish Sabharwal, Oyvind Tafjord, Peter Clark, and Hannaneh Hajishirzi. Unifiedqa: Crossing format boundaries with a single qa system. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1896–1907, 2020b.

Tushar Khot, Peter Clark, Michal Guerquin, Peter Jansen, and Ashish Sabharwal. Qasc: A dataset for question answering via sentence composition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8082–8090, 2020.

Byoungjae Kim, KyungTae Chung, Jeongpil Lee, Jungyun Seo, and Myoung-Wan Koo. A bi-lstm memory network for end-to-end goal-oriented dialog learning. *Computer Speech & Language*, 53:217–230, 2019.

Bacil F Kirtley. *A motif-index of traditional Polynesian narratives*. University of Hawai'i Press, 1971.

Dan Klein and Christopher D Manning. Accurate unlexicalized parsing. In *Proceedings of the 41st annual meeting of the association for computational linguistics*, pages 423–430, 2003.

Claire Kramsch. Language and culture. *AILA review*, 27(1):30–55, 2014.

Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. Building machines that learn and think like people. *Behavioral and brain sciences*, 40:e253, 2017.

Edward P Lazear. Culture and language. *Journal of political Economy*, 107(S6):S95–S126, 1999.

Edmund Ronald Leach. *A runaway world?* Oxford University Press, 1968.

Douglas B Lenat. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38, 1995.

Hector Levesque, Ernest Davis, and Leora Morgenstern. The winograd schema challenge. In *Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning*. Citeseer, 2012.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, 2020.

Jiwei Li, Rumeng Li, and Eduard Hovy. Recursive deep models for discourse parsing. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2061–2069, 2014.

Nai Li and Gill Kirkup. Gender and cultural differences in internet use: A study of china and the uk. *Computers & Education*, 48(2):301–317, 2007.

Hongyu Lin, Le Sun, and Xianpei Han. Reasoning with heterogeneous knowledge for commonsense machine comprehension. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2032–2043, 2017.

Ziheng Lin, Min-Yen Kan, and Hwee Tou Ng. Recognizing implicit discourse relations in the penn discourse treebank. In *Proceedings of the 2009 conference on empirical methods in natural language processing*, pages 343–351, 2009.

Lia Litosseliti. *Research methods in linguistics*. Bloomsbury Publishing, 2018.

Chia-Wei Liu, Ryan Lowe, Iulian V Serban, Michael Noseworthy, Laurent Charlin, and Joelle Pineau. How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. *arXiv preprint arXiv:1603.08023*, 2016.

Hugo Liu and Push Singh. Conceptnet—a practical commonsense reasoning tool-kit. *BT technology journal*, 22(4):211–226, 2004.

Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. K-bert: Enabling language representation with knowledge graph. In *AAAI*, pages 2901–2908, 2020.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

Howard Lune and Bruce L Berg. *Qualitative research methods for the social sciences*. Pearson, 2017.

Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60, 2014.

Gary Marcus. Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*, 2018.

Jennifer Mason. *Qualitative researching*. sage, 2017.

Zoe Matthews, Jayashree Ramakrishna, Shanti Mahendra, Asha Kilaru, and Saraswathy Ganapathy. Birth rights and rituals in rural south india: care seeking in the intrapartum period. *Journal of Biosocial Science*, 37(4):385–411, 2005.

Tim May. *Social research: Issues, methods and research*. McGraw-Hill Education (UK), 2011.

Elizabeth Merkhofer, John Henderson, David Bloom, Laura Strickhart, and Guido Zarrella. Mitre at semeval-2018 task 11: Commonsense reasoning without commonsense knowledge. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 1078–1082, 2018.

Martina Miliani, Lucia C Passaro, and Alessandro Lenci. Text frame detector: Slot filling based on domain knowledge bases. In *CLiC-it*, 2019.

George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

Michelle R Nelson and Sameer Deshpande. Love without borders: An examination of cross-cultural wedding rituals. In *Contemporary Consumption Rituals*, pages 151–174. Psychology Press, 2004.

Thanh-Tung Nguyen, Xuan-Phi Nguyen, Shafiq Joty, and Xiaoli Li. Rst parsing from scratch. *arXiv preprint arXiv:2105.10861*, 2021.

Kristoffer L Nielbo and Jesper Sørensen. Prediction error during functional and non-functional action sequences: A computational exploration of ritual and ritualized event processing. *Journal of Cognition and Culture*, 13(3-4):347–365, 2013.

Kristoffer L Nielbo, Donald M Braxton, and Afzal Upal. Computing religion: A new tool in the multilevel analysis of religion. *Method & Theory in the Study of Religion*, 24(3): 267–290, 2012.

Dov Neuman Noy. *Motif-index of Talmudic-Midrashic literature*. Indiana University, 1954.

Nir Ofek, Sándor Darányi, and Lior Rokach. Linking Motif Sequences with Tale Types by Machine Learning. In Mark A Finlayson, Bernhard Fisseni, Benedikt Löwe, and

Jan Christoph Meister, editors, *Proceedings of the 4th Workshop on Computational Models of Narrative (CMN'13)*, volume 32, pages 166–182, Hamburg, Germany, 2013. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi: http://dx.doi.org/10.4230/OASIcs.CMN.2013.166.

Alexandra Olteanu, Onur Varol, and Emre Kiciman. Distilling the outcomes of personal experiences: A propensity-scored analysis of social media. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 370–386, Portland, OR, 2017.

AN Oppenheim. *Questionnaire design, interviewing and attitude measurement*. Pinter Publishers, 1992.

Harinder Pal et al. Demonyms and compound relational nouns in nominal open ie. In *Proceedings of the 5th Workshop on Automated Knowledge Base Construction*, pages 35–39, 2016.

Martha Palmer, Daniel Gildea, and Nianwen Xue. Semantic role labeling. *Synthesis Lectures on Human Language Technologies*, 3(1):1–103, 2010.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70:153–163, 2017.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Pro-*

cessing (EMNLP), pages 1532–1543, 2014. URL http://www.aclweb.org/anthology/D14-1162.

Matthew E. Peters, Waleed Ammar, Chandra Bhagavatula, and Russell Power. Semi-supervised sequence tagging with bidirectional language models. In *ACL*, 2017.

Chen Qu, Liu Yang, W Bruce Croft, Falk Scholer, and Yongfeng Zhang. Answer interaction in non-factoid question answering systems. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, pages 249–253, Glasgow, Scotland, 2019.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683*, 2019.

Altaf Rahman and Vincent Ng. Resolving complex cases of definite pronouns: the winograd schema challenge. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 777–789, 2012.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*, 2016.

J. Ramírez-Rivera, B. Klein, and J. Slemko. *Puerto Rican Tales: Legends of Spanish Colonial Times*. Ediciones Libero, 1977. ISBN 9780960170036. URL https://books.google.ca/books?id=XaJtAAAAMAAJ.

Guozheng Rao, Weihang Huang, Zhiyong Feng, and Qiong Cong. Lstm with sentence representations for document-level sentiment classification. *Neurocomputing*, 308:49–57, 2018. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2018.04.

045. URL https://www.sciencedirect.com/science/article/pii/S092523121830479X.

Hannah Rashkin, Maarten Sap, Emily Allaway, Noah A Smith, and Yejin Choi. Event2mind: Commonsense inference on events, intents, and reactions. *arXiv preprint arXiv:1805.06939*, 2018.

Nils Reiter, Oliver Hellwig, Anand Mishra, Anette Frank, and Jens Burkhardt. Using nlp methods for the analysis of rituals. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, 2010.

Nils Reiter, Oliver Hellwig, Anette Frank, Irina Gossmann, Borayin Maitreya Larios, Julio Rodrigues, and Britta Zeller. Adapting nlp tools and frame-semantic resources for the semantic analysis of ritual descriptions. In Caroline Sporleder, Antal van den Bosch, and Kalliopi Zervanou, editors, *Language Technology for Cultural Heritage*, pages 171–193, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ISBN 978-3-642-20227-8.

Matthew Richardson, Christopher JC Burges, and Erin Renshaw. Mctest: A challenge dataset for the open-domain machine comprehension of text. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 193–203, Seattle, WA, 2013.

Sophie Roche and Sophie Hohmann. Wedding rituals and the struggle over national identities. In *The Transformation of Tajikistan*, pages 141–156. Routledge, 2013.

Melissa Roemmele, Cosmin Adrian Bejan, and Andrew S Gordon. Choice of plausible alternatives: An evaluation of commonsense causal reasoning. In *AAAI spring symposium: logical formalizations of commonsense reasoning*, pages 90–95, 2011.

Peter H Rossi, James D Wright, and Andy B Anderson. *Handbook of survey research*. Academic press, 2013.

Natalya Sadomskaya and Tamara Dragadze. Soviet anthropology and contemporary rituals [with discussion]. *Cahiers du Monde russe et sovietique*, pages 245–255, 1990.

Swarnadeep Saha, Harinder Pal, et al. Bootstrapping for numerical open ie. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 317–323, 2017.

Swarnadeep Saha et al. Open information extraction from conjunctive sentences. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2288–2299, 2018.

Maarten Sap, Saadia Gabriel, Lianhui Qin, Dan Jurafsky, Noah A Smith, and Yejin Choi. Social bias frames: Reasoning about social and power implications of language. *arXiv preprint arXiv:1911.03891*, 2019a.

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A Smith, and Yejin Choi. Atomic: An atlas of machine commonsense for if-then reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3027–3035, Honolulu, HI, 2019b.

Maarten Sap, Hannah Rashkin, Derek Chen, Ronan Le Bras, and Yejin Choi. Social IQA: Commonsense Reasoning about Social Interactions. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4453–4463, Hong Kong, China, 2019c.

Maarten Sap, Vered Shwartz, Antoine Bosselut, Yejin Choi, and Dan Roth. Commonsense reasoning for natural language processing. In *Proceedings of the 58th Annual*

Meeting of the Association for Computational Linguistics: Tutorial Abstracts, pages 27–33, Online, 2020.

Edward Sapir. *Culture, language and personality: Selected essays*, volume 342. Univ of California Press, 1985.

Sotirios Sarantakos. *Social research*. Macmillan International Higher Education, 2012.

Edgar H Schein. What is culture. *Newbury Park, CA: Sage*, pages 243–253, 1991.

Karin Kipper Schuler. *VerbNet: A broad-coverage, comprehensive verb lexicon*. University of Pennsylvania, 2005.

Feng Shi, Liuqing Chen, Ji Han, and Peter Childs. A data-driven text mining and semantic network analysis for design information retrieval. *Journal of Mechanical Design*, 139 (11):111402, 2017.

Peng Shi and Jimmy Lin. Simple bert models for relation extraction and semantic role labeling. *ArXiv*, abs/1904.05255, 2019.

Brian K Smith. The unity of ritual. *Indo-Iranian Journal*, 29(2):79–96, 1986.

Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. Matching the blanks: Distributional similarity for relation learning. *arXiv preprint arXiv:1906.03158*, 2019.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642, 2013.

Radu Soricut and Daniel Marcu. Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 228–235, 2003.

Robert Speer and Catherine Havasi. Conceptnet 5. *Tiny Transactions of Computer Science*, 2012.

Robyn Speer, Joshua Chin, and Catherine Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge. *arXiv preprint arXiv:1612.03975*, 2016.

Speer, Robyn. ConceptNet in Context. Invited talk at the Reasoning for Complex QA (RCQA) Workshop at AAAI 2020, 2020.

Helen Spencer-Oatey and Peter Franklin. What is culture. *A compilation of quotations. GlobalPAD Core Concepts*, 1:22, 2012.

Gabriel Stanovsky, Julian Michael, Luke Zettlemoyer, and Ido Dagan. Supervised open information extraction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 885–895, 2018.

Alon Talmor, Jonathan Herzig, Nicholas Lourie, and Jonathan Berant. Commonsenseqa: A question answering challenge targeting commonsense knowledge. *CoRR*, abs/1811.00937, 2018. URL `http://arxiv.org/abs/1811.00937`.

Niket Tandon, Gerard De Melo, and Gerhard Weikum. Webchild 2.0: Fine-grained commonsense knowledge distillation. In *Proceedings of ACL 2017, System Demonstrations*, pages 115–120, Vancouver, Canada, 2017.

Niket Tandon, Bhavana Dalvi Mishra, Joel Grus, Wen-tau Yih, Antoine Bosselut, and Peter Clark. Reasoning about actions and state changes by injecting commonsense knowledge. *arXiv preprint arXiv:1808.10012*, 2018a.

Niket Tandon, Aparna S Varde, and Gerard de Melo. Commonsense knowledge in machine intelligence. *ACM SIGMOD Record*, 46(4):49–52, 2018b.

Paul H Thibodeau and Frank H Durgin. Metaphor aptness and conventionality: A processing fluency account. *Metaphor and Symbol*, 26(3):206–226, 2011.

Stith Thompson. *Motif-index of folk-literature: a classification of narrative elements in folktales, ballads, myths, fables, mediaeval romances, exempla, fabliaux, jest-books and local legends*, volume 4. Indiana University Press, 1960.

Stith Thompson. *The folktale*. Univ of California Press, 1977.

Erik F. Tjong Kim Sang and Fien De Meulder. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pages 142–147, 2003. URL `https://aclanthology.org/W03-0419`.

Trieu H Trinh and Quoc V Le. A simple method for commonsense reasoning. *arXiv preprint arXiv:1806.02847*, 2018.

Adam Trischler, Tong Wang, Xingdi Yuan, Justin Harris, Alessandro Sordoni, Philip Bachman, and Kaheer Suleman. Newsqa: A machine comprehension dataset. *arXiv preprint arXiv:1611.09830*, 2016.

Victor W Turner. Symbols in african ritual. *Science*, 179(4078):1100–1105, 1973.

Wendy L Ulrich. Hrm and culture: History, ritual, and myth. *Human resource management*, 23(2):117–128, 1984.

Anne Underhill. An analysis of mortuary ritual at the dawenkou site, shandong, china. *Journal of East Asian Archaeology*, 2(1):93–127, 2000.

Princeton University. About wordnet, 2010. Retrieved on May 9, 2016 from: http://wordnet.princeton.edu.

Hans-Jörg Uther. *The types of international folktales: a classification and bibliography, based on the system of Antti Aarne and Stith Thompson*. Suomalainen Tiedeakatemia, Academia Scientiarum Fennica, 2004.

Edwin R Van Teijlingen and Vanora Hundley. The importance of pilot studies. 2001.

Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R Bowman. Glue: A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv:1804.07461*, 2018.

Linlin Wang, Zhu Cao, Gerard De Melo, and Zhiyuan Liu. Relation classification via multi-level attention cnns. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1298–1307, 2016.

Ralph Weischedel, Martha Palmer, Mitchell Marcus, Eduard Hovy, Sameer Pradhan, Lance Ramshaw, Nianwen Xue, Ann Taylor, Jeff Kaufman, Michelle Franchini, et al. Ontonotes release 5.0 ldc2013t19. *Linguistic Data Consortium, Philadelphia, PA*, 23, 2013.

Johannes Welbl, Nelson F Liu, and Matt Gardner. Crowdsourcing multiple choice science questions. *arXiv preprint arXiv:1707.06209*, 2017.

Carla Willig. *EBOOK: introducing qualitative research in psychology*. McGraw-hill education (UK), 2013.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*, 2019.

Phillip Wolff and Dedre Gentner. Structure-mapping in metaphor comprehension. *Cognitive science*, 35(8):1456–1488, 2011.

Shanchan Wu and Yifan He. Enriching pre-trained language model with entity information for relation classification. In *Proceedings of the 28th ACM international conference on information and knowledge management*, pages 2361–2364, 2019.

Hu Xu, Bing Liu, Lei Shu, and Philip S. Yu. Bert post-training for review reading comprehension and aspect-based sentiment analysis. *arXiv preprint arXiv:1904.02232*, 2019.

Kutlay Yagmur, Kees de Bot, and Hubert Korzilius. Language attrition, language shift and ethnolinguistic vitality of turkish in australia. *Journal of multilingual and multicultural development*, 20(1):51–69, 1999.

W Victor H Yarlott, Armando Ochoa, Anurag Acharya, Laurel Bobrow, Diego Castro Estrada, Diana Gomez, Joan Zheng, David McDonald, Chris Miller, and Mark Alan Finlayson. Finding trolls under bridges: Preliminary work on a motif detector. In *Proceedings of the Ninth Annual Conference of Advancements in Cognitive Science (ACS) 2021*, 2021. URL `https://advancesincognitivesystems.github.io/acs2021/data/ACS-21_paper_23.pdf`.

Zhi-Xiu Ye and Zhen-Hua Ling. Multi-level matching and aggregation network for few-shot relation classification. *arXiv preprint arXiv:1906.06678*, 2019.

Jung-Soo Yi and Soonae Park. Cross-cultural differences in decision-making styles: A study of college students in five countries. *Social Behavior and Personality: an international journal*, 31(1):35–47, 2003.

Hamid Zafar, Mohnish Dubey, Jens Lehmann, and Elena Demidova. Iqa: Interactive query construction in semantic question answering systems. *Journal of Web Semantics*, 64:100586, 2020.

Sheng Zhang, Rachel Rudinger, Kevin Duh, and Benjamin Van Durme. Ordinal commonsense inference. *Transactions of the Association for Computational Linguistics*, 5:379–395, 2017.

Zhuosheng Zhang, Junjie Yang, and Hai Zhao. Retrospective reader for machine reading comprehension. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14506–14514, 2021.

Ofer Zwikael, Kazuo Shimizu, and Shlomo Globerson. Cultural differences in project management capabilities: A field study. *International Journal of Project Management*, 23(6):454–462, 2005.

APPENDIX A - SELF IDENTIFICATION QUESTIONNAIRE

This questionnaire, or a subset of it, was used in all the surveys as the self-identification questions.

PART I: ETHNICITY, RACE, AND CULTURAL GROUP QUESTIONS

1. How would you describe yourself?

    a. American Indian or Alaskan Native

    b. Native Hawaiian or Pacific Islander

    c. Black or African American

    d. Jewish

    e. White or Caucasian

    f. Asian

    g. Hispanic or Latino

    h. Irish

    i. Other (Please specify: _______)

2. Some people consider themselves to be part of a specific cultural group (e.g. Celtic, Taino, Amish, Cherokee). Do you identify with any specific cultural group? If so, which one?

    - No

    - Yes (Please specify _____)

3. If you answered "Yes" on Question #3, how strongly do you identify with said cultural group?

Barely Identify				Strongly Identify
o	o	o	o	o

4. Please circle the region of the world in which you were born.

    a. North America

    b. South America

    c. Europe

    d. Asia

    e. Africa

    f. Australia

    g. Other (Specify: ______)

5. What country do you currently live in? _____

6. What region of said country do you live in (territory, state, province, etc.)? _____

7. If you live in a country other than the country in which you were born, please specify the name of the country in which you were born: _____

8. If you live in a country other than the country in which you were born, please indicate how old you were when you first moved to your country of residence: _____

PART II: CAREGIVER QUESTIONS

Caregivers are defined as the primary people who raised you. In many cases, the caregivers are your father and mother, but could be grandparents, friends of the family, nannies, etc.

9. Please describe your first caregiver (e.g. father, mother, grandfather, etc.): ________
10. Please describe your second caregiver: ________
11. Please circle the region of the world in which caregiver #1 was born.
    a. North America
    b. South America
    c. Europe
    d. Asia
    e. Africa
    f. Australia
    g. Other (Specify: ______)
12. Is caregiver #1 from the same country you reside in?
    a. Yes
    b. No
    c. I don't know
13. If caregiver #1 is not from the same country you reside in, please specify their country of origin: _____
14. If caregiver #1 is not from the same country you reside in, please specify what region of their country of origin: _____
15. Please circle the region of the world in which caregiver #2 was born.
    a. North America
    b. South America

c. Europe

d. Asia

e. Africa

f. Australia

g. Other (Specify: ______)

16. Is caregiver #2 from the same country you reside in?

    a. Yes

    b. No

    c. I don't know

17. If caregiver #2 is not from the same country you reside in, please specify their country of origin: _____

18. If caregiver #2 is not from the same country you reside in, please specify what region of their country of origin: _____

19. How would you describe caregiver #1?

    a. American Indian or Alaskan Native

    b. Native Hawaiian or Pacific Islander

    c. Black or African American

    d. Jewish

    e. White or Caucasian

    f. Asian

    g. Hispanic or Latino

h. Irish

i. Other (Please specify: _______)

20. How would you describe caregiver #2?

a. American Indian or Alaskan Native

b. Native Hawaiian or Pacific Islander

c. Black or African American

d. Jewish

e. White or Caucasian

f. Asian

g. Hispanic or Latino

h. Irish

i. Other (Please specify: _______)

PART III: NATIVE LANGUAGE QUESTIONS

Please use the following list of language codes to answer questions 3-5. On the line provided in front of each question, please write in the code number which corresponds to the appropriate language.

a. English e. Hebrew i. Gaelic

b. Spanish f. Arabic j. Italian

c. Mandarin Chinese g. Russian k. Language not listed: _____

d. Portuguese h. French l. Unknown/Not Available

21. ______ Your own native language (mother tongue)

22. ______ Caregiver #1's native language

23. ______ Caregiver #2's native language

24. ______ Language used by your friends to you

25. ______ Language used by your family to you

26. ______ Language used by your neighbors to you

27. ______ Language used by you to your friends

28. ______ Language used by you to your family

29. ______ Language used by you to your neighbors

PART IV: RELIGION QUESTIONS

Please use the following list of religion codes to answer questions 6-11. On the line provided in front of each question, please write in the code number which corresponds to the appropriate religion.

A. Christian D. Muslim

A1. Roman Catholic D1. Shia Muslim

A2. Protestant Christian D2. Sunni Muslim

B. Jewish E. Buddhist

B1. Sephardi Jewish F. Non-Religious

B2. Ashkenazi Jewish G. Religion not listed: ________

C. Hindu H. Unknown/Not Available

30. _____ What religion did you practice growing up?

31. _____ What religion do you practice now?

32. _____ What religion has Caregiver #1 practiced?

33. _____ What religion has Caregiver #2 practiced?

34. _____ What religion do your friends practice?

APPENDIX B - INSTRUMENTS OF THE MOTIF UNDERSTANDING STUDY

Main Study Questionnaire

Irish Motif Protocol

Salmon of Wisdom

No Motif

While getting coffee, James says to his coworker: "I've been trying to figure out a way to reduce the cost of our product." His coworker turns to him and says, "Management would love that."

Motif alone

While getting coffee, James says to his coworker: "I've been trying to figure out a way to reduce the cost of our product." His coworker turns to him and says, "You're hunting for the salmon of wisdom."

Implication: James is on a fool's errand.

Combined

While getting coffee, James says to his coworker: "I've been trying to figure out a way to reduce the cost of our product." His coworker turns to him and says, "You're hunting for the salmon of wisdom. Management would love that."

Attribute Questions

Based on the conversation above, how much would you think each of the following terms could be used to describe James (not his coworker):

1. Greedy (D)
2. Foolish (D)
3. Popular (N)
4. Weird (N)

5. Smart (RH: salmon of wisdom grants wisdom, but is a foolish thing to seek)

6. What other attributes do you think James might have, based on the conversation?

7. What attributes might he lack, based on the conversation?

Inferences

Based on the conversation above, how would you answer the following questions:

1. Is James talking to someone? (L1)

2. Is James currently at work? (L1)

3. Does the coworker mention management as a result of James' plan? (L2)

4. Do you think James' plan will work out? (L3)

5. Do you think James is likely to be successful in this job? (L3)

Finn McCool

No Motif

A group of friends are talking about something that happened at work: their coworker, William, took control of a major project. One of them says: "How do you think the project is going to work out?" Another replies: "Yeah, William jumps around between a lot of projects."

Motif

A group of friends are talking about something that happened at work: their coworker, William, took control of a major project. One of them says: "He's like Finn McCool himself."

Implication: William will have great success on the project due to his incredible talent.

Combined

A group of friends are talking about something that happened at work: their coworker, William, took control of a major project. One of them says: "How do you think the project is going to work out?" Another replies: "William has taken over a lot of projects. He's like Finn McCool himself."

Attribute Questions

Based on the conversation above, how much would you think each of the following terms could be used to describe William (not his coworkers):

1. Talented (D)
2. Smart (D)
3. Lazy (N)
4. Greedy (N)
5. Popular (RH: the reading of McCool with focus on "cool")
6. What other attributes do you think William might have, based on the conversation?
7. What attributes might he lack, based on the conversation?

Inferences

Based on the conversation above, how would you answer the following questions:

1. Does William work at the same company as the friends? (L1)
2. Is William part of the conversation? (L1)
3. Do the friends bring up William because of a project at work? (L2)
4. Do you think the project will succeed? (L3)
5. Do you think William is respected by his team? (L3)

Tir na nog

No Motif

Sally and her friends are looking for a new bar after their favorite bar has closed. Sally says, “I hope we can find somewhere as good.” Her friend says, “That place was so convenient!”

Motif

Sally and her friends are looking for a new bar after their favorite bar has closed. Sally says, “I hope we can find somewhere as good.” Her friend says, “We might be looking for Tir na nog.”

Implication: Sally and her friends are looking for a mythical place that they will never find.

Combined

Sally and her friends are looking for a new bar after their favorite bar has closed. Sally says, “I hope we can find somewhere as good.” Her friend says, “That place was so convenient! We might be looking for Tir na nog.”

Attribute Questions

Based on the conversation above, how much would you think each of the following terms could be used to describe Sally (not her friends):

1. Greedy (D)
2. Smart (D)
3. Popular (N)
4. Loving (N)
5. Lucky (RH: misinterpretation of tir na nog as a bar name, lucky that she just found one)
6. What other attributes do you think Sally might have, based on the conversation?

7. What attributes might she lack, based on the conversation?

Inferences

Based on the conversation above, how would you answer the following questions:

1. Is the bar still in business? (L1)
2. Are Sally and her friends currently at the bar? (L1)
3. Are Sally and her friends looking for a new bar? (L2)
4. Does Sally's friend have an idea for a new bar? (L3)
5. Do you think that Sally will be able to find a new bar? (L3)

Shamrock

No Motif

Jerry is talking about a presentation someone selling software gave at work. He recalls a particularly notable graphic and thinks: "That was a very modern graphic."

Motif

Jerry is talking about a presentation someone selling software gave at work. He recalls a particularly notable graphic and thinks: "It sure was St. Patrick's shamrock."

Implication: The graphic really demonstrated what the software did, the way St. Patrick used the shamrock to demonstrate the trinity.

Combined

Jerry is talking about a presentation someone selling software gave at work. He recalls a particularly notable graphic and thinks: "That was a very modern graphic. It sure was St. Patrick's shamrock."

Attribute Questions

Based on the conversation above, how much would you think each of the following terms could be used to describe the software salesman:

1. Smart (D)
2. Easy-to-understand (D)
3. Friendly (N)
4. Hard-working (N)
5. Lucky (RH: lucky is the more common meaning, but references the St. Patrick story)
6. What other attributes do you think the software salesman might have, based on the conversation?
7. What attributes might he lack, based on the conversation?

<u>Inferences</u>

Based on the conversation above, how would you answer the following questions:

1. Did Jerry attend the presentation? (L1)
2. Does the software salesman work at the same company as Jerry? (L1)
3. Do you think that Jerry understood what the software does? (L2)
4. Do you think that Jerry was convinced the software was useful? (L3)
5. Do you think the software salesman made a successful pitch? (L3)

300 years as a swan

No Motif

Jessica is online debating the utility of vaccination, when another commenter takes objection with what she's saying. Someone tells her: "You're just trying to make people mad."

Motif

Jessica is online debating the utility of vaccination, when another commenter takes objection with what she's saying. Someone tells her: "That sort of thinking will have your kids spend 300 years as swans."

Implication: Jessica's stance will lead to a tragic fate for her kids, like the children of Lir.

Combined

Jessica is online debating the utility of vaccination, when another commenter takes objection with what she's saying. Someone tells her: "You're just trying to make people mad. That sort of thinking will have your kids spend 300 years as swans."

Attribute Questions

Based on the conversation above, how much would you think each of the following terms could be used to describe Jessica:

1. Educated (D)
2. Naïve (D)
3. Friendly (N)
4. Antagonistic (N)
5. Protective (RH: interpretation based on length of time + kids + animal)
6. What other attributes do you think Jessica might have, based on the conversation?
7. What attributes might she lack, based on the conversation?

Inferences

Based on the conversation above, how would you answer the following questions:

1. Is Jessica talking to people in person? (L1)

2. Is Jessica in an argument? (L1)

3. Does the other commenter think Jessica is trying to make people mad? (L2)

4. Is Jessica in favor of vaccination? (L3)

5. Does the other commenter believe Jessica's children are healthy? (L3)

Wren

No Motif

Charlie is out with his wife, Courtney, for dinner at a very nice restaurant. Courtney is shocked he was able to get a reservation, but remembers he made a call earlier.

Motif

Charlie is out with his wife, Courtney, for dinner at a very nice restaurant. Courtney is shocked he was able to get a reservation, but remembers he is one of the wren boys.

Implication: Charlie is well-connected to some form of secret society.

Combined

Charlie is out with his wife, Courtney, for dinner at a very nice restaurant. Courtney is shocked he was able to get a reservation, but remembers he is one of the wren boys and made a call earlier.

Attribute Questions

Based on the conversation above, how much would you think each of the following terms could be used to describe Charlie:

1. Sociable (D)

2. Resourceful (D)

3. Generous (N)

4. Jealous (N)

5. Weird (RH: talking about birds without prior context)
6. What other attributes do you think Charlie might have, based on the conversation?
7. What attributes might he lack, based on the conversation?

Inferences

Based on the conversation above, how would you answer the following questions:

1. Is Charlie in a restaurant? (L1)
2. Are Charlie and Courtney in a relationship? (L1)
3. Does Courtney know how Charlie got a reservation? (L2)
4. Was there anything unusual about Charlie making the reservation? (L3)
5. Is Charlie a well-connected individual? (L3)

Leprechaun

No Motif

Fred's acquaintances are discussing the number of things he's brought home from his workplace. One of them says, "It's amazing how much stuff he has here."

Motif

Fred's acquaintances are discussing the number of things he's brought home from his workplace. One of them says, "He's like a leprechaun."

Implication: Fred is a mischievous person who questionably acquires things.

Combined

Fred's acquaintances are discussing the number of things he's brought home from his workplace. One of them says, "It's amazing how much stuff he has here. He's like a leprechaun."

Attribute Questions

Based on the conversation above, how much would you think each of the following terms could be used to describe Fred:

1. Tricky (D)
2. Mischievous (D)
3. Skillful (N)
4. Strong (N)
5. Lucky (RH: leprechaun -> lucky charms)
6. What other attributes do you think Fred might have, based on the conversation?
7. What attributes might he lack, based on the conversation?

<u>Inferences</u>

Based on the conversation above, how would you answer the following questions:

1. Is Fred present for the conversation? (L1)
2. Did Fred get his stuff from his workplace? (L1)
3. Are Fred's acquaintances commenting because Fred took items from work? (L2)
4. Is Fred a well-liked person? (L3)
5. Was Fred authorized to take the items he has from his workplace? (L3)

King Conchobar

No Motif

After Tim takes over as CEO of a faltering company, the performance of the company improves. Some employees of the company are around a water cooler when one says, "Tim sure has taken a firm hold of things."

Motif

After Tim takes over as CEO of a faltering company, the performance of the company improves. Some employees of the company are around a water cooler when one says, "Tim sure is acting like King Conchobar."

Implication: Tim is a strong leader.

Combined

After Tim takes over as CEO of a faltering company, the performance of the company improves. Some employees of the company are around a water cooler when one says, "Tim sure has taken a firm hold of things. He's acting like King Conchobar."

<u>Attribute Questions</u>

Based on the conversation above, how much would you think each of the following terms could be used to describe Tim:

1. Smart (D)
2. Talented (D)
3. Short-tempered (N)
4. Strange (N)
5. Tyrannical (RH: king)
6. What other attributes do you think Tim might have, based on the conversation?
7. What attributes might he lack, based on the conversation?

<u>Inferences</u>

Based on the conversation above, how would you answer the following questions:

1. Is Tim present in the conversation? (L1)
2. Has Tim always been CEO? (L1)
3. Are the employees talking about Tim because of the recent takeover? (L2)

4. Is Tim a good leader? (L3)
5. Do the employees like Tim? (L3)

Jewish Motif Protocols

1. **Amalek**

No Motif:

A group of neighbors are out one night. Conversation turns to their recently elected Mayor: Christopher. One of them says: “That new Mayor, Christopher, what do you think of him?” Another says, “His policies are very different from the last mayor’s policies.”

Motif alone:

A group of neighbors are out one night. Conversation turns to their recently elected Mayor: Christopher. One of them says: “That new Mayor, Christopher, what do you think of him?” Another says “I looked at his policies. I think he’s the Amalek.”

Combined:

A group of neighbors are out one night. Conversation turns to their recently elected Mayor: Christopher. One of them says: “That new Mayor, Christopher, what do you think of him?” Another says “I looked at his policies and they alarmed me. I think he’s the Amalek.”

<u>Attribute Questions:</u>

Based on the conversation above, how much would you think each of the following terms could be used to describe Christopher:

1. Smart (N)
2. Misogynistic (N)
3. Honest (D)

4. Anti-Semitic (D)

5. Evil (RH: Amalek is evil, but more so towards Jews)

6. What other attributes do you think Christopher might have, based on the conversation?

7. What attributes might he lack, based on the conversation?

<u>Inferences:</u>

There are two level 1 questions (perception), one level 2 question (comprehension), and two level 3 questions (projection). L1/L2 is expected to be similar for all groups with differences mostly on L3.

Based on the conversation above, how would you answer the following questions:

8. Do the speakers live in the same town? (L1)

9. Is Christopher present for the conversation? (L1)

10. Do you think Christopher's policies favorable to the neighbors? (L2)

11. Do you think Christopher will be a good Mayor? (L3)

12. Do you think the other people in the conversation are likely to vote for Christopher in the future? (L3)

1. **Tower of Babel**

No Motif:

One evening over drinks, Maria says "My brother wanted to take his current business international. It's currently based in the United States."

Motif alone:

One evening over drinks, Maria says "My brother wanted to take his current business international. I think it's the Tower of Babel."

Combined:

One evening over drinks, Maria says "My brother wanted to take his current business international. I think his finances say otherwise. It's the Tower of Babel."

<u>Attribute Questions:</u>

Based on the conversation above, how much would you think each of the following terms could be used to describe Maria's brother (not Maria):

1. Greedy (D)
2. Friendly (N)
3. Overambitious (D)
4. Smart (RH)
5. Loving (N)
6. What other attributes do you think Maria's brother might have, based on the conversation?
7. What attributes might he lack, based on the conversation?

<u>Inferences:</u>

Based on the conversation above, how would you answer the following questions:

8. Does the business belong to Maria? (L1)
9. Is Maria an only child? (L1)
10. Do you think the business has enough money to go international? (L2)
11. Do you think that the business will succeed if it goes international? (L3)
12. Does Maria think an additional loan would help her brother? (L3)

2. **Milk with Meat**

No Motif:

Juan, a baseball correspondent, just started his graduate work in Physics. He has been juggling both works.

Motif alone:

Juan, a baseball correspondent, just started his graduate work in Physics. His friends say he always mixes milk with meat.

Combined:

Juan, a baseball correspondent, just started his graduate work in Physics. His friends say he always talks about his research at the games, mixing milk with meat.

Attribute Questions:

Based on the story above, how much would you think each of the following terms could be used to describe Juan:

1. Smart (R)
2. Friendly (N)
3. Context-aware (D)
4. Mindful (D)
5. Loving (N)
6. What other attributes do you think might describe Juan, based on the story?
7. What attributes might definitely not be characteristics of Juan, based on the story?

Inferences:

Based on the story above, how would you answer the following questions:

8. Does Juan go to baseball games regularly?
9. Has Juan joined graduate school?

10. Do you think the friends are interested in Physics?

11. Do you think Juan needs to stop talking about his research in the games in the future?

12. Do you think Juan will have problems with friends if he keeps talking about Physics during the games?

3. **Seventy Languages**

No Motif:

One evening over drinks, Maria says "My brother's business was doing well, so he expanded it recently. Now it has a whole host of departments."

Motif alone:

One evening over drinks, Maria says "My brother's business was doing well, so he expanded it recently. It's like he's speaking seventy languages."

Combined:

One evening over drinks, Maria says "My brother's business was doing well, so he expanded it recently. Now it has a whole host of departments. It's like he's speaking seventy languages."

<u>Attribute Questions:</u>

Based on the conversation above, how much would you think each of the following terms could be used to describe Maria's brother (not Maria):

1. Successful (D)

2. Lucky ~~Friendly~~ (N)

3. Overambitious (D)
4. Smart (RH)
5. Loving (N)
6. What other attributes do you think Maria's brother might have, based on the conversation?
7. What attributes might he lack, based on the conversation?

Inferences:

Based on the conversation above, how would you answer the following questions:

8. Does the business belong to Maria? (L1)
9. Is Maria an only child? (L1)
10. Do you think Maria's brother spends more time on his business after the expansion? (L2)
11. Do you think that the business is running smoothly after the expansion? (L3)
12. Does Maria think her brother will be successful in managing the business after the expansion? (L3)

4. **Haman**

No Motif:

Two friends are talking over lunch. One says, "I was listening to a speech the other day by the new CEO of the pharmaceutical company. She is making a lot of policy changes."

Motif alone:

Two friends are talking over lunch. One says, "I was listening to a speech the other day by the new CEO of the pharmaceutical company. I'm convinced she's Haman."

Combined:

Two friends are talking over lunch. One says, “I was listening to a speech the other day by the new CEO of the pharmaceutical company. She is making a lot of policy changes. I’m convinced she’s Haman.”

<u>Attribute Questions:</u>

Based on the conversation above, how much would you think each of the following terms could be used to describe the CEO:

1. Charismatic (N)
2. Angry (N)
3. Conniving (D)
4. Progressive (R)
5. Evil (D)
6. What other attributes do you think CEO might have, based on the conversation?
7. What attributes might they lack, based on the conversation?

<u>Inferences:</u>

Based on the conversation above, how would you answer the following questions:

8. Has the CEO been at his position for long? (L1)
9. Do you think the friends know the CEO personally? (L1)
10. Do you think the friend approves of the way CEO’s new policy? (L2)
11. Do you think the new policies will be detrimental to the public? (L3)
12. Do you think that the new CEO will be likeable? (L3)

5. **Dove**

No Motif:

Two colleagues are talking over dinner. Alex says, “I saw the new government’s proposal. The new Prime Minister is a real leader.”

Motif alone:

Two colleagues are talking over dinner. Alex says, “I saw the new government’s proposal. The new Prime Minister is a Dove.”

Combined:

Two colleagues are talking over dinner. Alex says, “I saw the new government’s proposal. The new Prime Minister is a real leader and a Dove.”

Attribute Questions:

Based on the conversation above, how much would you think each of the following terms could be used to describe the Prime Minister:

1. Careful (N)
2. Inconsiderate (D)
3. Overambitious (RH)
4. Peaceful (D)
5. Loving (N)
6. What other attributes do you think the Prime Minister might have, based on the conversation?
7. What attributes might she lack, based on the conversation?

Inferences:

Based on the conversation above, how would you answer the following questions:

8. Do you think they’re talking about an old proposal? (L1)

9. Is this conversation happening over a meal? (L1)
10. Do you think Alex approves of the new PM? (L2)
11. Do you think that the PM will have peaceful policies in the future? (L3)
12. Do you think the PM will remain loyal to his voters? (L3)

6. **Kiddush**

No Motif:

A group of friends decides to throw a summer party for the neighborhood. They agree it should be a big event, with no expense spared. One of them suggests that some homeless people who are known to hang around the neighborhood should be invited.

Motif alone:

A group of friends decides to throw a summer party for the neighborhood. They agree it should be a kiddush hashem.

Combined:

A group of friends decides to throw a summer party for the neighborhood. They agree it should be a big event, with no expense spared — a kiddush hashem. One of them suggests that some homeless people who are known to hang around the neighborhood should be invited.

Attribute Questions:

Based on the story above, how much would you think each of the following terms could be used to describe the neighborhood party:

1. Charitable (D)
2. Devout (RH)
3. Musical (N)
4. Energetic (N)

5. Crowded (D)
6. What other attributes do you think might describe the party, based on the story?
7. What attributes might definitely not be characteristics of the party, based on the story?

Inferences:

Based on the story above, how would you answer the following questions:

8. Is the party planned for the winter? (L1)
9. Will there be a lot of people invited to the party? (L1)
10. Do you think the party will be held at the neighborhood bar? (L2)
11. Will they agree to invite the homeless people? (L3)
12. Do you think there will be a prayer offered at the party? (L3)

7. **Behemoth**

No Motif:

Two engineers are talking over lunch. One of them says, “Google just bought three new promising startups.”

Motif alone:

Two engineers are talking over lunch. One of them says, “Google just acquired three new promising startups. It will be a behemoth soon.”

Combined:

Two engineers are talking over lunch. One of them says, “Google just acquired three new promising startups. It’s becoming even bigger. It will be a behemoth soon.”

Attribute Questions:

Based on the story above, how much would you think each of the following terms could be used to describe Google:

1. Undefeatable ~~Powerful~~ (D)
2. Conscientious (D)
3. Prosperous (N)
4. Successful (RH)
5. Charitable (N)
6. What other attributes do you think might describe Google, based on the story?
7. What attributes might definitely not be characteristics of Google, based on the story?

<u>Inferences:</u>

Based on the story above, how would you answer the following questions:

8. Are the people in the conversation medical doctors? (L1)
9. Were the companies Google acquired big companies? (L1)
10. Do you think the startups will add to Google's influence? (L2)
11. Do you think Google will implement policies that will benefit the public? (L3)
12. Do you think Google is likely to have an unnecessarily large influence on people? (L3)

Puerto Rican Motif Protocol

CHARACTER MOTIF: Jibaro

Stimulus story

No Motif:

To help decide what laptop to pick at the store, Felipe fills out a quiz that suggests a few models based on the user's needs and background. When the store manager finds out

Felipe is a farmer looking for a laptop, the manager tells one of the store assistants "That man is a farmer; it's amazing that he drove all the way here to the city!"

Motif alone:

To help decide what laptop to pick at the store, Felipe fills out a quiz that suggests a few models based on the user's needs and background. When the store manager finds out Felipe is a farmer looking for a laptop, the manager tells one of the store assistants "In that case, let's make sure we give him a special jibaro deal on a laptop!"

Combined:

To help decide what laptop to pick at the store, Felipe fills out a quiz that suggests a few models based on the user's needs and background. When the store manager finds out Felipe is a farmer looking for a laptop, the manager tells one of the store assistants "That man is a farmer; it's amazing that he drove all the way here to the city! In that case, let's make sure we give him a special jibaro deal on a laptop!"

[Implication: Jibaro means dumb countryside person; Felipe is going to be taken advantage of.]

Attribute Questions

Based on the passage above, how much would you think each of the following terms could be used to describe ***how the manager perceives Felipe***:

1. Intelligent (D)
2. Naïve (D)
3. High-regard (RH, may assume manager holds felipe in high regard for driving to the city)
4. Enthusiastic (N)
5. Hard-working (N)

Inference Questions

Based on the passage above, how would you answer the following questions:

6. Is Felipe a farmer? (L1)
7. Is Felipe looking for a laptop? (L1)
8. Was the manager reacting to Felipe being a farmer? (L2)
9. Does the manager think highly of Felipe? (L3)
10. Is Felipe getting a good deal? (L3)

CHARACTER MOTIF: El guaraguao y el pitirre

Stimulus story

No Motif:

Juan is discussing with his grandfather how some of the younger generation of Puerto Ricans want independence from the US. His grandfather responds, “I have noticed that.”

Motif alone:

Juan is discussing with his grandfather how some of the younger generation of Puerto Ricans want independence from the US. His grandfather responds, “The younger generation must realize that every hawk has its kingbird.”

Combined:

Juan is discussing with his grandfather how some of the younger generation of Puerto Ricans want independence from the US. His grandfather responds, “I have noticed that. The younger generation must realize that every hawk has its kingbird.”

[Implication: His grandfather is for the independence, implying that although Puerto Rico is a small poor territory they can still win]

Attribute Questions

Based on the passage above, how much would you think each of the following terms could be used to describe ***Juan’s grandfather***:

11. Compliant (D)

12. Submissive (D)

13. Anti-Independence (RH, may assume he is against independence)

14. Attentive (N)

15. Focused (N)

Inference Questions

Based on the passage above, how would you answer the following questions:

16. Was Juan speaking to his grandfather? (L1)

17. Did Juan mention independence? (L1)

18. Was Juan's grandfather responding to what Juan said? (L2)

19. Is Juan's grandfather disagreeing with the pro-independence view? (L3)

20. Is Juan's grandfather impressed by the younger generation? (L3)

PROP MOTIF: The sentry box of the devil

Stimulus story

No Motif:

Sebastian is running late to work for the third time this week because he can't find his car keys. When asking his wife if she has seen his car keys, she responds "You lost track of your keys again?!"

Motif alone:

Sebastian is running late to work for the third time this week because he can't find his car keys. When asking his wife if she has seen them, she responds "They are in the devil's sentry box, as always."

Combined:

Sebastian is running late to work for the third time this week because he can't find his car keys. When asking his wife if she has seen them, she responds "You lost track of your keys again?! They are in the devil's sentry box, as always."

[Implication: Sebastian will never find his keys again.]

Attribute Questions

Based on the passage above, how much would you think each of the following terms could be used to describe ***Sebastian's wife's response***:

21. Cooperative (D)
22. Informative (D)
23. Useful (RH, may assume his wife was being serious)
24. Interesting (N)
25. Charming (N)

Inference Questions

Based on the passage above, how would you answer the following questions:

26. Does Sebastian have a wife? (L1)
27. Is Sebastian able to find his car keys? (L1)
28. Was Sebastian's wife talking to Sebastian? (L2)
29. Did Sebastian's wife give him useful information to find his car keys? (L3)
30. Will Sebastian find his car keys successfully using what his wife said? (L3)

PROP MOTIF: La pava

Stimulus story

No Motif:

Amanda is trying to teach her younger sister how to open a coconut. When her younger sister attempts to do it herself, Amanda laughs and says “Oh gosh!”

Motif alone:

Amanda is trying to teach her younger sister how to open a coconut. When her younger sister attempts to do it herself, Amanda laughs and says “You should be wearing a pava with the way you’re handling this.”

Combined:

Amanda is trying to teach her younger sister how to open a coconut with a knife. When her younger sister attempts to do it herself, Amanda laughs and says “Oh gosh! You should be wearing a pava with the way you’re handling this.”

[Implication: Amanda’s sister is handling the coconut well, like a true countryside person which is why she would wear the hat.]

Attribute Questions

Based on the passage above, how much would you think each of the following terms could be used to describe ***how Amanda’s sister is handling the coconut***:

31. Skillfully (D)
32. Incorrectly (D)
33. Fooling around (RH, may assume she is being silly with her sister)
34. Slowly (N)
35. With a knife (N)

Inference Questions

Based on the passage above, how would you answer the following questions:

36. Does Amanda have a sister? (L1)
37. Is Amanda opening a coconut? (L1)

38. Does Amanda's sister make her laugh? (L2)

39. Is Amanda's sister opening the coconut correctly? (L3)

40. Is Amanda impressed by her sister? (L3)

CHARACTER MOTIF: JUAN BOBO

Stimulus story

No Motif:

The principal of a school walks into a classroom to pick out students for a news interview. The principal wants the best, good-looking students to represent the school well. When the principal goes to pick out Jimmy for the interview, another student proclaims "No! Don't pick Jimmy!"

Motif alone:

The principal of a school walks into a classroom to pick out students for a news interview. The principal wants best, good-looking students to represent the school well. When the principal goes to pick out Jimmy for the interview, another student proclaims "No! Jimmy is the Juan Bobo of the classroom."

Combined:

The principal of a school walks into a classroom to pick out students for a news interview. The principal wants the best, good-looking students to represent the school well. When the principal goes to pick out Jimmy for the interview, another student proclaims "No! Don't pick Jimmy, he is the Juan Bobo of the classroom."

[Implication: Juan Bobo is the typical class clown character.]

Attribute Questions

Based on the passage above, how much would you think each of the following terms could be used to describe ***Jimmy***:

41. Responsible (D)

42. Intelligent (D)

43. Good Looking (RH, talking about finding good-looking students in the story)

44. Funny (N)

45. Hyperactive (N)

Inference Questions

Based on the passage above, how would you answer the following questions:

46. Do you think the principal is searching for a good-looking student? (L1)

47. Do you think the principal cares much about the types of students for an interview? (L1)

48. Do you think that Jimmy is a good candidate for the interview? (L2)

49. Are other students just jealous of Jimmy? (L3)

50. Is Jimmy a class clown type character? (L3)

EVENT MOTIF: MIRACLE OF HORMIGUEROS

Stimulus story

No Motif:

Gabriel has been searching for his phone for 10 days, which contains all his precious photos and memories from the past 5 years. After asking his sister for help, she stops looking after an hour and says, ”Your phone stored so many precious memories.”

Motif alone:

Gabriel has been searching for his phone for 10 days, which contains all his precious photos and memories from the past 5 years. After asking his sister for help, she stops looking after an hour and says, ”It's as if the lady of Hormigueros has visited us.”

Combined:

Gabriel has been searching for his phone for 10 days, which contains all his precious photos and memories from the past 5 years. After asking his sister for help, she stops looking after an hour and says, ”Your phone stored so many precious memories... it's as if the lady of Hormigueros has visited us.”

[Implication: There are two stories of miracles from Hormigueros; one of a bull being prevented from attacking a farmer, one of a lost daughter being found. In both, the miracle is believed to be due to intercession by the Virgin of Montserrat.]

Attribute Questions

Based on the passage above, how much would you think each of the following terms could be used to describe ***Gabriel finding his phone again***:

1. Unlikely (D)
2. Expected (D)
3. Impossible (RH since they may interpret it as though a lady has made it impossible to find or put a curse on them)
4. Miraculous (N)
5. Insignificant (N)

Inference Questions

Based on the passage above, how would you answer the following questions:

6. Is Gabriel searching for his phone? (L1)
7. Has Gabriel been searching for his phone for a long time? (L1)
8. Is Gabriel's sister helping him as a result of him losing his phone? (L2)
9. Will Gabriel find his phone? (L3)
10. Is the Lady of Hormigueros preventing Gabriel's phone being found? (L3)

Experiment Script for the Pilot

Assumed setting: Office or Lab

***Subject** joins the Zoom call. Record the time that they join.*

Researcher: Hi, thanks for taking the time to take this survey. I've posted the link to the survey in the Zoom chat. Let me know once you've opened it.

*Wait for the **subject** to open the survey.*

Researcher: Thank you for being willing to participate on our study on story understanding. This study will take about an hour and mainly involves you reading some simple stories and answering some questions about them.

Researcher: There is no time limit as we are not recording how long the questionnaire takes, so there's no need to rush.

Researcher: Once you leave, please do not discuss the contents of the questionnaire with other potential participants.

Researcher: Please read the letter on the first page of the survey, which explains the details of the study you are about to participate in, and let me know when you have finished reading it. If you would like a copy of the letter, please contact `motifsurvey@cs.fiu.edu` and we will send it to you.

*Wait for the **subject** to finish reading the letter.*

Researcher: Do you have any questions regarding this study?

*Wait for the **subject** to ask questions.*

(If asked about the focus of the study) Researcher: I'm sorry, but I can't tell you anything more about the focus of the study at this time other than we are interested in the kinds of information and assumptions people make when given stories of different types.

Researcher: As stated, your participation in this research is voluntary and there is no penalty for refusing to participate or deciding to stop. Do you agree to participate in this study?

*Wait for the **subject** to confirm intent to participate.*

Researcher: During the questionnaire, I will have my audio/video off and you're free to do the same. If I have to leave at any point in time, I'll notify you via Zoom chat.

Researcher: While taking the questionnaire, please refrain from looking up any of the information in it. Please fill out this questionnaire to the best of your abilities and let me know when you have finished the questionnaire.

*Leave the **subject** alone with the questionnaire.*

*During the questionnaire, **subjects** may ask questions. Answering questions that clarify wording on questions is fine. Do not answer questions regarding the focus of the study beyond the previous suggested answer.*

(If asked why we are collecting demographic information) Researcher: Part of what we are interested in is how people with different backgrounds respond to the stories presented. If you do not feel comfortable providing this information, as stated earlier, your participation is voluntary.

*Once the **subject** indicates they are finished:*

Researcher: Thanks for participating. Would you mind if I ask you a few feedback questions about the survey?

*If **subject does mind:***

Researcher: Alright, that's fine. Again, thanks for your time. You will receive a separate email containing your payment for participating in the study.

*If **subject does not mind:** ask the **subject** the questions for the exit interview.*

Researcher: Thanks for taking the time to participate in this study.

Record the time the subject left, as well as whether or not they completed the survey.

APPENDIX C - INSTRUMENTS OF THE MOTIF RECOGNIZABILITY STUDY

This appendix includes the instruments of the motif recognizability study, in the following order:

1. Template of Questionnaire
2. Motific Prompts
    (a) Irish Prompts
    (b) Jewish Prompts
    (c) Puerto Rican Prompts
3. Baseline Prompts

See the following pages for the above-mentioned documents.

Recognizability of motifs for out-culture people

Section I – Self Identification

…

Section II (A)

Read the following text and answer the following questions:

We are in bad psychic shape. The behemoth is in his bathrobe. We need to suit up, get a grip. Because if there are more 9/11s coming at us, or retaliation here over Iraq.

1. What is your understanding of the overall text?
   <Open Text>

2. How much confidence do you have in your understanding of the text?

Extremely Confident	Very Confident	Somewhat confident	Neither confident nor unconfident	Somewhat unconfident	Very unconfident	Extremely unconfident

3. How much of the meaning of the overall text do you think you understood?

Entirely understood	Mostly understood	Somewhat understood	Understood about half	Somewhat didn't understand	Mostly didn't understand	Did not understand at all

4. Was there any portion of the text you did not understand?
   <Yes/No>

5. Highlight the portion you understood the least.

6. Do you think you are missing some information to understand this clearly?
   <Yes/No>

7. How much confidence do you have that you are missing some information to understand this sentence clearly?

Extremely Confident	Very Confident	Somewhat confident	Neither confident nor unconfident	Somewhat unconfident	Very unconfident	Extremely unconfident

8. What information do you think you would need to understand this better / more clearly?
   <Open Text>

9. How much confidence do you have that you are need the information you provided in #10 to understand this sentence better/clearly?

Extremely Confident	Very Confident	Somewhat confident	Neither confident nor unconfident	Somewhat unconfident	Very unconfident	Extremely unconfident

Section II (B)

10. Had you ever heard of the term “Behemoth” before?
<Open Text>

11. How familiar are you with the term "Behemoth"?

Extremely Confident	Very Confident	Somewhat confident	Neither confident nor unconfident	Somewhat unconfident	Very unconfident	Extremely unconfident

12. What do you think is the meaning of “Behemoth” in the given text?
<Open Text>

13. How much confidence do you have in your understanding of the meaning of "Behemoth"?

Extremely Confident	Very Confident	Somewhat confident	Neither confident nor unconfident	Somewhat unconfident	Very unconfident	Extremely unconfident

14. Based on the provided text, please rate the following attributes in terms of how likely they are to fit to describe the entity referred to as “behemoth”:

	Completely fits	Somewhat fits	Nether fits or doesn’t fit	Somewhat doesn’t fit	Completely doesn’t fit
Important					
Powerful					
Good					
Bad					
Evil					
Unprepared					
Alert					

Updated attributes for Jewish motifs

Amalek #1

Teitelbaum addresses a large rally on Saturday evening in Mea She'arim, organized by the Hassidic umbrella organization Edah Haredit, and told the crowd that the State of Israel is this generation's Amalek, and the Zionists are the offspring of Amalek.

	Attribute Category
Evil	Correct
Anti-Semitic	Correct Nuance
Honest	Opposite
Nice	Opposite Nuance
Powerful	Situational
Misogynistic	Neutral
Smart	Neutral

Amalek #2

A fly is powerless to bore a hole in a piece of flesh. Only once the fly finds a lesion or an open cut does it have the ability to further the damage. Thus- says Rabbi Luntschitz, Amalek has no power against a righteous person.

	Attribute Category
Evil	Correct
Anti-Semitic	Correct Nuance
Honest	Opposite
Nice	Opposite Nuance
Weak	Situational
Misogynistic	Neutral
Smart	Neutral

Amalek #3

Nonetheless, there were many important rabbis such as Maimonides in his Sefer Hamitzvot and Mishneh Torah and Rabbi Pinhas Halevi of Barcelona in his Sefer Hahinuch(13th century) who rules that Amalek still exists and we are still commanded to remember their deed and to destroy

	Attribute Category
Evil	Correct
Anti-Semitic	Correct Nuance
Honest	Opposite
Nice	Opposite Nuance
Sinful	Situational
Misogynistic	Neutral
Smart	Neutral

Amalek #4

On this basis, my revered rebbe Rav J.B. Soloveitchik quoted his grandfather, Rav Haim of Brisk, who distinguished between the physical nation of Amalek that once lived near Canaan and the concept of evil identified with Amalek. This latter Amalek exists in every generation - and must be continuously destroyed.

	Attribute Category
Evil	Correct
Anti-Semitic	Correct Nuance
Honest	Opposite
Nice	Opposite Nuance
Greedy	Situational
Misogynistic	Neutral
Smart	Neutral

Amalek #5

Whether the Nazis were Amalek, and I believe they were in the last generation, or Iran is Amalek today, and I believe they are, is not the point. The Jews have had many enemies over the millennia. Our enemies all wanted to destroy every one of us.

	Attribute Category
Evil	Correct
Anti-Semitic	Correct Nuance
Honest	Opposite
Nice	Opposite Nuance
Powerful	Situational
Misogynistic	Neutral
Smart	Neutral

Tower of Babel #1

The Tower of Babel once more looms . In which metafiction God , speaking in the plural of majesty , decides to thwart man 's titanic project by creating the languages of the world whereas there had been originally only one and thus sowing confusion that made further progress impossible .

	Attribute Category
Overambitious	Correct
Grandiose	Correct Nuance
Successful	Opposite
Fortunate	Opposite Nuance
Unprepared	Situational
Clever	Neutral
Alert	Neutral

Tower of Babel #2

Barford has photographed over 6,000 shop fronts in the process of making his ' Tower of Babel' , putting a 21st century spin on the biblical story from the Book of Genesis .

	Attribute Category
Overambitious	Correct
Grandiose	Correct Nuance
Successful	Opposite
Fortunate	Opposite Nuance
Festive	Situational
Clever	Neutral
Alert	Neutral

Tower of Babel #3

FRANK Dunlop , the Edinburgh Festival 's retiring director , last week accused its Fringe of degenerating into \" a third - rate circus \" , and being reminiscent of a modern Tower of Babel , Coney Island and Alton Towers .

	Attribute Category
Overambitious	Correct
Grandiose	Correct Nuance
Successful	Opposite
Fortunate	Opposite Nuance
Idiotic	Situational
Clever	Neutral
Alert	Neutral

Tower of Babel #4

Quentin Hardy of The New York Times said it well: The tech industry is trying \" to topple the Tower of Babel . \" He said that 80 to 90 percent of the web is in just 10 languages . Google , for one , has made it known that it is working toward this goal , and will soon announce updates to its translation app for phones .

	Attribute Category
Overambitious	Correct
Grandiose	Correct Nuance
Successful	Opposite
Fortunate	Opposite Nuance
Difficult	Situational
Clever	Neutral
Alert	Neutral

Tower of Babel #5

So , if enterprise service management feels like building the tower of Babel , just remember how Belgians have made that tower work in Brussels . Figure out how the other side works , and let the other side know how you work as well . Then you 're breaking down silos instead of building a tower .

	Attribute Category
Overambitious	Correct
Grandiose	Correct Nuance
Successful	Opposite
Fortunate	Opposite Nuance
Effective	Situational
Clever	Neutral
Alert	Neutral

Behemoth #1

Ocasio - Cortez said she has no regrets about the fight. The tech behemoth 's planned headquarters would helped drive up rents and drive longtime Queens’s residents out of their homes, she said.

	Attribute Category
Powerful	Correct
Evil	Correct Nuance
Good	Opposite
Improved	Opposite Nuance
Innovative	Situational
Beautiful	Neutral
Alert	Neutral

Behemoth #2

A first - time candidate, he said considered it his " civic duty " to make sure Gianaris had a challenge, and felt the incumbent opposed the tech behemoth for political reasons

	Attribute Category
Powerful	Correct
Evil	Correct Nuance
Good	Opposite
Improved	Opposite Nuance
Creative	Situational
Beautiful	Neutral
Alert	Neutral

Behemoth #3

The first real appearance of the Behemoth is not the proper entrance. A stiff, immobile, and completely unconvincing puppet begins dismantling a small boat (an obvious miniature).

	Attribute Category
Powerful	Correct
Evil	Correct Nuance
Good	Opposite
Improved	Opposite Nuance
Intimidating	Situational
Beautiful	Neutral
Alert	Neutral

Behemoth #4

We are in bad psychic shape. The behemoth is in his bathrobe. We need to suit up, get a grip. Because if there are more 9/11s coming at us, or retaliation here over Iraq.

	Attribute Category
Powerful	Correct
Evil	Correct Nuance
Good	Opposite
Improved	Opposite Nuance
Unprepared	Situational
Beautiful	Neutral
Alert	Neutral

Behemoth #5

In general, the PEs of the new Behemoth stocks were higher than those that were already Behemoths three years ago, pulling the median at least one multiple turn higher.

	Attribute Category
Powerful	Correct
Evil	Correct Nuance
Good	Opposite
Improved	Opposite Nuance
Profitable	Situational
Beautiful	Neutral
Alert	Neutral

Golem #1

ERLIN - A gigantic golem made out of wooden Hebrew letters lies motionless on the ground, yet it seems as if only a few magic whispers are needed to bring the creature to life.

	Attribute Category
Artificial	Correct
Driven	Correct Nuance
Compassionate	Opposite
Emotional	Opposite Nuance
Magical	Situational
Beautiful	Neutral
Temporary	Neutral

Golem #2

There are Golem hotels; Golem door - making companies; Golem clay figurines (made in China) ; a recent musical starring a dancing] Golem ; and a Czech strongman called the Golem who bends iron bars with his teeth.

	Attribute Category
Artificial	Correct
Driven	Correct Nuance
Compassionate	Opposite
Emotional	Opposite Nuance
Strong	Situational
Beautiful	Neutral
Temporary	Neutral

Golem #3

One of my favorite really early golem stories is about two rabbis, Rabbi Hanina and Rabbi Oshaya, who make a golem calf and they bring it to life and then they eat it 'cause they 're hungry.

	Attribute Category
Artificial	Correct
Driven	Correct Nuance
Compassionate	Opposite
Emotional	Opposite Nuance
Delicious	Situational
Beautiful	Neutral
Temporary	Neutral

Golem #4

Isaac Bashevis Singer once said, " I am not exaggerating when I say that the Golem story appears less obsolete today than it seemed 100 years ago. What are computers and robots of our time, if not golems?" If you think about it, Singer is correct. The myriad of ethical questions raised by artificial intelligence were hashed out centuries ago by rabbinic literature.

	Attribute Category
Artificial	Correct
Driven	Correct Nuance
Compassionate	Opposite
Emotional	Opposite Nuance
Intelligent	Situational
Beautiful	Neutral
Temporary	Neutral

Golem #5

Few here dispute that the Golem, who is often depicted as a menacing brown blob or an artificial humanoid, has become a lucrative global brand. But it is a profound irritation to Prague 's Jewish leaders that the Maharal 's legacy has been hijacked by a powerful dunce whom the Talmud characterizes as a fool.

	Attribute Category
Artificial	Correct
Driven	Correct Nuance
Compassionate	Opposite
Emotional	Opposite Nuance
Skilled	Situational
Beautiful	Neutral
Temporary	Neutral

Haman #1

I also recommend that he read the Parsha of Ki Tetzei over and over again. Perhaps, he will understand better the commandment to " Remember, and to not forget" that Amalek applies to Iran and its proxies today. Even Mordechai lamented the fact that there were Jews in his day who did not see Haman 's evil and plan for " the Final Solution.

	Attribute Category
Evil	Correct
Conniving	Correct Nuance
Honest	Opposite
Caring	Opposite Nuance
Clever	Situational
Powerful	Neutral
Intelligent	Neutral

Haman #2

Rabbi Alexander Zusslin HaCohen, a German Rabbi from the 14th century, taught that when we examine the passage in the Talmud that says a person is supposed to get so intoxicated that he doesn't know the difference between " Blessed is Haman " and " Cursed in Mordechai," the phrase " Blessed is Mordechai " is equal to 502 and that " Cursed is Haman" is also equal to 502.

	Attribute Category
Evil	Correct
Conniving	Correct Nuance
Honest	Opposite
Caring	Opposite Nuance
Holy	Situational
Powerful	Neutral
Intelligent	Neutral

Kiddush Hashem #1

It is unfortunate Jesus also died as a Jewish martyr for Kiddush Hashem, for the sanctification of His Divine Name.

	Attribute Category
Honorable	Correct
Devout	Correct Nuance
Stingy	Opposite
Subdued	Opposite Nuance
Sad	Situational
Popular	Neutral
Important	Neutral

Kiddush Hashem #2

Those who say that European Jews went to their deaths like sheep to the slaughter are guilty of a horrible calumny. Animals are not capable of dying al kiddush Hashem, for the sanctity of God 's name.

	Attribute Category
Honorable	Correct
Devout	Correct Nuance
Stingy	Opposite
Subdued	Opposite Nuance
Violent	Situational
Popular	Neutral
Important	Neutral

Kiddush Hashem #3

These boys, our boys have died Al Kiddush Hashem in sanctification of God 's name simply because they are Jews.

	Attribute Category
Honorable	Correct
Devout	Correct Nuance
Stingy	Opposite
Subdued	Opposite Nuance
Violent	Situational
Popular	Neutral
Important	Neutral

Kiddush Hashem #4

Acknowledging that Gil- Ad, Eyal and Naftali died al Kiddush Hashem helps us to see them as heroes and to be inspired by their sacrifice, but it doesn't remove the awful pain - we remain broken- hearted.

	Attribute Category
Honorable	Correct
Devout	Correct Nuance
Stingy	Opposite
Subdued	Opposite Nuance
Tragic	Situational
Popular	Neutral
Important	Neutral

Kiddush Hashem #5

Whether it was in Jewish public life, in the House of Lords, in the pulpit of the synagogue, at the podium at a lecturer on medical ethics, in the study hall of Torah and Talmud, he was a symbol of what the definition of kiddush Hashem was in the modern rabbinate. As such he was and will remain an inspiration for the countless rabbis who were inspired by his presence on the Jewish scene.

	Attribute Category
Honorable	Correct
Devout	Correct Nuance
Stingy	Opposite
Subdued	Opposite Nuance
Pure	Situational
Popular	Neutral
Important	Neutral

Leviathan #1

Tablelands resident Philip Jones discovered the winged leviathan] and could not stop himself from snapping a photo as he offered up his own $ 20 campaign contribution .

	Attribute Category
Powerful	Correct
Evil	Correct Nuance
Good	Opposite
Improved	Opposite Nuance
Beautiful	Situational
Innovative	Neutral
Alert	Neutral

Atabey – 1

There are figures of Atabey, and her son, Yocahu, as well as other cemis, or spirits.

Attributes:

1. Motherly
2. Fertile
3. Sickly
4. Corrupt
5. Worshipped
6. Old
7. Impartial

Chupacabra – 1

For “Street Smarts – Boliche Boulevard and Beyond”, we posed six pretty easy Latino heritage questions to more than a dozen residents and tourists on Seventh Avenue in Ybor City. Good sports all, sadly, most didn’t know a Chupacabra from a bowl of arroz con pollo.

Attributes:

1. Legendary
2. Mythical
3. Tangible
4. Existent
5. Appetizer
6. Nocturnal
7. Interesting

Chupacabra – 2

Chimichurri is a sauce. A spicy and piquant green sauce from our friends in South America. They use it mostly to marinate or serve with grilled beef, but it goes great with just about anything else: chicken, fish, vegetables, Chupacabra. Anything.

Attributes:

1. Legendary
2. Mythical

3. Tangible
4. Existent
5. Protein
6. Nocturnal
7. Interesting

Coqui – 1

Legend has it that the coqui will die of a broken heart if taken off its island. But the little frog has no such affection for its own tadpole – it doesn't have one.

Attributes:
1. Adored
2. Significant
3. Bad Omen
4. Unpleasant
5. Unusual
6. Loud
7. Small

Coqui – 2

You can go, for different reasons, but when you go back to the island it's always the same, you fill up with a strange feeling, a melancholy, because you sart making the film of everything you've lived, streets, and I meet the neighbors who saw me very little. Ahhh… I'm from there like the coqui, so you know.

Attributes:
1. Adored
2. Significant
3. Rare
4. Strange
5. Native
6. Loud
7. Small

Coqui – 3

The frog was also unique to Puerto Rico, until it was transported to Hawaii. It flourishes there, too, but that little frog sings all night long. Coqui, coqui.

Attributes:

1. Adored
2. Significant
3. Annoying
4. Unpleasant
5. Loud
6. Unique
7. Small

Coqui – 4

As the coqui signals yet another night of darkness, Sylvia Martizen sees from her kitchen window the glowing lights of San Lorenzo in the valley. She takes advantage of the last bit of daylight to reheat the dinners her family prepares for her. She puts a bag of ice in a cooler.

1. Adored
2. Significant
3. Bad Omen
4. Unpleasant
5. Time-Watch
6. Loud
7. Small

Coqui – 5

We got lost driving back to Rincon but just like every other time, we found our way after a few wrong turns and directions from the locals. An early flight the next morning precluded a last swim in the sea, but we did have one final experience: The coquis made noise in the predawn darkness.

Attributes:

1. Adored
2. Significant
3. Annoying
4. Unpleasant
5. Loud

6. Impressive
7. Small

Divina Providencia – 1

A painting of La Virgen de la Divina Providencia watched over the elements for the Eucharist with the island's flag – red and white stripes with a blue triangle emblazoned with a white star – draped behind her.

Attributes:
1. Protective
2. Kind
3. Evil
4. Bad Omen
5. Symbolic
6. Uninterested
7. Impressive

Juan Bobo – 1

The Adventures of Juan Bobo: Open Eye Figure Theatre performs this puppet show about Juan Bobo's quest to set things right after three devils steal love, food, and music from the world.

Attributes:
1. Funny
2. Foolish
3. Responsible
4. Intelligent
5. Good Natured
6. Hyperactive
7. Good Looking

Juan Bobo – 2

Open Eye Figure Theatre's original show, performed with live music in a mix of English and Spanish, is based on Juan Bobo. When the world is turned upside down by three devils, Juan Bobo is sent off to make things right, getting into scrapes and mischief along the way.

Attributes:
1. Funny
2. Foolish
3. Responsible
4. Intelligent
5. Good Natured
6. Hyperactive
7. Good Looking

Juan Bobo – 3

The Adventures of Juan Bobo is an original story inspired by a Puerto Rican folk character, Juan Bobo. The world turns upside down when three devils come to his village and Juan Bobo must set things right again.

Attributes:
1. Funny
2. Foolish
3. Responsible
4. Intelligent
5. Good Natured
6. Hyperactive
7. Good Looking

Reyes Magos – 1

Today, long after gifts from Old Saint Nick have been unwrapped and Christmas trees have been put away, many Hispanic children are preparing for a visit from Los Reyes Magos.

Attributes:
1. Festive
2. Magical
3. Tangible
4. Existent
5. Famous

6. Important
7. Powerful

Reyes Magos – 2

Original Story: The family put hay and water under their beds Saturday night to feed the camels that would bring Los Reyes Magos. They also put their best shoes under the bed in anticipation of gifts.

Attributes:

1. Festive
2. Magical
3. Tangible
4. Existent
5. Animal
6. Important
7. Powerful

Three Kings – 1

“If you go to Puerto Rico now, you see the Three Kings signs all over.”, says Ortiz. “It’s still the most important day there.” Richard Vasquez of the newsletter Las Culturas says many immigrants hold to the Three Kings tradition when they arrive in the U.S.

Attributes:

1. Festive
2. Magical
3. Tangible
4. Existent
5. Monarchy
6. Important
7. Powerful

Three Kings – 2

These families still observe Christmas Eve and Christmas Day with reverence and solemnity and presents, but on January 6th, they rejoice in the visit of the three kings.

Attributes:
1. Festive
2. Magical
3. Tangible
4. Existent
5. Monarchy
6. Important
7. Powerful

Three Kings – 3

Carlos Martinez, executive director of the Chicano Humanities and Arts Council in Dever, remembers putting out his shoes as a child in Texas. But the pull of two traditions puzzled him, and three kings faded from his family celebrations.

Attributes:
1. Festive
2. Magical
3. Tangible
4. Existent
5. Monarchy
6. Important
7. Powerful

Three Kings – 4

"Besides the regular Nativity scene, it has the Three Kings.", he said. "One is carrying a small tree frog – a coqui – that is capable of seeing at nighttime. Another is bringing a Puerto Rican parrot. Then there are children with chickens, tropical fruits, and other things they are offering to the child.

Attributes:
1. Festive
2. Magical
3. Tangible
4. Existent
5. Worshipped
6. Important

7. Powerful

Three Kings – 5

Many of the santos are grouped in such scenes as the flight into Egypt, the Holy Family, the Nativity, the three kings and the Crucifixion.

Attributes:
1. Festive
2. Magical
3. Tangible
4. Existent
5. Worshipped
6. Important
7. Powerful

Yocahu – 1

There are figures of Atabey, and her son, Yocahu, as well as other cemis, or spirits.

Attributes:
1. Holy
2. Religious
3. Evil
4. Corrupt
5. Large
6. Impartial
7. Old

Children of Lir

Story

My Ireland can be hard to take, asks, ' Did St Patrick banish all the snakes? ' My Ireland is the Children of Lir, a herd of deer and a Connacht brogue.

Attributes

Attribute	Order
Tragic	Correct
Jealousy	Correct Nuance
Joyful	Opposite
Familial	Opposite Nuance
Young	Situational
Beautiful	Neutral
Charming	Neutral

Cuchulainn

Story 1

The Northern Ireland Tourist Board leaflet chirpily points out, even Cuchulainn would want to arrange a few overnight stops before setting out.

Attributes 1

Attribute	Order
Powerful	Correct
Monstrous	Correct Nuance
Weak	Opposite
Mortal	Opposite Nuance
Celebrity	Situational
Rare	Neutral
Annoying	Neutral

Story 2

Turn the figure left through 90 degrees and it looks like a pieta; right way up, Sheppard 's Cuchulainn with his pageboy hair, like a Victorian stage wig, most of all has the appearance of a tableau from a pageant, something Patrick Pearse, an admirer of Sheppard 's work, might have put on at St Enda 's.

Attributes 2

Attribute	Order
Powerful	Correct
Monstrous	Correct Nuance
Weak	Opposite

Mortal	Opposite Nuance
Artwork	Situational
Rare	Neutral
Annoying	Neutral

Story 3

The legends have been divided into cycles, and the Ulster cycle here is dominated by the figure of Cuchulainn, who remains a powerful symbolic force in contemporary Ireland. A statue of Cuchulainn stands in the General Post Office in Dublin to commemorate the 1916 rebellion. But, ironically, for certain mythologically minded Protestants in Northern Ireland, especially among the UDA, Cuchulainn is an important symbol, too: northern, ruthless, fearless, murderous, implacable in the face of death. I often feel that if we had to sacrifice one figure from these legends, I could easily do without Cuchulainn.

Attributes 3

Attribute	Order
Powerful	Correct
Monstrous	Correct Nuance
Weak	Opposite
Mortal	Opposite Nuance
Statue	Situational
Rare	Neutral
Annoying	Neutral

Story 4

Meanwhile, Cuchulainn abides, overseeing people buying stamps in the GPO, edging into the imagination of poets , playwrights and novelists . Carson notes that Horslips produced an album called the Tain in 1973. But musically there was more.

Attributes 4

Attribute	Order
Powerful	Correct
Monstrous	Correct Nuance
Weak	Opposite
Mortal	Opposite Nuance
Celebrity	Situational
Rare	Neutral
Annoying	Neutral

Finn McCool

Story 1

In the shadow of Harland and Wolff 's Samson and Goliath a site for a new public art project is being discussed, and I have submitted my own proposal for a 10 - meter red - hand sculpture. It will have titanic proportions with a hint of Finn McCool for a country that thrives on the giant and the mythical.

Attributes 1

Attribute	Order
Talented	Correct
Smart	Correct Nuance
Lazy	Opposite
Greedy	Opposite Nuance
Sculpted	Situational
Loud	Neutral
Festive	Neutral

Story 2

The grounds of Glenarm Castle Estate are set to quake this July as teams of leading strongmen go head-to-head in a battle worthy of the legendary Finn McCool.

Attributes 2

Attribute	Order
Talented	Correct
Smart	Correct Nuance
Lazy	Opposite
Greedy	Opposite Nuance
Strongman	Situational
Loud	Neutral
Festive	Neutral

Story 3

I do enjoy the odd holiday abroad and I 'm not ashamed to admit I have a sports car that drinks petrol like Shane McGowan would down a pint of Guinness. I have to be honest that my carbon footprint would be the size of Finn McCool 's right foot.

Attributes 3

Attribute	Order
Talented	Correct
Smart	Correct Nuance
Lazy	Opposite
Greedy	Opposite Nuance
Polluting	Situational

Loud	Neutral
Festive	Neutral

Story 4

In those days, Shark was no more than the schoolboy evoked by Mullins, sooner into hurling than horses. In fact, that was where he acquired the nickname, playing for the parish in a Co Kilkenny semi - final, a young Finn McCool with limbs that might bestride a bullock.

Attributes 4

Attribute	Order
Talented	Correct
Smart	Correct Nuance
Lazy	Opposite
Greedy	Opposite Nuance
Sports Player	Situational
Loud	Neutral
Festive	Neutral

Story 5

So, what in the name of Finn McCool is going on here? Something is definitely afoot at the Giants office.

Attributes 5

Attribute	Order
Talented	Correct
Smart	Correct Nuance
Lazy	Opposite
Greedy	Opposite Nuance
Coach	Situational
Loud	Neutral
Festive	Neutral

Leprechaun

Story 1

“The Leprechaun 's Gold " by Pamela Duncan Edwards (Harper Trophy, $ 15.99): In this story, an Irish musician tries to ruin his friend 's chances of winning a harp contest . Everything changes when a magical leprechaun comes into play.

Attributes 1

Attribute	Order
Tricky	Correct
Mischievous	Correct Nuance
Skillful	Opposite
Strong	Opposite Nuance
Novel	Situational
Festive	Neutral
Religious	Neutral

Story 2

The Irish Times headlined a story about the museum before it opened in 2010 as " The Louvre of Leprechauns”, while the journalist wrote: " Frankly, I was expecting it to be a little more leprechauny , and a lot less Neolithic.

Attributes 2

Attribute	Order
Tricky	Correct
Mischievous	Correct Nuance
Skillful	Opposite
Strong	Opposite Nuance
Historic	Situational
Festive	Neutral
Religious	Neutral

Story 3

But Houston 's local leprechaun is calling it quits after this year, the 50th anniversary of the hotel 's opening.

Attributes 3

Attribute	Order
Tricky	Correct
Mischievous	Correct Nuance
Skillful	Opposite
Strong	Opposite Nuance
Texan	Situational
Festive	Neutral
Religious	Neutral

Story 4

Michael Connolly, CSO spokesman, insisted their figures were correct, and had been in the past, despite international derision from economists such as Paul Krugman, who dubbed 2015 's report of 26 % growth ' leprechaun economics ' based on figures that included multinational companies.

Attributes 4

Attribute	Order
Tricky	Correct
Mischievous	Correct Nuance
Skillful	Opposite
Strong	Opposite Nuance
Theoretical	Situational
Festive	Neutral
Religious	Neutral

Story 5

The Leprechaun Who Wished He wasn't by Siobhan Parkinson is about an 1,100 - year - old leprechaun who wants to be tall.

Attributes 5

Attribute	Order
Tricky	Correct
Mischievous	Correct Nuance
Skillful	Opposite
Strong	Opposite Nuance
Old	Situational
Festive	Neutral
Religious	Neutral

Shamrock

Story 1

Chinese authorities are obviously far more worried about matters jasmine - related than anything to do with the shamrock] , so the police withdrew permission for the annual parade in Shanghai because they do not want any public gatherings taking place in sensitive areas .

Attributes 1

Attribute	Order
Symbolic	Correct
Christian	Correct Nuance
Bad Omen	Opposite
Made for Holidays	Opposite Nuance
Illegal	Situational

Plant-like	Neutral
Lucky	Neutral

Story 2

In his work , Boke of the Histories of Irelande, he claims that the ' wild Irish ' were given to eating the plant . He wrote : ' Shamrotes , watercresses , rootes and other herbes they feed upon . ' / The shamrock features on the passport stamp of the Caribbean island of Montserrat , which has a large proportion of citizens of Irish descent .

Attributes 2

Attribute	Order
Symbolic	Correct
Christian	Correct Nuance
Bad Omen	Opposite
Made for Holidays	Opposite Nuance
Edible	Situational
Plant-like	Neutral
Lucky	Neutral

Story 3

While people associate the three - leafed shamrock with Ireland , the Irish Republic 's official harp , appears on Irish coins and government publications . However , Ireland 's tourism board and national airline , Aer Lingus , use shamrocks.

Attributes 3

Attribute	Order
Symbolic	Correct
Christian	Correct Nuance
Bad Omen	Opposite
Made for Holidays	Opposite Nuance
Corporate	Situational
Plant-like	Neutral
Lucky	Neutral

Story 4

But the shamrock caused no fear in Washington DC , where President Bush accepted a bowl of shamrock from Taoiseach Bertie Ahern] .

Attributes 4

Attribute	Order

Symbolic	Correct
Christian	Correct Nuance
Bad Omen	Opposite
Made for Holidays	Opposite Nuance
Presidential	Situational
Plant-like	Neutral
Lucky	Neutral

Story 5

The flag of Montreal has a shamrock in the lower right quadrant to represent the Irish population - one of the four major ethnic groups living in the city in the 19th Century] . The other images on the flag are a blue fleur - de - lys (to represent the French) , a red Rose of Lancaster (the English and Welsh) and a thistle (the Scots) .

Attributes 5

Attribute	Order
Symbolic	Correct
Christian	Correct Nuance
Bad Omen	Opposite
Made for Holidays	Opposite Nuance
Canadian	Situational
Plant-like	Neutral
Lucky	Neutral

Tir na Nog

Story

My Ireland can be hard to take, asks, ' Did St Patrick banish all the snakes? ' My Ireland is the Tir na nOg a herd of deer and a Connacht brogue.

Attributes

Attribute	Order
Fantastical	Correct
Tragic	Correct Nuance
Down-to-earth	Opposite
Joyful	Opposite Nuance
Full of Nature	Situational
Beautiful	Neutral
Charming	Neutral

Baseline for motif recognizability

S.NO.	MOTIF	CONTROL PROMPT –NON-MOTIFIC USAGE
1.	Amalek	“Defeat Of Amalek” from the album “Not Dead Yet” has one of my favorite metal riffs.
2.	Atabey	Atabey is looking into expanding its facilities by building two new cancer research centers.
3.	Behemoth	Behemoth is a Polish extreme metal band from Gdask, considered to have played an important role in establishing the Polish extreme metal underground.
4.	Chupacabra	Jose got a new dog for his birthday and named him Chupacabra.
5.	Coqui	Scientists have been studying the unique reproduction cycle of Coquis.
6.	Cuchulainn	“The Boys' Cuchulain” by Eleanor Hull, with fifteen illustrations in colour by Stephen Reid, was one of the new additions to the library.
7.	Divina Providencia	Our Lady of Providence Church is open every day from 8am-7pm with daily mass starting at 10am.
8.	Finn McCool	“Find out what happened when Gordon Ramsay visited Finn McCool's and read about when and why Finn McCool's closed.”
9.	Golem	"The Golem", directed by Doron and Yoav Paz, released in 2018.
10.	Haman	The family moved to Seoul from a small town of Haman to look for additional job opportunities.
11.	Juan Bobo	The university recently hired Dr. Juan Bobo as a new faculty member.
12.	Kiddush	The Rambam, based on the Talmud, says the mitzvah of kiddush Hashem is fulfilled when a person is placed in a situation where they have to give up their life for the sake of God.
13.	Three Kings	In English history, The Year of the Three Kings may refer to the years 1066, 1483, or 1936.
14.	Leprechaun	People love to dress as leprechauns for Halloween.
15.	Leviathan	During WWI the Leviathan ship carried thousands of American troops across the Atlantic Ocean, infested with submarines and mine fields, the ship remained unscathed by any attack.
16.	Shamrock	Shamrock refers to either the lesser clover or the white clover, in addition to some other three-leaved plants.
17.	Children of Lir	The children of Lir, 32, from Ireland, became famous due to a viral video on YouTube.
18.	Tir Na Nog	“Tir Na Nog Bistro & Bar is a hip neighborhood restaurant that boasts a relaxing, yet upscale environment."
19.	Tower of Babel	The Tower of Babel by Bruegel is a masterpiece.
20.	Yocahu	If you are looking for a more exciting vacation, Villa Yocahu also offers a full bar with kegarator and cinema room with reclining leather seats and 150" screen.

APPENDIX D - INSTRUMENTS OF THE RITUAL SURVEY

For this study, the same questionnaire was repeated for the six different rituals:

1. Birth
2. Coming of Age
3. Birthday
4. New Year
5. Wedding
6. Funeral

Please see the following page for the questionnaire.

Ritual Survey

General Event Related Questions

1. Where does this event typically happen? Textbox
2. When does this event typically happen? Textbox
3. How long does this event typically last? Textbox + Dropdown for unit of time
4. Who are the main people involved in this event? Option to add as many people as needed

   Person1 → Textbox

   Person 2 → Textbox

   Person3 → Textbox

   …

   PersonN → Textbox

5. Is there one primary person for this event? Yes/No
6. Who is the primary person for this event? Select from options of Q4

PersonN

Before the event

7. Does PersonN typically have an intent in causing the event? Yes/No
8. What is PersonN's typical intent in causing the event? Textbox
9. Does PersonN typically need to do anything before this event? Yes/No
10. What does PersonN typically need to do before this event? Textbox

During the event

11. Does PersonN typically use something during the event? Yes/No
12. What things does this person typically use during this event? Textbox

After the event

13. How would PersonN be described as a consequence of the event? Textbox
14. Does PersonN typically want to do something after this event? Yes/No
15. What does PersonN typically do after this event? Textbox
16. What does PersonN typically feel after this event? Textbox

Repeat these questions for each person mentioned in Q4.

VITA

ANURAG ACHARYA

January 1, 1993	Born; Butwal, Lumbini, Nepal
2015	B.A., English & Political Science Ratna Rajya Campus, Tribhuvan University Kathmandu, Bagmati, Nepal
2016	B.E., Computer Engineering Kathmandu Engineering College, Tribhuvan University Kathmandu, Bagmati, Nepal
2021	M.S., Computer Science Florida International University Miami, Florida, USA

PUBLICATIONS AND PRESENTATIONS

Acharya, A., Talamadupula, K., & Finlayson, M. A. (2021). *Towards an Atlas of Cultural Commonsense for Machine Reasoning*. Proceedings of the Workshop on Common Sense Knowledge Graphs (CSKGs), The Thirty-Fifth AAAI Conference on Artificial Intelligence.

Yarlott, W. V. H., Ochoa, A., Acharya, A., Bobrow, L., Castro-Estrada, D., Gomez, D., Zheng, J., McDonald, D., Miller, C. & Finlayson, M. A. (2021). *Finding Trolls Under Bridges: Preliminary Work on a Motif Detector*. Advances in Cognitive Systems.

Yarlott, W. V. H., Ochoa, A., Acharya, A., Bobrow, L., Castro-Estrada, D., Gomez, D., Zheng, J., McDonald, D., Miller, C. & Finlayson, M. A. (2021). *Finding Trolls Under Bridges: Preliminary Work on a Motif Detector*. Advances in Cognitive Systems.