2013

Toward digitizing the human experience : a new resource for natural language processing

Jerry Scott Weltman
Louisiana State University and Agricultural and Mechanical College

TOWARD DIGITIZING THE HUMAN EXPERIENCE: A NEW RESOURCE FOR NATURAL LANGUAGE PROCESSING

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

In

The School of Electrical Engineering and Computer Science

by
Jerry Scott Weltman
B.A., University of Texas, 1984
M.S., Stanford University, 1985
May 2013

ACKNOWLEDGEMENTS

I would like to thank the faculty and staff in computer science and linguistics at Louisiana State University for their enthusiastic support of my interdisciplinary studies, and I especially thank my co-chairs Sitharama Iyengar and Michael Hegarty for their willingness to share mentorship across departments. I also thank my committee member, Jianhua Chen, for her invaluable counsel. I very much appreciate Coretta Douglas, David Trammel, Vera Watkins, and Maggie Edwards, who gave generously of their time and helped me avoid the pitfalls of graduate study. I am indebted to Lisi Oliver, Ann Mulkern, and Sukhamay Kundu, who helped shape this dissertation in both their seminars and their extra consultations. As a member of two departments, I was fortunate to have two communities of students, and I send thanks to Claudia Gil, Timur Gilmanov, Harry Moore, Alfred Samman, Forrest Osterman, Danielle Alfandre, Mona Alkadi, Clevia Pérez, Jane Richardson, Tracy Wories, Tina Villa, Anne Reynolds-Case, Jaïmé Dubé, Daniela Katunar, Tena Gnjatovic and Srivathsan Srinivas, with a special shout out to Ni Tyjah Thigpen for letting me know when to shut up.

A dissertation can be a lonely business, and I am eternally grateful to Margit Link-Rodrigue, Stephen Elmer, and Mike Boechler. Each one of them contributed key insights to this project and kept me from falling down the rabbit hole when trying to model something so "simple" as common sense.

Finally, I thank my wife Sharon and children Alex and Elizabeth, for accepting a newcomer into the family – a little boy named Max who likes to drop vases. With their love and encouragement, I was able to switch careers and explore the fascinating and maddening effects when language meets computation.

TABLE OF CONTENTS

ABSTRACT

A long-standing goal of Artificial Intelligence is to program computers that understand natural language. A basic obstacle is that computers lack the common sense that even small children acquire simply by experiencing life, and no one has devised a way to program this experience into a computer. This dissertation presents a methodology and proof-of-concept software system that enables non-experts, with some training, to create simple experiences. For the purposes of this dissertation, an experience is a series of time-ordered comic frames, annotated with the changing intentional and physical states of the characters and objects in each frame. Each frame represents a small action and the effects of that action. To create an annotated experience, the software interface guides non-experts in identifying facts about experiences that humans normally take for granted. As part of this process, it uses the Socratic Method to help users notice difficult-to-articulate commonsense data. The resulting data is in two forms: specific narrative statements and general commonsense rules. Other researchers have proposed similar narrative data for commonsense modeling, but this project opens up the possibility of non-experts creating these data types. A test on ten subjects suggests that non-experts are able to use this methodology to produce high quality experiential data. The system's inference capability, using forward chaining, demonstrates that the collected data is suitable for automated processing.

INTRODUCTION

In the early days of artificial intelligence, AI researchers were optimistic about being able to create robots or intelligent agents that could engage in human-level conversation. But, as AI critic and philosopher Hubert Dreyfus points out, they soon realized that the problem was much more difficult than they had imagined:

> The trouble started with the failure of attempts to program an understanding of children's stories. The programs lacked the common sense of a four-year-old, and no one knew how to give them the background knowledge necessary for understanding even the simplest stories (Dreyfus, 1992).

A significant problem blocking progress in story understanding is that virtually all natural language communication, and especially stories, contain gaps between what is stated and what is understood. Consider this simple story:

1) *Max was on the sofa, bored, all by himself. There was a pretty vase on a little side table. He went there and picked it up. He dropped it. Crash! This was fun!*

This story has a lot of missing information that even the youngest readers could fill in from their own life experiences, but a computer cannot:

- Max is probably a little boy
- Max is probably in a room
- At the start of the story, Max is probably in a sitting position[1]
- Max sees the vase before going to it
- In order to pick up the vase, Max probably walks to the side table
- To walk to the vase, Max first stands up

Children have accumulated a vast bank of life experiences by the time they start reading. They may not even know what porridge is, but they know what it is like to be hungry. They may never have met a wolf, but they have faced fears. They may never have picked up a vase, but they have

[1] In some cultures, the presence of a sofa might indicate royalty or wealth, and the usual position might be lying down rather than sitting. Thus, filling in the gaps depends partially, but significantly, on a presumed shared culture.

desired things and have strategized about how to get them. In short, they are able to project their own experiences onto plots of stories to understand the goals of characters and to fill in the unstated information.

Recognizing that assumed life experiences are inherent in normal human communication, can we program simple experiences into a computer to help it understand children's stories? Progress in this area would benefit many practical natural language processing tasks such as language translation, text summarization, information extraction, and general interactions between humans and computers. This area of study falls under the category of natural language processing (NLP), a subfield of artificial intelligence (AI).

The technology to record human experiences is already available. More and more people are recording their life activities on social media with pictures and videos. Disk space is cheap enough that a person could capture every moment of his or her life. But these recordings are "analog" information, a term borrowed from the ideas of Dretske (1999) about how humans convert ordinary external information into digital information. Straight recordings, even on a digital medium, would not be useful for NLP unless the salient aspects of the human experience are captured and given a coherent structure. The goal of a literal digitizing process is to extract enough features of the analog signal so that it can be reproduced. It periodically samples the signal and converts salient continuous data into time-ordered discrete values. The goal of a "digitized" human experience would be to extract features that explain what is happening in an everyday experience. The digitizing metaphor here emphasizes the idea that experiential narratives would not be traditional literary stories and plots[2], but a series of situational time-

[2] Of course, all actions have some type of plot. The "Max breaks the vase" experience could be analyzed in terms of a variation on Lehnert's "Fleeting Success" plot unit (Lehnert 1981): "Max achieves his quest for amusement; however, when Mommy sees what he has done, he will be sorry." However, the goal in this dissertation is to capture experiential details rather than plot.

slices annotated with the critical data needed to explain who is doing what, and why they are doing it.

I.1 The Human Experience Project

Would it be possible to create a project analogous to Wikipedia, where instead of encyclopedic articles about human knowledge, people create digitized human experiences that would be useful for NLP? There would be many challenges to such a project besides the usual ones associated with any large-scale Internet collaboration:

- When people describe experiences, they naturally leave out the mundane detail that is obvious to humans but critical to NLP. How can we get people to articulate assumptions that they may not even realize they are making?
- Depending on the circumstance, every sight and sound could be important, from the dust on a window sill to the sound of a person's footsteps as they cross a room. A random thought or distant memory could turn out to be critical. How do we decide what data are important to capture in an experience?
- Professional philosophers, psychologists, and AI workers are mired in modeling and formalizing experiential knowledge. How can we expect non-experts to interpret experiences?

This dissertation proposes to help non-experts create experiential narratives towards digitizing the human experience. To help non-experts overcome the challenges of articulating experiential data, deciding what is relevant, and interpreting events, the proposed method employs three novel techniques: 1) It models experience as a sequence of still frames, like the still images of an animation flip book. When animated, the frames help users notice missing details about actions; if the movement appears too abrupt, more frames (with more detail) need to be added. 2) It asks users to focus on intention, location, and movement – information that is critical to a commonsense understanding of the situation. 3) It asks users to explain the reason behind each description. Similar to the Socratic Method, it displays the users' answers as a general rule, which exposes their assumed commonsense information and encourages deeper explanations.

The overall project, called the Human eXperience Project (HXP), is a methodology and corresponding software framework that enables non-experts to create detailed narratives of simple everyday experiences. Using Singh's ideas to collect specific experiences rather than abstract life scripts (Singh & Barry, 2003, Singh et al. 2004), HXP asks users to describe situations at a specific time and place. There is no "right" way to describe an experience; all interpretations are welcome. In line with the goals of McCarthy et al. (2002), HXP focuses on *simple* experiences—activities and naïve mental states that one would expect to find in stories at the level of kindergarten or first grade. As the McCarthy group suggests, concentrating on the knowledge found in children's stories helps make story understanding more tractable. HXP would eventually be part of a wiki web site, which means contributors view, discuss, and edit each other's work. The experiential narratives that contributors create would be open for discussion and would undergo many refinements as contributors hash out their meaning.

I know of no prior work that specifically focuses on collecting highly detailed child-centered experiences from non-experts. I believe that such a corpus would be a boon to statistically-oriented NLP, providing valuable training data and new correlations between actions, intention, and location. It would also help provide the raw data to develop new types of architectures for deep semantics and commonsense reasoning algorithms (Aamodt et al., 1994; Fahlman, 2011; Laird, 2012; Minsky, 2007; Mueller, 2006; Schubert, 2006; Zarri, 2010).

For this dissertation, I have implemented a proof-of-concept software program to explore whether non-experts can provide the sort of detailed commonsense data that traditionally has been manually encoded by experts. I then conducted a test on ten subjects. The results indicate that some non-experts are able to use this method to produce high quality experiential data.

I.2 Chapter Descriptions

Chapter 1 presents an overview of the methodology and the types of data to be collected. It shows the comic frames of a sample experiential narrative "Max breaks the vase" that will be referred to throughout this dissertation. Starting with the background, characters, and props, it takes the reader through the entire annotation process. The overview includes a sample screen with a series of drop-down selections to guide the user in creating statements that describe the frame's action and results of the action. It also provides two extended examples with sample screens showing how HXP converts specific explanations into general If-Then rules. These rules are part of a Socratic Method that restates a person's belief as a general rule in order to highlight inconsistencies or missing information. When users see their explanations in this form, they often add more detailed statements to the narrative. The result of this process is that the user creates highly detailed narratives that provide useful commonsense data. The narrative statements are also likely to be relevant to the experience because the statements perform a dual role: (1) they describe the current key frame, and (2) they provide an explanation for subsequent statements.[3] As an important parallel activity, users also create rules that generalize a specific situation to a more general one, crucial for applying the narrative data to a wide variety of situations.

[3] In the future, I suspect we may want to notify users when a statement is irrelevant (not needed to explain the experience). Similar to a compiler detecting unused variables in a computer program, the system could detect unneeded statements by checking if they are part of at least one explanation.

Chapter 2 discusses research related to collecting narratives and modeling commonsense data, especially narratives. In the early 1990s, large corpora of labeled text and digitized lexical resources helped shift NLP research away from deep semantics towards statistical methods. However, recent success of IBM's Watson program (Baker, 2011) shows that statistical methods make use of a wide variety of resources, including logical rules. Offering both logical rules and structured narrative data, HXP would be a valuable new resource. The idea of collecting structured narrative data for NLP is not new, starting with scripts and narrative frames (Schank & A belson, 1977; Minsky,1974). More recently, Elson & McKeown (2010) proposed collecting a corpus of annotated stories, but their structures and methodology are geared toward narratology, not commonsense modeling. HXP directly inherits Singh's idea to collect everyday narratives from untrained users (Singh, 2005), but HXP provides a path to collect the detailed narrative statements and commonsense rules that have been previously created only by experts.

Chapter 3 presents HXP's underlying model of experience. It first describes specific narrative statements and addresses the many design decisions involved in asking non-experts to describe an experience. What are the critical components to capture? How much detail is necessary and practical? How do we model what is true at any given point of the narrative? Unlike a story or newspaper article that focuses on what is interesting, HXP targets the mundane. However, rather than collect objective data that could be obtained from a 3D visual model, HXP seeks subjective data that, while imprecise, nevertheless drives the actions in the narrative. A community of users will have different points of view about a situation, but rather than try to get a consensus, HXP allows the same actions to have multiple interpretations. Redundant data from the community are also valuable because they help strength associations between words and their most common usage.

The second part of the chapter describes the structure of HXP rules. The rules are defeasible Horn clauses qualified with a level of confidence. To demonstrate that the rules are suitable for automatic processing, the HXP implementation includes a simple inference process that predicts what else might be true from a given statement. This process brings out issues about how to match specific statements with general rules, and how to assign priorities when there is more than one candidate rule. In order to match statements, HXP defines a partial order, based on the hypernym relation. Using this partial order, HXP computes the closest match, taking into account the rules' confidence levels.

Chapter 4 describes the design of the HXP user interface, with the goal of making it easy for users to describe experiences through a series of drop-down selections. For this purpose, HXP uses a controlled natural language with a limited number of predicates and argument structures. In an effort to help the user find the appropriate predicate, HXP has several categories and subcategories of predicates. However, the users unanimously preferred to simply type in a word rather than drill down through a series of choices. The chapter shows screen shots of different input screens and discusses how HXP prompts for different information depending on the type of object or the type of predicate. To make the user interface able to handle a wide variety of predicates, as well as new ones that will be added, the software is designed around modular structures called *value templates* that can be combined to issue the appropriate input prompt at the appropriate time. The top-level value templates are presented.

Chapter 5 describes a ten person user test to evaluate the methodology. After about two hours of training and about 1.5 hours of independent annotation, two subjects were not able to contribute quality data, four were able to contribute but found the process very tedious, and four were able contribute and found the process challenging and fun. A panel of three judges (two of the test subjects and myself) evaluated the quality of the 96 total commonsense rules that the subjects had qualified with *probably* or *definitely*, and found that 83 (86%) were acceptable and could be applied to other situations. Furthermore, the HXP methodology appeared to help subjects capture mundane detail about location, movement, and intention. Supporting the idea that the HXP data is understandable by non-experts, subjects could easily understand the annotations of others and the two judges found it enjoyable to identify the problems in the other subjects' rules (as well as in their own rules). Within the constraints of this limited test, the results are encouraging and indicate that some non-experts, with training, are able to use the HXP methodology to create detailed narratives explained with general rules of common sense. Although a larger test in the future would produce interesting data about different types of users' capabilities, this small-scale experiment is adequate for demonstrating the methodology's potential to collect data from non-experts.

Chapter 6 covers future work, discussing several improvements to HXP and follow-up projects. In the short-term, the HXP user interface would be improved by having more explicit prompts and offering a way for users to add new predicates to the system. Later, the HXP controlled natural language should be expanded to provide more expressive annotations via adverbs, simple dialog, social relationships, comparisons, representations of abstract concepts, time durations, and repetitive actions. The inference capability should also be expanded to cover backward chaining, and multi-step inference. Finally, to take the project from proof-of-concept to fully functioning wiki collaboration, the software should be implemented as a thin client (i.e., the program should run from an Internet browser) with a host of features to make it more like a social networking site. To make the project more enjoyable for users, it could be integrated with a 3D model. Over the long term, the HXP methodology could be integrated with other NLP projects to make use of experiential data. Going beyond NLP, the HXP methodology could be an exciting tool of inquiry for event semantics and cognitive psychology.

HXP addresses the open research question of how to program common sense into computers. I seek the middle ground between formal commonsense rules built by experts on the one hand and loosely structured commonsense relationships collected from untrained volunteers on the other. I have developed a methodology and proof-of-concept software system that enables non-experts, with some training, to create simple, yet highly structured experiential narratives. The methodology shows promise in solving the inherent difficulties of having non-experts create data that traditionally has been the product of expert analysis. The resulting narratives and rules would be a valuable resource for both statistical and deep semantic analysis.

CHAPTER 1: OVERVIEW OF HXP USER INTERACTIONS

This chapter presents an overview of how a user creates an experience in HXP. It begins with a description of *key frames,* sequential slices in time. Then it describes the annotation process. This process starts by having the user add a background, character, and props. Then the user adds statements by following a series of drop-down menu selections. After adding each statement, the user explains the statement with a generalized commonsense rule.

1.1 Creating an Experience

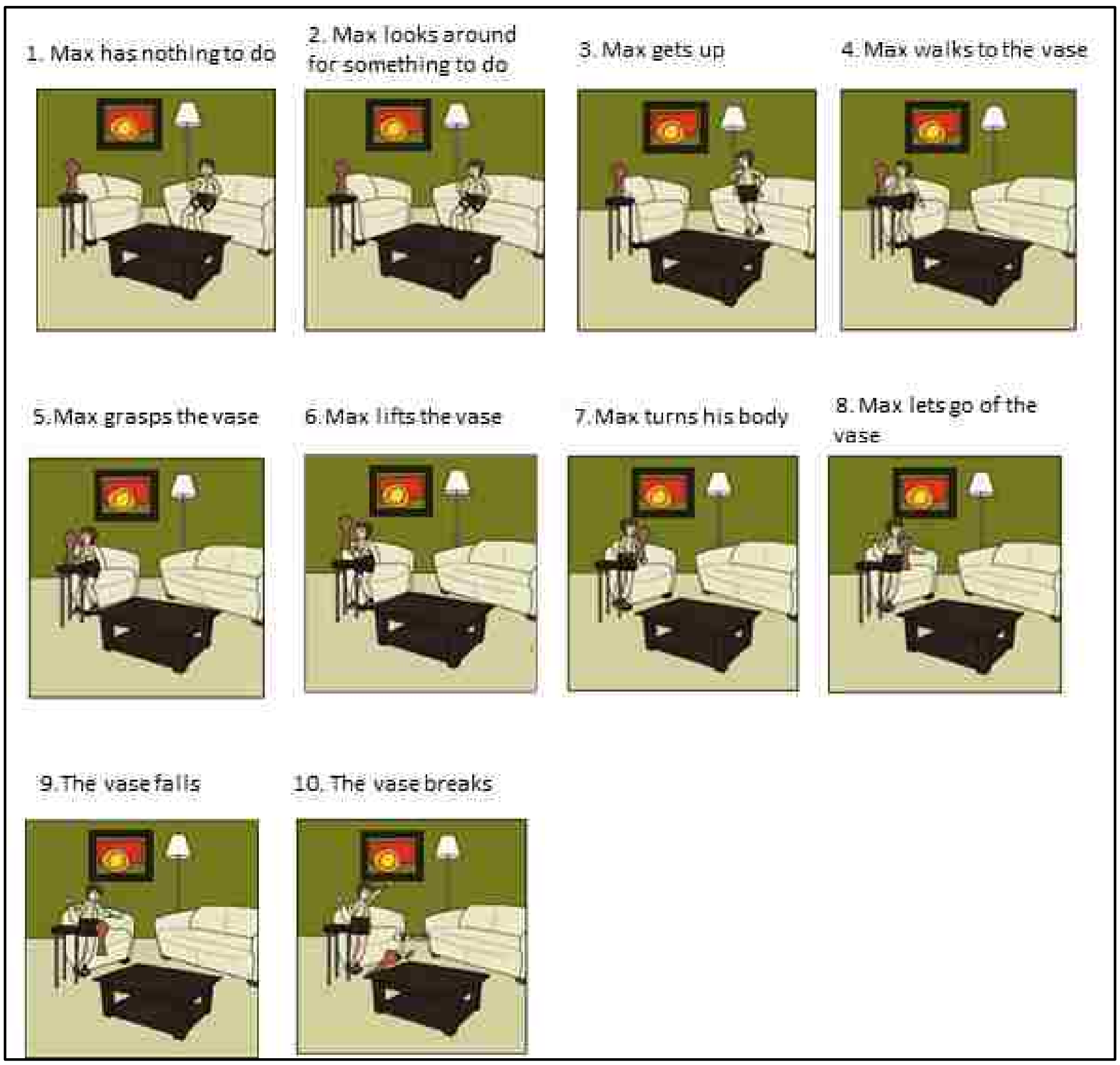


Figure 1: The “Max Breaks the Vase” experience

To create an HXP experience, a user first creates a series of comic strip frames, where each frame represents some small slice in time (e.g. one second or less). Figure 1 shows the frames of an experience, entitled “Max Breaks the Vase.” The frames depict a little boy named Max sitting in the living room with nothing to do. He notices a vase, picks it up, and drops it on the floor with a crash.

Each frame depicts an outwardly visible action. They are called *key frames* in HXP to distinguish them from other AI and linguistic uses of *frame*. (Minsky 1974, Fillmore 2003). Taken from the field of animation, key frames are images that define the movement and provide the anchor points between smooth transitions. Narratives using key frames offer an intuitive visual reminder to keep the actions small. The format of the images is not important at this point in the project; they can be drawings (either 2-D or 3-D), a series of photographs, or even a series of stills from a video. The images are not currently tagged; they have no inherent meaning. When shown one after the other, the images should give the illusion of animation. That is, if the images between key frames do not flow smoothly, the user knows more key frames need to be added, which will add more narrative detail. For example, an image of a boy on a sofa, followed by one of a boy next to a table, would be too abrupt; it would leave out the commonsense knowledge about how the boy stands up and walks to the table. Each frame also has an informal phrase that describes it, shown above each image in Figure 1. This phrase is not currently parsed, but it helps users create a structured caption, as described later.

For the experience showing in Figure 1, I used Pixton, a free comic editing/sharing web site with reportedly hundreds of thousands of participants.[4] The Pixton web site allows users to create comic frames and download them as JPEG files. Creating comic narratives suitable for

[4] http://pixton.com

HXP is a task manageable by non-experts, as evidenced by the fact that many hobbyists create comics for fun. Furthermore, I asked two undergraduates to use Pixton to create some simple experiences along the style of Figure 1. After a few minutes of training, they were able to create a ten-frame narrative in about an hour.

In this dissertation, all comic frames are assumed to be created in Pixton, or some other system, and manually imported into HXP. The focus of this dissertation is on annotating each frame, the most challenging aspect of the data collection.

1.2 The Annotation Process

The key frame images by themselves, like an animation or movie, are not useful for advancing NLP. It is quite difficult to extract situational data from unstructured data; otherwise, we would simply mine existing videos and text to model the human experience.[5] Therefore, HXP guides the user to convert the raw key frames into structured data. That is, the user first identifies the background settings, characters and props. Then the user creates statements to describe the action in each key frame and the effects of those actions. Finally, the user provides commonsense rules that explain each statement.

1.2.1 Background, Characters, and Props

Similar to comics or movie editing software, the HXP software interface provides a set of stock background settings, characters, and props from which a contributor can populate a comic frame. Example backgrounds are a living room, kitchen, classroom, or park. Example props are a ball or vase, and example characters are a boy, a woman, and a dog. There are currently a handful of stock backgrounds, characters, and props. The stock choices are pre-configured, but

[5] Pangburn (2002), however, shows the possibility of acquiring objects and actions from constrained video sequences.

they can be edited by advanced users. (Capabilities for advanced users are currently not implemented.)

In our example experience, a user would select *living room* as the background setting to match the background in Figure 1. This setting comes pre-configured with many objects such as a floor, ceiling, four walls, and some furniture such as a coffee table. Similarly, the user adds a vase to the frame by selecting from a list of stock props and then adds a boy from the stock characters, naming the boy *Max*.

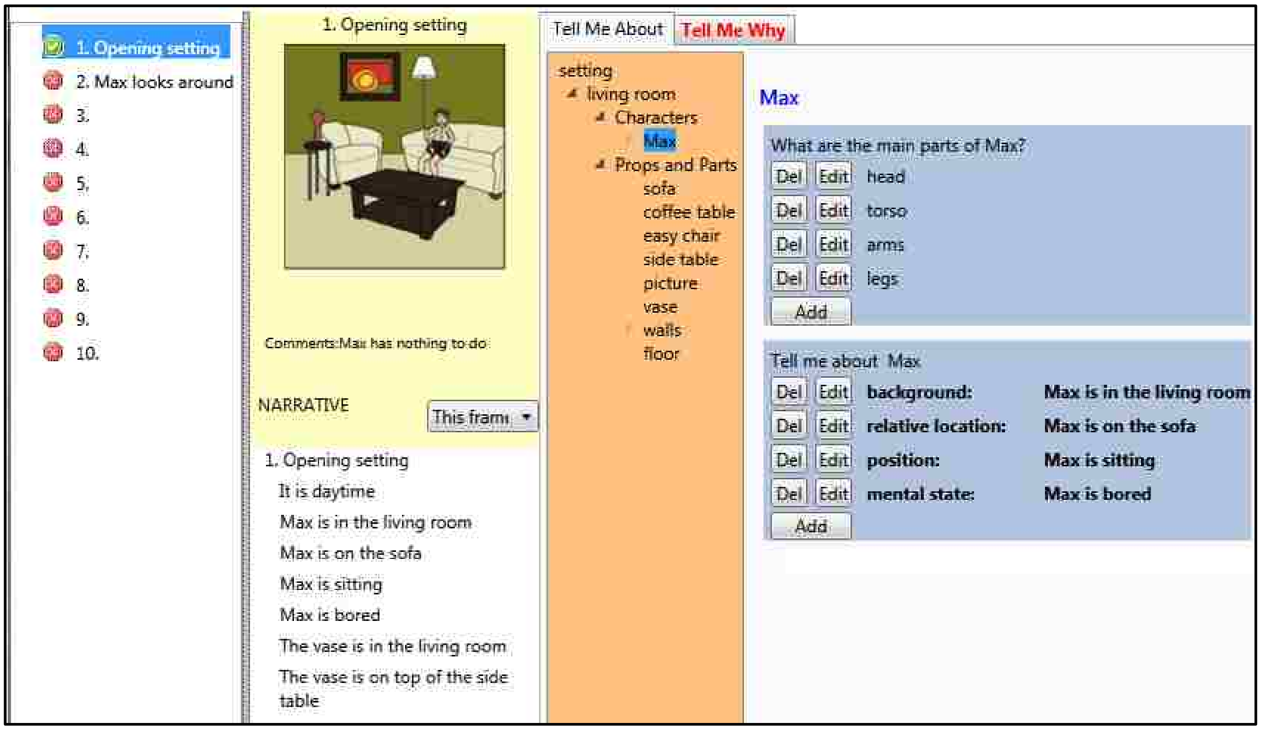


Figure 2: The screen for annotating key frame 1 "Opening setting." The right side of the screen shows the annotations for Max.

1.2.2 Statements

At this point, the user is prompted to input key commonsense information about the background, characters and props. The user chooses a time of day, identifies Max's room location, his body pose, and his mental state. Then the user identifies the room location of the vase. Figure 2 shows the screen after the user has selected daytime for the time of day, on the sofa for the location of Max, sitting for the body position, bored for his mental state, and on top of the side table for the room location of the vase.

When the user adds a statement like “It is daytime” or “Max is on the sofa,” HXP does not parse free text, but rather structures the user input by using drop-downs and selections from controlled vocabularies. It provides feedback in natural language to show the meaning of the user’s choices. This type of interface is called WYSIWYM for “What You See Is What You Mean” (Power et al., 1998). It controls the input so that all statements are unambiguous. All words in a statement are linked either to the WordNet standard ontology (Fellbaum, 1998) or to the HXP database (for words and concepts not found in WordNet).

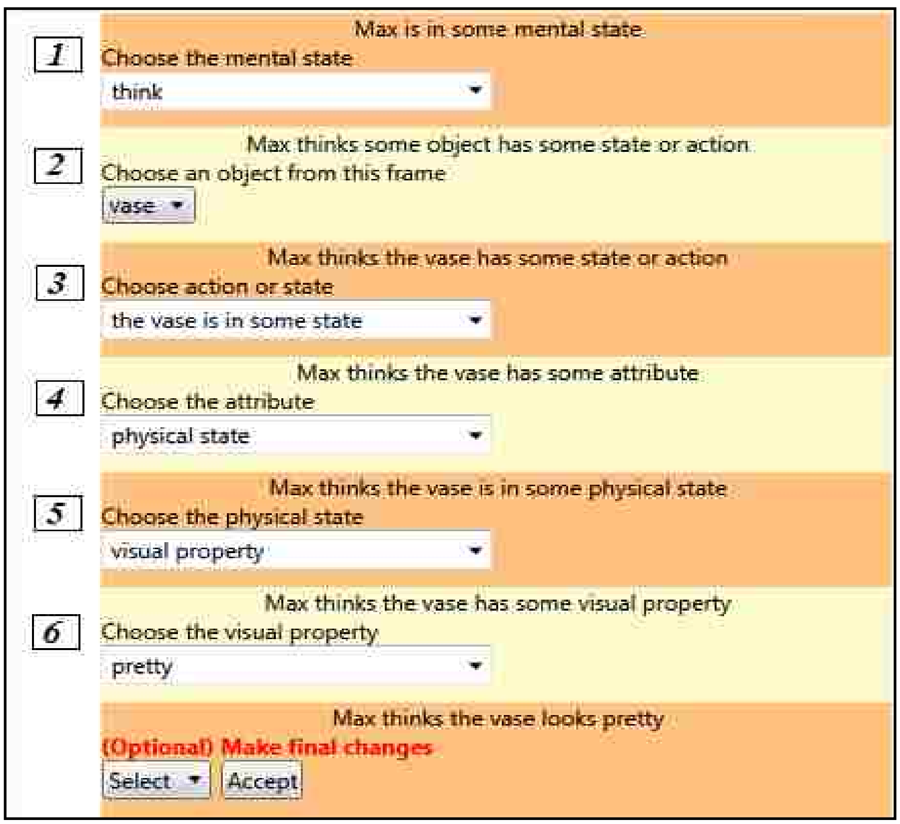


Figure 3: The user interface presents a series of drop-down selections to help the user enter the statement "Max thinks the vase looks pretty."

As an example of how this input works for a fairly complex sentence, Figure 3 shows a screen to input "Max thinks the vase is pretty." Each of the six inputs is a drop-down choice based on the current state of the input.

In step 1, the user chooses from a list of mental states, including emotions such as *angry* and *glad* as well as complex states such as *belief* and *desire*. There are about 60 mental states currently in the system. Once the user has chosen *think*, the system displays "Max thinks some object has some state or action" and prompts the user to choose an object from this key frame**.**

In step 2, the user chooses among all the objects that are in the key frame, including all the props, parts of props, and characters. In this example, the user chooses *vase*, and the system displays "Max thinks the vase has some state or action."

In step 3 the user is prompted to specify whether the vase is in some state or is doing some action. The user chooses *the vase is in some state* and continues on with the rest of the steps to drill down to *pretty*. Note, as a short cut, the user can simply type in *pretty* at any of the steps, starting at 3.

The underlying data structure representing each statement is a clause, containing a subject, predicate, optional arguments, and an optional subclause. In this example,

- Clause (subject="Max", predicate="think")
- Subclause (subject="vase", predicate="visual attribute", argument="pretty")[6]

Different predicates require different input screens and argument structures. HXP has about a dozen general-purpose templates that control the structure of the predicate, and each predicate maps to a template. For example, there is a template for enumerated types like colors and shapes, and another template for relative location predicates like *next to*. In this example,

---

[6] HXP considers single-argument predicate adjectives such as *pretty* to be semantic field values, as discussed in Chapter 4.

think is mapped to a template that requires a subclause. Of course, there are many synonyms for *think*, such as *believe*, and *consider*. Users choose the most appropriate synonym, and different synonyms could map to different templates. The templates and statement structure are described in more detail in Chapter 4.

1.2.3 Commonsense Rules

We have seen how a user creates statements that describe the objects in a frame. Now we will see how a user creates generalized commonsense rules that explain each statement. Going back to Figure 2, we see that the user has created seven statements, starting at the top of the Narrative section with "It is daytime." The Tell Me Why tab at the top of the figure is red, indicating that the user has not explained the reason behind these statements. Figure 4 shows the corresponding Tell Me Why screen.

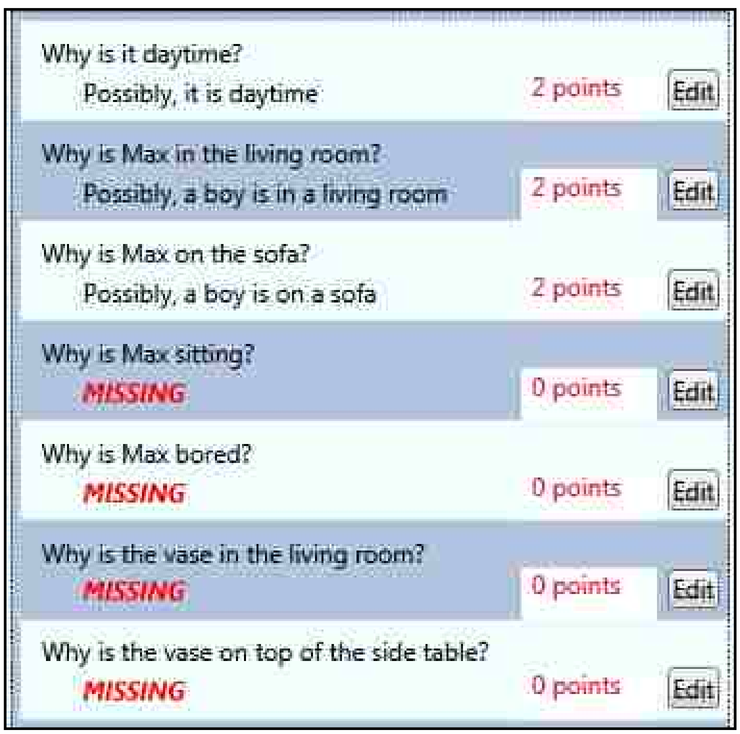


Figure 4: The Tell Me Why screen asks the user to explain each statement. The first three statements have been answered as simply "one of many possibilities." The other statements are as yet unexplained.

As with any "Tell me why…" question, sometimes the answer is simply, "Just because I said so!" That is, the reason is too difficult to explain. In this example, there really is no good reason as to why daytime was chosen as the time of day. Therefore, the user chooses *This is just one of many possibilities* – the polite equivalent of "Just because!" Users always have the options of answering in this way, and this is perfectly fine, especially in the opening scene where the characters and setting are just being introduced.

However, even in the opening scene it is possible to provide a more informative answer to some questions. Let us look at the fourth question, "Why is Max sitting?" This statement can be explained in terms of the previous statement "Max is on the sofa." The relationship between being on a sofa and being in a sitting position is an unstated, but understood, rule of common sense. The HXP user interface guides the user through a series of screens to create this rule.

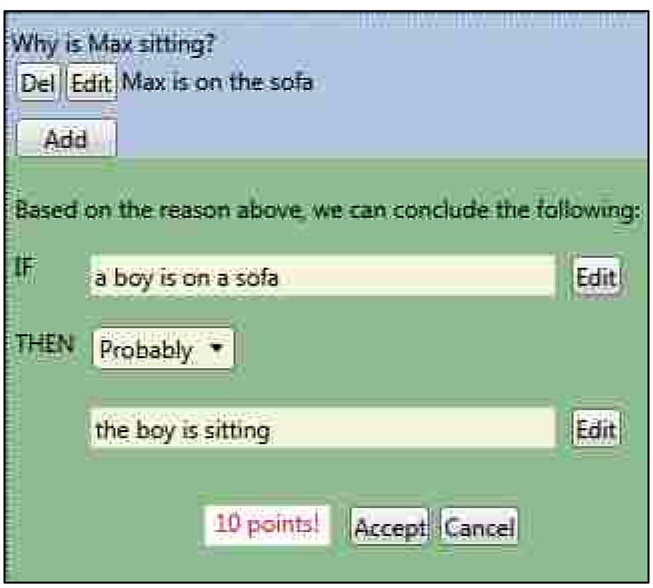


Figure 5: The user explains that we know that Max is sitting because Max is on the sofa. This explanation is then displayed as a general rule of common sense.

First the user selects which of the previous statements allow us to assume that Max is sitting. In this case, the user chooses the statement "Max is on the sofa." The system now restates this explanation as a general If-Then rule, as shown in Figure 5. The user has the option of editing the rule to make it more general. In this case, the user edits “a boy is on a sofa” and generalizes by choosing from the hypernyms for each noun. *Boy* has many hypernyms, including *male*, *child*, *person*, *living thing*, and *object*.[7] Likewise, *sofa* has many hypernyms, including *seat*, *piece of furniture*, *furnishing*, *man-made object*, and *object*. The hypernyms are taken from WordNet, with a few modifications.[8] In this example, the user’s best generalization would be from *boy* to *person* and from *sofa* to *seat*.

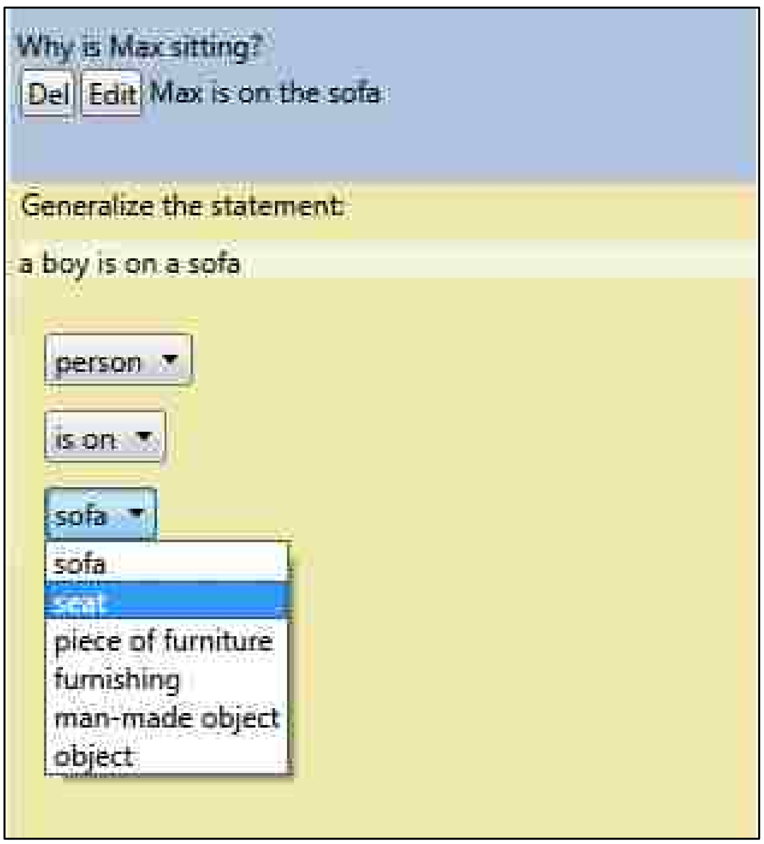


Figure 6: The user chooses hypernyms to generalize *a boy is on a sofa*. The user has already generalized *boy* to *person* and is now generalizing *sofa* to *seat.*

[7] The hypernyms are shown as a sequence, but they would more appropriately be displayed as a tree or lattice. That is, *boy→male→person* is one branch. The other branch as *boy→child→person*. The current HXP user interface shows a single sequence in order not to overwhelm users with detail.
[8] Modifications include changing WordNet’s *artifact* to be *man-made object* to make it accessible to non-technical audiences. Similarly, I removed technical WordNet terms like *physical entity* and *entity* from the hypernym tree. Finally, when necessary, I added hypernyms. For example, I added *child* to the hypernym tree of *boy*.

Figure 6 shows the input screen for generalizing the statement. After generalization, the rule would be displayed as:

1) IF a person is on a seat
THEN probably the person is sitting

To recap, a user first answered the question "Why is Max sitting?" by choosing a previous statement "Max is on the sofa." HXP then restated this explanation in terms of the general If-Then rule, shown in Figure 5. Restating the explanation in this form exposes gaps in commonsense information and encourages users to add more detail. Once satisfied that the rule seems to answer the question, the user proceeded to generalize it further, arriving at rule (1) above. At this point, not only has the user explained how we know that Max is sitting in this specific situation, but also the user has generated a useful rule for NLP, qualified by *probably*. In other words, given a situation where a person is on a seat, an NLP application could infer that the person is sitting, and the qualification of *probably* could be used to prioritize this rule over other possible rules. The example illustrates the gap between what is stated and what is understood. In typical narratives, most people would assume that someone is in a state of sitting if the narrative simply says that the person is on a sofa and no other postures are specified.

It is instructive to look at another example to show the efficacy of this method. In Frame 2 of Figure 1, Max looks around for something to do. During the training portion of the user test, when asked why Max looks around, every one of the test subjects easily answered, "Max is bored," which is Max's mental state from Frame 1. This answer leads to rule (2), which seems correct.[9]

[9] This rule assumes the boy is able to see. Chapter 3 discusses the issue of deciding what assumptions to make when formulating rules.

2) IF a boy is bored
   THEN probably the boy looks around

Moving on to Frame 3 of Figure 1, the question is much more difficult: Why does Max stand up? At first, each test subject tried to answer again that Max is bored, generating rule (3).[10]

3) *IF a boy is bored
   THEN probably the boy stands up

The subjects were generally unhappy about this rule because the mere fact that a boy is bored does not generally lead to the boy standing up. In fact, we know that Max is standing up to go to the vase. To capture this intention, one subject added the statement to Frame 2: "Max is curious about the vase" and gave this new statement as the reason for why Max gets up, generating rule (4) below.

4) *IF a boy is curious about a vase
   THEN probably the boy stands up

When the subject saw rule (4), the subject realized that curiosity was not enough to explain why Max stands up. Other subjects added statements about Max's intention, such as "Max desires to play with the vase" or "Max desires to examine the vase." But even these statements were judged to be insufficient as soon as they were presented as, shown in rule (5).

5) *IF a boy desires to examine a vase
   THEN probably the boy stands up

In order for a rule to make sense, the subjects had to add a statement to the narrative about Max's relative location. He is not near the vase. With this newly uncovered detail, along with the already existing statement about Max's sitting position, users created rule (6).[11]

6) IF a boy is sitting
   AND the boy desires to examine a vase
   AND the boy is not near the vase
   THEN probably the boy stands up

[10] The asterisk (*) in rule (4) indicates the subjects found the rule unacceptable and subsequently changed it.
[11] One test subject added a separate mental state "Max desires to walk to the vase" with a similar explanation.

When the subjects viewed rule (6), it seemed right; it seemed to reasonably explain why Max stands up before walking to the vase. Once the subjects felt that the rule reasonably explains why Max stands up, they generalized it to rule (7).

7) IF a person is sitting
AND the person desires to examine an object
AND the person is not near the object
THEN probably the person stands up

Generalized rules like (7) allow specific situations to be compared to a class of situations – very useful for commonsense modeling. But the process of creating the rules itself is also useful for commonsense modeling because it leads users to add more details. In this case, it prompted the detail about Max's desire to examine the vase and the detail about Max's location. Of course, these additional details spawn even more statements. Why does Max desire to examine the vase? It is because Max sees the vase and perhaps Max thinks the vase is pretty. It is important to note, however, that the user can stop the Tell Me Why cycle at any time by choosing *This is just one of many possibilities*.

The confidence levels of *possibly*, *probably*, and *definitely* are deliberately non-precise. They represent the user's assessment of what most people would assume if they were to fill in the gap between what is stated and what is understood in the context of a typical narrative.

To motivate users to add detail and general rules, we established a simple point system on each explanation. If a user takes the easy way out with the "Just because" answer, they receive the minimum points. But if they can explain a statement in terms of a previous statement, they receive more points. And if they can generalize the statement from, say, *boy* to *person*, they get even more points. Finally, the most points are obtained by increasing the confidence level from *possibly* to *probably* or to the highest level of *definitely*.

1.3 Summary

This chapter has presented an overview of the HXP methodology, showing the main elements of an experiential narrative and how the HXP interface guides the user to create structured and detailed data about simple experiences. Narratives are organized around a sequence of key frames, where each key frame represents a small action and the effects of that action. Users add background settings, characters, and props, and then describe these elements as the narrative progresses with each key frame. Narrative statements are simple predicate-argument structures, which the user creates from a series of drop-down menu selections, so that the resulting statements are unambiguous, with vocabulary linked to a standard ontology. The HXP interface guides the users to explain each narrative statement with a general commonsense rule. Users build these rules inductively, by first selecting previous specific statements and then selecting hypernyms of the objects or predicates to make them apply to more general situations.

Quite often there are changes in mental states, body positions, or object locations that would naturally go unstated because they are so obvious to a human. However, when HXP asks users to explain each statement, they notice missing details and add them to the narrative. Thus, the HXP methodology, which prompts users to explain each narrative statement with a general commonsense rule, serves two purposes. It not only creates data that abstracts from specific situations to more general ones, but it also results in a more detailed and coherent narrative.

CHAPTER 2: RELATED WORK

Having presented an overview of HXP, we now can compare this project to others concerned with story understanding and NLP. There have been several projects engaged in deep semantics and story understanding (Charniak,1972; Lehnert,1978; Dyer,1983; Lenat et al., 1990; Mueller,1998). However, the problem has remained intractable, and attempts to create rules based on simplified, artificial worlds or stereotypical situations have proven to not scale up to more realistic situations. By the 1990s, the creation of large corpora of tagged sentences such as PropBank (Palmer et al., 2005) and the Penn Treebank (Marcus et al., 1993), in combination with lexical resources, especially WordNet (Fellbaum, 1998), provided more promising results through statistical methods. As a result, most NLP research has shifted away from deep semantics and moved towards statistically oriented approaches, and we have seen impressive advances in parsing, data extraction, machine translation, and question answering, as demonstrated recently when IBM's Watson program beat human champions at Jeopardy (Baker, 2011). Yet Watson draws upon many resources, including the logical rules of Cyc's OpenCyc database, PropBank, as well as a variety of knowledge banks formed from simple word associations. The successes of Watson's statistical approach suggest that NLP can benefit from many types of knowledge sources, and I would like to add HXP's structured experiential narratives to this mix.

Comparing HXP to previous work, this chapter first discusses projects that collect logical rules (Cyc, AceWiki, various case-based reasoning systems). Among these projects, HXP is most similar to the case-based reasoning systems because HXP statements and rules could be viewed as cases for experiential reasoning. Next, the chapter briefly touches on projects that collect lexically-oriented data (Penn Treebank, PropBank, WordNet, VerbNet, FrameNet) and projects that deal with discourse and rhetorical structure (Penn Discourse Treebank, TimeBank). This will bring us to the projects most closely related to HXP because they collect narratives (Scheherezade, OMCS narrative projects). Being able to collect chunks of commonsense knowledge in narrative form has long been a goal of NLP, and HXP offers a new approach. After looking at the data collection projects, we will see that projects that model story understanding and commonsense reasoning use data types that are very similar to HXP's, except that HXP allows non-experts to create the data. Finally, the chapter ends with a discussion of how HXP synergizes with projects that deal with visual representations of commonsense data (EBLA, WordsEye).

2.1 Logical Rules

2.1.1 Cyc

Probably the most famous attempt to build a program for deep semantic processing is the Cyc project. Started in 1984, its original ten-year goal was to "span human consensus reality knowledge: the facts and concepts that you and I know and which we each assume the other knows" (Lenat et al.,1990). Over two decades, the project employed a small team of AI researchers to carefully craft millions of logical assertions, predicates, and concepts, as well as an integrated logic system for understanding natural language. Unfortunately, this grand experiment has not accomplished its original ambition. In the late 2000s, Cyc's focus shifted to more reachable goals such as natural language queries into specific data domains like medicine or counterterrorism.

Cyc was designed to work with abstract, logical rules, independent of any particular situation. It partitions knowledge into microtheories, where each microtheory is locally consistent. It takes a great deal of expertise to create these abstract rules and keep them consistent. In contrast, HXP is to be a collection of specific narratives and rules built by non-experts. Unlike Cyc rules, the HXP rules are not intended to be consistent with one another, and they are defeasible. The HXP knowledge is informal and inconsistent, but still relates details of actions, agents, objects, and intentions in a structure suitable for NLP.

2.1.2 AceWiki and Attempto Controlled English (ACE)

The AceWiki (Kuhn, 2008) enables users to collaborate online to build ontologies and make logical assertions. Unlike the first-order logic language of Cyc, AceWiki uses a subset of English, making it easier for non-experts to understand. This language, Attempto Controlled English (ACE), provides user-friendly translation between natural language and formal logic (Fuchs & Schwitter, 1996). Version 6.6 of ACE (Fuchs et al., 2010) defines many formalisms for describing experiences, including verb constructions (active, indicative, third person), noun constructions (count, mass, plural), adverbs, and generalized quantifiers. It also represents modality and intentionality with subordinate clauses. However, as far as I am aware, ACE 6.6 is not designed for reasoning about narratives, and it does not provide a methodology for eliciting commonsense details from users. Furthermore, there are a few constructions that ACE does not represent, such as "Max is curious about the weight of the vase" and "Max is curious about why the vase falls." Nevertheless, AceWiki is an excellent example of a website that allows users to create structured knowledge using a natural language interface, and ACE is a powerful and user-friendly controlled natural language that has many constructions that HXP needs in the future. HXP would do well to incorporate many of its formalisms.

2.1.3 Case-Based Reasoning Systems

Case-based reasoning (CBR) is a type of machine learning based on the idea that new problems can be solved by re-using solutions to previous problems. As such, CBR systems are a good model for collecting specific experiences and generalizing them to new situations. In CBR terminology, a previous problem and its solution, in the form of a logical rule, are referred to as a *case*. Thus HXP specific situations and explanations in the form of general rules could be viewed as CBR cases, cases to reason about similar experiences. A critical component in any CBR system is how to compare cases to find the closest match. HXP's hypernym-based partial order, discussed in Chapter 3, provides this capability.

The idea of collecting If-Then rules from non-experts through an interactive user interface is somewhat related to Protos, a CBR system for classifying audiology cases (Porter et al., 1990). Protos holds training sessions with a human expert to learn how to classify new cases. During these sessions, it first attempts to classify a case using its current set of rules and relationships. Then it questions the human expert to get more information. Protos interactively guides the human towards a better audiology classification, and in the process, extracts useful information for reasoning about new cases. Protos displays to the users information in the form of If-Then rules. However, Protos is specifically geared towards the audiology domain, whereas HXP is a more general purpose system. Furthermore, Protos does not generalize the explanation to go much beyond the current audiology case.

Another system that acquires rules through interactive dialog is the Disciple system (Tecuci & Hieb, 1996), with many domain-specific implementations (Tecuci et al., 2005; Tecuci et al., 2008). Disciple is a tool for rapid development of learning agents for expert systems. A subject matter expert (SME), with help from a knowledge engineer, enumerates different cases of how to solve a specific problem, including alternative scenarios. In the process, the SME creates the ontology for the problem by defining the important concepts and relationships among those concepts. With this base of information, Disciple analyzes the data and automatically creates If-Then rules from example problems and solutions. It then refines the rules by interactively prompting the user to specify which other examples apply to the given rule. Interestingly, Disciple makes rules more general or specific via the *hypernymy* relationships of the SME's ontology. HXP inherits this idea of generalizing If-Then rules from hypernyms. However, since HXP models the problem domain of simple experience, it can re-use the hypernym relationships of WordNet. Unlike HXP, Disciple automatically generates the initial If-Then rules from the user's examples. Disciple's automatic rule generation is appropriate when the user is willing to do an initial deep analysis of a problem domain to identify important alternative paths and example solutions. But in HXP, it would be asking too much for non-experts to do this analysis. However, HXP could borrow Disciple's approach in the future to refine existing rules after more experiences have been collected.

2.2 Lexically-Oriented Data

Having examined related rule-based NLP and AI projects, we now turn to popular lexically-oriented data used by statistical NLP. Statistical NLP became feasible with the creation of tagged corpora, large collections of texts whose words are manually labeled with part of speech and other information. The Penn Treebank (Marcus et al., 1993), with tagged Wall Street Journal articles, is a popular resource for statistical parsers. PropBank (Palmer et al., 2005) tags verbs and their argument structure.

Besides part-of-speech tagging, a variety of NLP resources capture lexical relationships and verb argument structure. Many of these knowledge bases contain binary relations among words and phrases, such as the following:

- sofa is a subclass of furniture
- milk is used for drinking
- scientists publish papers

Others contain more complex relations, describing predicates and their argument structure. For example, the verb *eat* may be described as having an *agent* that performs the action and *patient* that is affected by the action. The patient is further described as *comestible* and *solid*.

The resources involving lexical structure can also be categorized according to how they are built: through manual labor or through automatic information extraction. In the manual labor class, crafted by experts, are WordNet (Fellbaum, 1998), VerbNet (Kipper et al., 2006), and FrameNet (Fillmore et al., 2003). WordNet is probably the most popular NLP resource in the world. Analogous to a machine-readable thesaurus, it contains hundreds of thousands of words with their lexical relationships. Words that are synonyms (e.g. *arise*, *get up*, *stand up*), are part of the same *synset*, and these are linked by lexical relationships such as hyponyms (e.g. dog and animal), and meronyms (e.g. wheel and car). VerbNet, as the name indicates, deals with verbs and their relations. For each verb, it defines general argument structure and thematic roles such as *agent* and *patient*. At the semantic frame level, FrameNet deal with standard situations and roles: Stealing a car fits into the *Theft* frame, which is a type of *Committing Crime* frame with core elements *Crime* and *Perpetrator*.

A slightly different class of manually built resource is ConceptNet (Liu & Singh, 2004). While it is constructed by manual labor, ConceptNet is not built by AI experts or highly trained graduate students, but rather it is a product of thousands of casual users online playing various simple games as part of the Open Mind Common Sense (OMCS) project at MIT (Singh et al. 2002). A pioneer of crowdsourcing methodology several years before the word *crowdsourcing* was coined, OMCS has shown that untrained participants can provide large quantities of commonsense data about words and phrases. More recently, the Games with a Purpose project (Von Ahn et al., 2006) has become a source for ConceptNet.

In the automatically extracted category are TextRunner, DIRT, Never-Ending Language Learner, Background Knowledge Base, and PRISMATIC, all described in (Fan et al., 2011). In general, these resources are created from programs that extract semantic information from millions of web pages. The programs parse texts, searching for repeated syntactic patterns from which to infer semantic relations. Their results are enhanced by leveraging the knowledge in the manually created resources.

In brief, resources of word relations are the backbone of statistical NLP. The manually built resources came first and allowed more accurate parsing, identification of verb arguments, named entity recognition, and other sub tasks, which in turn enabled advances in automatic information extraction. Like other manually built resources, HXP narratives will be expensive and time-consuming to produce. But I believe creating data of this sort is a necessary first step for identifying higher-level patterns that can later be applied to automated methods.

2.3 Discourse/Rhetorical Structure

Some tagged corpora go beyond syntactic tagging within a sentence. They label *discourse relationships*, how one sentence relates to another, such as causation, conjunction, and event ordering. The Penn Discourse Treebank (Prasad et al., 2008) labels intersentential relations and their rhetorical purpose, but as noted by (Elson, 2012), it is geared towards expository prose rather than narrative. The TIMEBANK Corpus (Pustejovsky et al., 2003) codes events and temporal relationships. However, there are no resources for narrative data (except for Elson's proposal based on Scheherezade, discussed below). In particular, there are no current projects geared towards collecting commonsense details in narrative form.

Would it be possible to construct discourse relationships, including narrative data, automatically? Chambers & Jurafsky (2011) automatically learn some semantic roles from raw text based on related verb patterns. In their automated learning system, verbs are considered related if they occur near each other more often than expected by chance. Also, verbs are part of a related event sequence if they have coreferring arguments. This work builds from their previous work learning narrative event chains from raw text (Chambers & Jurafsky, 2009). This type of work shows the potential for building NLP resources automatically containing valuable narrative data. The resulting data could construct templates about events such as kidnapping or elections and induce the order of events. These systems depend on manually built resources, in particular WordNet, treebanks, and the TIMEBANK Corpus. However, Li et al. (2012) describe the limitations to automatically extracting narrative-like knowledge from unlabeled news corpora and other online texts. In particular, there is the difficulty of unsupervised machine learning without a priori knowledge of the topic. There would have to be a large number of narratives on the same topic for a program to determine which events are relevant. I believe that a resource of detailed experiential data such as HXP that relates events and intentions would be make it easier for automated systems to recognize relevant information.

2.4 Narrative Data

Less constrained than formal logical rules, yet more extensive than lexical associations, narratives provide “contextualized knowledge” – chunks of coherent relationships among states and events (Singh & Barry, 2003). There have been several attempts to use narrative structures for commonsense reasoning. Schank, Minsky, and colleagues introduced the concept of scripts and frames in the 1970s (Schank & Abelson, 1977; Minsky,1974). Scripts are abstractions of event sequences based on many concrete experiences. For instance, a script for eating at a restaurant has roles such as the cook, waiter, and customer, and activities such as ordering food, eating, and paying the bill (Mueller, 2006). Mueller (1998) created a database of abstract scripts specifically for story understanding but noted the difficulty and tedium involved in this work. Creating a master script to generalize an activity is extremely difficult. Indeed, it requires a great deal of expertise to find the quintessential sequence of events and roles that make up a generalized *type* of activity such as going to a restaurant.

On the other hand, almost anyone can describe what happens in a particular experience at a particular time and place. As Singh and colleagues point out in their motivation for OMCS's StoryNet, "it may be easier to tell and explain a specific story, which focuses the user on a specific set of characters, objects, and events, and their relationships, then to ask them to make a general rule-based theory in the abstract of some domain"(2004).

This brings us to projects closely related to HXP that collect narrative data. First we present Scheherazade, a software system for annotating stories. Then we present several projects related to the OMCS that collect narrative data from untrained people via the Internet.

2.4.1 Scheherazade

Scheherazade (Elson & McKeown, 2010) seeks to create a bank of annotated stories to advance text understanding and narratology. HXP shares many attributes with the Scheherazade system (henceforth SCH). Both systems represent narrative statements as predicate-argument structures, represent intentions and goals with causal links, and specifically identify actors, locations, props, and narrative time slices (called *story points* in SCH). Finally, they both use a WYSIWYM user interface to guide the user input with minimum parsing. Nevertheless, SCH does not seek a reduced vocabulary as much as HXP. SCH allows users to paraphrase relatively complex statements and dialog from the story, whereas HXP seeks to represent experiential data as simply as possible. SCH focuses more on analyzing the structure of existing stories than on collecting commonsense data, while HXP seeks difficult-to-articulate commonsense data. These differences affect the visual representation of the narrative, the overall user interface, and the underlying data structures.

Although both systems visually represent a narrative in terms of time slices, the granularity of the time slices differ. The SCH user paraphrases a story in order to analyze the story's narrative structure. The user takes logically separate actions from the story and identifies them as separate story points. If the story says "there were three pigs living in a house, and a wolf came knocking on the door," there would logically be two story points: pigs living in a house and a wolf knocking on the door. The granularity of the story points may be large or small depending on the original author's rhetorical purpose. In this case, we don't know how the wolf gets to the door, and we really don't care. Perhaps the wolf walked from his own home, down some sort of woodland walk to the pigs' front yard, walked up the yard to the front porch, up the steps, and to the pigs' front door. This level of granularity is rather ridiculous for a fairy tale, but it is quite valuable for commonsense reasoning.

HXP asks users to provide this type of detail about a tiny scene from everyday life. To encourage this, HXP has a comic strip panel representation of time slices, specific prompts to annotate mental states, and a "Tell me why" interface to help users provide even more detail.

HXP has a two-tiered approach to the user interface. Novice users choose from stock backgrounds, characters, props, and vocabulary, and advanced users are able to create new stock data. The novice user has to choose from what is available, which is far less expressive than SCH but is simpler to use. I believe this trade-off in simplicity versus expressivity makes sense given that HXP is more concerned with commonsense details of simple situations than with translations of existing stories which may be arbitrarily complex. Finally, HXP users are asked to explain *every* statement in order to keep the narrative as coherent as possible and to elicit relevant commonsense detail whereas SCH seeks annotations that describe larger narrative goals.

In both systems, the underlying data structure is a set of predicate-argument structures linked by time, cause, and other relationships. Despite the similarities, SCH is designed to capture narratalogical structure, not mundane commonsense details. Also, the SCH interpretation layer links major plot points with complex causal graphs, but there is no structure for converting these into general commonsense rules.

As a final note, SCH is not merely a proof of concept; it is a mature, ready-to-use product. As HXP develops, some of the SCH formalisms for representing time, groups, and comparisons could be incorporated into HXP.

2.4.2 OMCS Narrative Projects

OMEX (Singh & Barry, 2003), StoryNet (Singh et al., 2004) and ComicKit (Williams et al., 2005) were all associated with MIT's OMCS project (Singh et al., 2002). All three attempted to collect stories specifically for commonsense reasoning. Similar in spirit to the OMCS project, but not part of the MIT initiative, (Li et al., 2012) uses crowdsourcing to collect narratives, and then parses the narratives to get a script of events for common activities. The next three subsections discuss these projects in detail.

OMEX and StoryNet

Singh & Barry (2003) created the Open Mind Experiences (OMEX) project to gather everyday experiences for commonsense reasoning. To structure the input, OMEX provides several very short story templates based on Lehnert's plot units (1981), which describe relations among stereotypic narrative situations. For example, the "Fleeting-Success" plot unit is about a situation that first seems to have a positive outcome but later has a negative outcome. Users select different plot units, and the OMEX system displays fill-in-the blank templates. For example, selecting "Fleeting-Success" displays the following propositions with empty blanks:

> Iris has a problem.
> Her _____ won't _______ .
> She _____ and it works perfectly until ________ .

The user fills in the blanks with free-format text. For example:

> Iris has a problem.
> Her pen won't write.
> She shakes it and it works perfectly until it explodes.

Using several pre-defined templates, OMEX prompts users with yes-no questions to explain the story. For example, OMEX might generate the following questions:

> Did the event shake the pen cause the event the pen explodes ?
> Can Iris see the pen?

According to Singh and colleagues, the OMEX data was too abstract to engage users and was not constrained enough for knowledge extraction. The lessons learned were incorporated into the follow-up StoryNet project (Singh et al., 2004). StoryNet makes use of existing OMCS data to offer an easy-to-use interface for creating structured stories. It uses a simple drag and drop interface to take statements from the OMCS database and put them together into a narrative. For example, from statements in the OMCS database involving airplanes, a StoryNet contributor could create this story:

> "I travel to an airport. I board a plane. I fly in an airplane. I put on safety equipment. I open a door. I see a cloud. I jump out of an airplane."

The StoryNet approach to structured input means users create very little free-format content. In the example above, the story statements are simply OMCS sentences arranged in ordered sequences. However, in exchange for this easy interface, the stories were not very engaging and therefore were probably less attractive for others to read. Singh et al. discussed ideas to add more interesting detail to the stories, but they were apparently not implemented.

In these projects, Singh and colleagues articulate the need for commonsense data in narrative form. HXP inherits their ideas to collect specific narratives and explanations. Like OMEX, HXP is envisioned as a wiki web site. However, HXP is not for casual users. The HXP annotation process requires training, and it is geared toward a smaller audience, hobbyists and people interested in advancing AI.

ComicKit

Another OMCS-associated experimental project was called ComicKit. It offered a comic strip interface for telling stories. Figure 7 shows an excerpt from a ComicKit. It is a story about Alice, who wakes up depressed and decides to go on a walk. The comic strip idea is a good way to engage users, as indicated by ComicKit's user test, where users reportedly reported a high degree of enjoyment. But ComicKit does not address one of the basic problems inherent with understanding comics: the stories require a lot of common sense to understand what happens in between the panels.[12] Although the comic strip format reportedly was fun for users, the lack of constraints on content and captions resulted in stories that were difficult for automated analysis. Unlike HXP, there was no mechanism for eliciting more details from users in order to make the narrative more detailed and coherent.

[12] The commonsense needed to interpret implied action in comics is called closure by McCloud (1994).

Figure 7: ComicKit story about Alice. The content is unstructured text.

HXP focuses on child experiences, which narrows the experiences to actions closely tied to simple activities. And actions are limited in time and space because contributors select detailed settings and show how objects change from one panel to the next. Finally, events in one panel are explicitly tied to those of the previous panel. In short, the HXP methodology helps contributors create highly structured scenes, yet there is a lot of freedom for specifying content.

Crowdsourced Event Scripts

Not related to the OMCS project, but similar in spirit, Li et al. (2012) collect everyday narratives in natural language from workers on Amazon's Mechanical Turk.[13] Rather than using drop-down menus for structured input, contributors write free narrative sentences. However, to make parsing easier, the contributors are provided a fixed set of character roles (e.g. customer, cashier) with proper names, and are instructed to use simple one-verb sentences with one activity per sentence. These constraints make it relatively easy to parse sentences and resolve noun

[13] mturk.com

references. However, the activities are not constrained to small actions and are not causally explained. Therefore, the narratives tend to leave out a great deal of detail.

Li et al.’s example narrative about a restaurant is a case in point. It begins with two activities: a) John drives to the restaurant and b) John stands in line. This level of granularity omits the detail of John's getting in the car, traveling along a road, parking the car, getting out, or going in the restaurant. It also leaves out commonsense information about why someone would go to a restaurant, use the car, stand in line, etc. In sum, Li's narrative data are valuable data about events that make up common activities and event ordering, but, unlike HXP, they do not attempt to represent the contextual details. As Li’s paper points out, the crowd workers habitually omit details they consider to be obvious.

The same problem was noted in another OMCS-related project called Open Mind Indoor Common Sense, which uses free-form text to describe common household activities such as sweeping the floor or bringing in the mail (Smith & Arnold, 2009). As with Li et al., contributors leave out obvious details when describing the activities. Also, different contributors use different levels of granularity to describe an activity, which makes it difficult to detect similarities among narratives. The HXP format that encourages smooth actions between frames helps solve this problem because all narratives are at about the same level of detail.

2.5 Story Understanding

Having considered projects closely related to HXP that collect narrative data, we now consider two projects relevant to HXP because they use narrative data to model story understanding. By the 1990s, AI researchers had largely abandoned deep semantics approaches to NLP in favor of statistical ones. However, Erik Mueller (2000) called for the AI community to once again tackle the very difficult problem of story understanding. He outlined new NLP tools and AI architectures and recommended a set of basic questions that understanding agents should be able to answer. McCarthy et al. (2002) answered the call, and they proposed a new research initiative in story understanding. They would focus on early reader texts designed for preschool and kindergarten in order to make the research more tractable. They would identify a small corpus of children's texts, manually annotate them, and then attempt to develop methods to understand the domains that occur most frequently.

As a first step, Mueller (2003) annotated one of these texts and modeled how a human would understand the story. The model has two main types of narrative data: story statements and commonsense axioms. With these data represented by formulas in event calculus, Mueller runs an efficient satisfiability solver that models the mental and physical states that occur during the narrative. Although his logical formulas were re-used and extended in later work (2004, 2007), the initiative towards children's stories seems to have run out of steam.

In a different approach to natural language processing, (Singh, 2005) envisioned a library of commonsense narratives to enable case-based reasoning (CBR). Singh's example narratives are highly detailed and manually created by Singh himself. Statements are predicate argument structures, and Singh explicitly represents different types of dependencies between statements such as *causes* and *implies*. Because the statements are so detailed, they contain a wealth of commonsense knowledge. For example, if the situation has a character that desires to hold a stick and then immediately grasps the stick, we could conclude that "If you want to be holding an object then you should try to grasp it" (p.38). Singh creates a CBR system where software agents called *critics* use hand-crafted rules to reason about an agent's situation. Singh believed it would be eventually possible to construct at least part of this knowledge from the general public.

Both Mueller and Singh use the two data types of data that HXP collects: specific narrative statements and generalized rules. However, their formalisms are expertly coded and cannot be generated from non-experts. Neither Mueller nor Singh explicitly mention defeasibility in their rules. HXP uses defeasible rules because they are easier for non-experts to construct from specific situations without considering all possible exception cases. Mueller's narrative statements are far less detailed, but his goal is to computationally add detail through a satisfiability solver and commonsense axioms. Singh's statements have much of the commonsense detail already, and his goal is to use the rules and statements as part of a CBR system to generalize to new situations. In this respect, the highly detailed statements collected by HXP are more similar to Singh's system. Furthermore, the HXP demonstration of forward chaining, discussed in Chapter 3, could be viewed as a CBR system that matches antecedents from specific cases and generalizes them to new situations. However, the HXP goal is not to build a CBR system, but rather to collect detailed experiential data suitable for both statistical and deep semantic analysis.

2.6 Experiential Data from Images

Images are an efficient way to represent commonsense data. For HXP, 2D images provide visual cues to help users detect missing actions, but they are not tagged or structured for analysis. The two projects in this subsection deal with visually-oriented representations of structured data.

2.6.1 EBLA

The goal of the Experience-Based Language Acquisition (EBLA) project is to acquire a childlike protolanguage based on the objects and actions detected in experiential videos (Pangburn, 2002). An EBLA experience is a short video of a simple action such as a hand picking up a ball. The project uses machine vision to detect objects and their relationships. Each EBLA video has a short natural language description like "hand picks up ball." The words of the description take on meaning as the system correlates the content and descriptions of many different videos.

EBLA's goals are to model language learning based on experience, whereas HXP seeks to collect experiential data. Therefore, there is a synergy between the two projects. The data collected by HXP could help EBLA and future machine vision projects by providing a context for the possible objects, intentions, and actions that might be detected. Conversely, advances in machine vision could improve the HXP annotation process, particularly in a future where users annotate actual video frames. If machine vision could detect objects and the starts of new actions, the system could automatically fill in some of the annotation information.

2.6.2 WordsEye and VigNet

WordsEye automatically converts natural language descriptions of objects and simple spatial relations into 3D scenes (Coyne & Sproat, 2001). WordsEye needs a lot of default information about how to place objects, and VigNet (Coyne et al., 2011) is a proposed resource to provide this type of data. VigNet extends the semantic information of FrameNet by adding visually-oriented information called *vignettes*. For example, the *At Counter* vignette represents a situation where two parties interact over a counter. The underlying vignette construction has a role for two actors, the counter, and items on or near it.

The purpose of VigNet is to provide data for text-to-scene generation, whereas HXP's purpose is to provide data for NLP. However, VigNet needs common sense in order to choose backgrounds for described objects and provide reasonable locations and body positions, so it share's HXP's goal to collect commonsense data. Rouhizadeh et al. (2011) describe a VigNet collection experiment where untrained workers were able to provide lists of objects typically found at a location and typical object parts. However, VigNet does not seem to be attempting to collect the type of detailed narrative data that HXP seeks about intentions and actions.

HXP could benefit from WordsEye and VigNet because users would enjoy constructing key frame data by using natural language descriptions. Conversely, VigNet would benefit from the details about actions and intentions provided through the HXP methodology.

2.7 Summary

This chapter discussed previous research related to collecting commonsense data for NLP. Several widely used manually-constructed resources are available that have enabled significant advances in statistical machine translation, question answering, and information extraction. However, most of the knowledge in these resources is based on tagged corpora and simple word associations. Larger chunks of knowledge, which link events and intentions, would both enable more complex reasoning and improve statistical processing. Singh (2005) envisioned a database of narratives that would serve this purpose. He believed that simple narratives would be easier to acquire from the general public than logical rules or abstract scripts. However, narratives collected from untrained people over the Internet, even when structured as simple event sequences, do not contain the sort of mundane details needed for commonsense reasoning. HXP addresses these issues with a methodology to enable non-experts, with some training, to create highly structured narratives.

CHAPTER 3: MODELING SIMPLE EXPERIENCES

While Chapter 1 presented an outline of how to collect the data, this chapter provides more detail about how these data types are modeled. First, it gives an overview of the experiential narrative structure. Then it presents the design of narrative statements. Finally, it describes the structure HXP commonsense rules and how they are used during inference.

3.1 Overview of the Narrative Structure

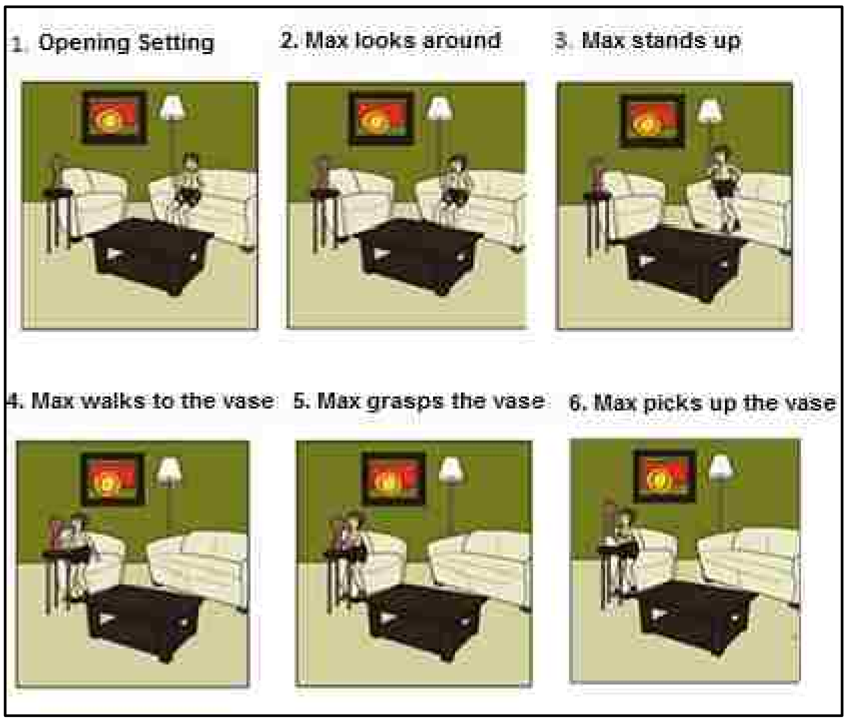


Figure 8: The first six frames of “Max breaks the vase” from Chapter 1

As discussed in Chapter 1, HXP displays an experiential narrative as a sequence of action pictures, like the still images of an animation flip book. Figure 8 shows a fragment of “Max breaks the vase,” taken from Figure 1 of Chapter 1 and repeated here for convenience. Each key frame represents a completed action over a small interval of time. For example, in the third key frame, Max has just stood up. In the fourth key frame, Max has completed walking to the vase. The user annotates each key frame with the effects of the action.[14]

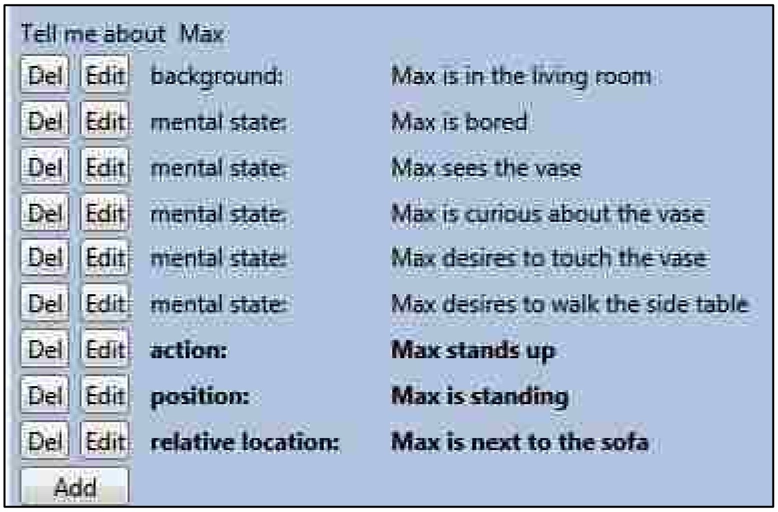


Figure 9: Annotations for key frame 3: "Max stands up"

Figure 9 shows annotations for key frame 3. The statements set in regular type are statements that still hold true from previous key frames. The statements set in boldface are new information for this key frame. In this figure, the first seven statements, starting with "Max is in the living room" are in regular type, indicating they were originally stated in a previous key frame, but are still true in key frame 3. The eighth statement, "Max stands up," is new in key frame 3. It is an action, and its effects are that "Max is standing" and "Max is next to the sofa." (In the previous key frame, Max had been in a sitting position and his location was on the sofa.) Because they are new to this key frame, all three statements about action, position, and relative location appear in boldface. The key frames, with their actions and effects, comprise a specific narrative experience, one that occurs at a particular time and place and with specific objects.

In addition to creating specific narratives, the user also creates general commonsense rules to explain each specific statement. Chapter 1 describes how the HXP user interface guides

[14] Although the key frames represent completed actions, the key frame captions are presented in present tense, in the style of journalism photographic captions.

the user to create these rules. First, HXP asks the user to explain each statement, either in terms of previous statements, or with a "Just because!" answer that means the explanation is too difficult. If a user explains a statement in terms of previous statements, the user can generalize the explanation in terms of the hypernyms of the statements' predicates and arguments. Thus, a statement such as "Max drops the vase" can be generalized in many ways, such as "A boy drops a man-made object" or "A person moves an object" because hypernyms of *Max* include *boy* and *person*; hypernyms of *drop* include *move*; and hypernyms of *vase* include *man-made object* and *object*.

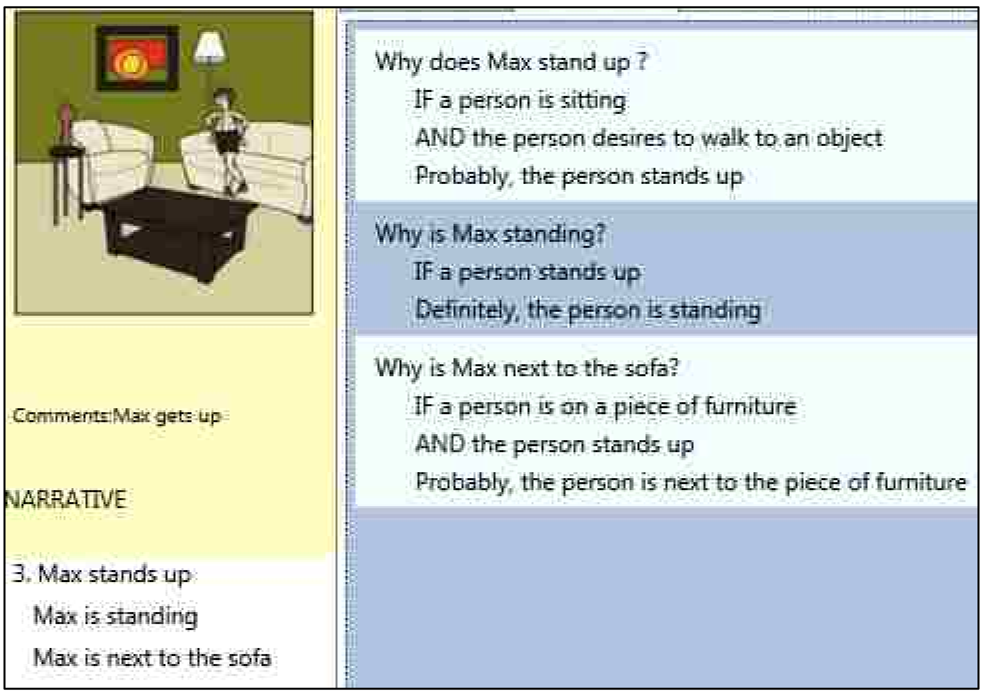


Figure 10: Three statements and If-Then rules that explain them

As a larger example, Figure 10 shows annotations for key frame 3, with three narrative statements and their corresponding commonsense rules.

3.2 Narrative Statements

Having reviewed the dynamic relationship between narrative statements and commonsense rules, we now present the elements used to model specific narratives. As discussed above, narratives contain a series of annotated key frames. Using the digitization metaphor, a key frame is a discretized sample of experiential data, a snapshot of experience. But what should be in this snapshot? First we discuss what data are most important, how much detail is practical, and how HXP captures the data. Then we focus on the structure of the HXP statement, which represents a single action or state. Next, we discuss how to handle different interpretations and redundant data that arise when a community of users collaborates to create narratives. Finally, we present the HXP frame problem – how to determine what statements are true at any given point in the narrative.

3.2.1 Key Frame Data

Key frames correspond to small physical changes in the narrative. If the key frame images do not flow smoothly when displayed in rapid sequence, the user has a visual clue that more key frames need to be added, which will result in more action detail. However, images are not labeled; they have no inherent meaning. They provide no help in describing agents, actions, goals, and other situational data. Like an animation or movie, the unlabeled images are not useful for advancing NLP. The situational data must be explicitly identified, so we must decide what situational data to model.

Since the goal is to advance story understanding, the data in each key frame should be geared towards helping AI programs answer questions about narratives. Mueller (2000) proposed a list of questions that understanding agents should be able to answer. Not surprisingly, these questions cover the Five Ws of journalism: who, what, when, where, why. In addition, they also specifically target commonsense data: intention, locations, and movements.

To get at most of the Five W data, HXP leverages what users already know about the conventions of telling stories through pictures. *Where* refers to the background setting, *what* to the objects in the scene, *who* to the characters, *when* to the time setting.

The *why* question is somewhat different. In journalism and narratology, *why* focuses on salient events – events that make the story interesting. But HXP seeks commonsense, mundane data that is often not salient. To get at this detail, users are prompted to explain every action and effect. Users can either choose to explain in terms of previous statements, or they can fall back on generic explanations such as *This is just one of many possibilities* when it becomes too difficult to find an explanation. Other narrative systems distinguish between different types of explanations, such as cause, partial cause, implication, dependency, or motivation (Singh & Minsky, 2005, Elson, 2012, Zarri, 2010). In contrast, HXP does not explicitly distinguish among causal types because these differences are subtle and difficult for non-experts to work with.

HXP explicitly prompts users to annotate mental states, goals, and intentions. This information is among the most difficult for non-experts to articulate, but the process of explaining every action guides the user into searching for intentions and desires that would motivate these actions.[15]

To capture information about location and movement, several issues arise. Should the system mark the location of every object in the room, and if so, should it use a commonsense grid like (Mueller, 1998), or possibly an absolute coordinate system? And when Max picks up the vase, should the system capture the fact that his fingers are curled, his arms are raised, his

[15] The original HXP design asked users to explicitly organize actions into hierarchies of subactions and subgoals, but early tests indicated this was too abstract and difficult. As discussed in Chapter 6, in the future I hope to automatically capture action/subaction relationships from intentional annotations.

legs are slightly bent and his feet are apart? In sum, how much detail on location and movement should be captured?

In considering these issues, it became clear that many of the body position and object topology questions would be best represented by an objective 3D model, as suggested by (Weltman, 2009). In contrast, even the most detailed 3D model will not have information about the subjective aspects of locations and positions. For example, whether or not Max is near the vase depends on the objects involved and agent intention. If Max wants to see the vase, "near" could mean several feet. If he wants to touch it, "near" could mean one foot. Therefore, the current HXP focus is on subjective annotations that have causal or motivational effects. Rather than specifying that Max is 3.5 meters from the vase or that the vase is at 3D coordinate (4,3,5), HXP simply models subjective room-scale locations like "Max is not near the vase" or "the vase is above the side table." While imprecise, these expressions are valuable because they can explain the causes and effects of actions.

The HXP model ties all specific objects to a background location. That is, when a user adds an object to a key frame, the containing background must be specified. A key frame can have more than one background setting. For example, there might be a scene where Max is in the living room, the dog is outside, and Mommy is in the bedroom. Each of these is backgrounds contains separate objects. Within a background location, HXP has specific prompts to remind the user to specify at least one relative location for each object contained and at least one body pose (e.g. *sitting*, *standing*) or functional state (e.g. *off*, *on*, *broken*).

Unlike (Mueller, 2003), which categorizes representations into different spaces such as room-scale and object-scale, HXP does not have separate predicates for different scales. Users would use identical *next to* predicates to express "Max is next to the table" and "The school is next to the park" even though the objective meaning of *next to* is different. The subtleties of these distinctions would be a burden to non-experts. On the other hand, when the scale is important, users will naturally create rules with different levels of generalizations. That is, there would probably be different consequences depending on whether Max is next to a piece of furniture, versus next to a building. In the future, HXP could be combined with a 3D model to provide easy-to-use mappings between precise, objective data and context-dependent subjective expressions.

Regarding the question about how much movement precision to represent, HXP seeks to depict actions with enough detail so that the movement appears somewhat smooth when the key frames are shown one after the other. Discussion with users indicates that this guideline is clear. Furthermore, keeping the transitions smooth between key frames helps capture the mundane details about an experience that people typically leave out of normal narratives. Finally, this guideline will help ground all narratives to a common level of granularity.

3.2.2 Structure of Statements

The previous section discussed what sort of data should be captured in each key frame and the desired level of granularity. There are still representational issues. How much freedom do users have to express actions and effects? The goal here is to balance structured input with enough expressivity to represent users' intuitions about naïve child experiences.

Each data item in a key frame is in the form of a simple subject-predicate statement. Users create statements via a series of drop-down selections so that there is little or no parsing and thus no ambiguities. As described in Chapter 1, a user first populates a key frame with objects – background settings, props, and characters. Then the user is able to select predicates to annotate the objects. This structured input allows all objects and predicates to be linked either to entries in the standard WordNet ontology, or to the HXP database (for words and concepts not found in WordNet). Internally, a statement is a clause, containing a subject, predicate, optional arguments, and an optional recursive subclause.

There is a conflict in the goals regarding freedom of expression. Natural language can be wonderfully nuanced, but fine shades of meaning and nearly synonymous expressions make it difficult to find patterns in data. Although users should be able to freely express what is happening in the narrative, a child-like, naïve vocabulary would be easier to analyze. Toward this end, I populated the HXP database with a limited number (about 150) stative predicates that I believe are important to story understanding: mental states, relative location, body positions, weather, time of day, etc. However, in the near future, users will be able to add new predicates to the HXP database if there is no way to approximate what they wish to express. In contrast to stative predicates, I do not see a good way to reduce the types of actions that users can select. Therefore, virtually every WordNet verb is available for selection. This freedom is problematic because WordNet has many fine shades of verb meanings. As discussed in Chapter 6, in the near future versions of HXP, I will explore options to have coarser verb meanings.

Another way in which HXP reduces the vocabulary of narratives is by defining a primary sense for each concept. All concepts, including predicates, map to WordNet synsets, which represents a group of synonymous word senses. The HXP primary sense for each synset is either the word with the highest frequency in the WordNet source texts, or it is the word that I believe is the simplest to understand. HXP displays all statements using the primary sense, even if the user had selected a different sense for that synset.

HXP displays the statements as simply as possible, with no conjunctions, a limited use of pronouns, and limited use of subclauses. Having independent, simple statements makes it easier to prompt the user to explain the reason for the statement. The child-centered experiences make it more acceptable to see this stilted writing style which may be familiar to users from their own memory of primary-level reading material. Indeed, in the HXP user test, users were to imagine they are communicating at a child's level when they make annotations, which helped them adjust to expressing themselves as simply as possible.

3.2.3 Different Interpretations and Redundant Data

Even in a simple narrative, people have different points of views about how to interpret an action. In a community of users, one person may think that Max drops the vase because he is naughty; another may think that it is an accident. Some users might think Max is thrilled by the sound of the crash and some might think he is frightened by the sound. Rather than try to get a community of users to agree on a correct interpretation, HXP welcomes all interpretations as long as users can justify the their statements with acceptable commonsense rules.

A community of users will also create redundant data or data that differ only slightly from one interpretation to the next. Redundant data is not necessarily bad. Since our goal is to collect data that link words and situations, redundancy will strengthen the associations that occur most frequently. On the other hand, it is not beneficial to have redundant rules, so future versions of HXP will automatically check for similar rules and suggest to users that they be merged.

3.2.4 The HXP Frame Problem

If Max is near the table in key frame 5 and picks up the vase in key frame 6, is he still picking up the vase in key frame 7? Is he still near the table? The HXP frame problem, related to the epistemological frame problem and other related frame problems (Shanahan, 2009) is to know what is true in any given key frame. One way to do this would be to allow users to mark each statement with the time when it is true and when it stops being true, similar to (Elson, 2012).

Making a trade-off for simplicity rather than power of expressivity, HXP assumes actions are true only in the current key frame, and states (*fluents* in AI terminology) are true until the user explicitly changes them. Furthermore, HXP associates each key frame with exactly one outwardly manifested action.[16] This action, represented by the key frame's caption, is assumed to begin and end inside the key frame, and the key frame's picture is assumed to depict the state of affairs immediately after the action occurs. For example the caption, "Max stands up" has a picture of Max standing because Max has just stood up. "Max walks to the side table" has a picture of Max standing at the side table because Max has just completed walking to the side table.

All the other annotation statements of the key frame are interpreted as effects of this completed action. HXP considers these effect statements to be fluents which remain true in subsequent key frames unless the user specifically changes them. To remind users to check for changes, HXP explicitly prompts the user to verify whether previous states still hold any time a character or object is an argument of the caption's predicate. For example the caption, "Max lifts up the vase" would cause HXP to prompt the user to check both Max and the vase to verify that their previous states still holds.[17]

3.3 Commonsense Rules

Besides specific narratives, the other basic HXP data type is the general commonsense rule. As summarized above, users construct these rules from specific statements, and then

[16] The opening setting key frame is an exception to this rule because the opening of a story is conventionally an ongoing state.
[17] These prompts are not yet fully implemented in the HXP software

generalize them by selecting hypernyms of the objects or predicates. To demonstrate how HXP rules and statements could be used for NLP, I have implemented a simple inference system. It uses single step forward chaining to suggest what else might be true after the user adds a new statement. This section presents the structure and interpretation of the rules and how they are used during inference. Of course, an inference system with a large number of rules is not computationally tractable. The goal here is strictly to demonstrate that HXP rules are well enough structured for automatic analysis. A large number of structured rules would be an excellent test bed for developing new algorithms for deep semantics as well as a valuable training corpus for statistically oriented NLP.

3.3.1 Structure and Interpretation of Rules

Rules are simply Horn clauses displayed as If-Then statements. However unlike Horn clauses, rules are modally qualified by either *definitely*, *probably*, or *possibly*. The confidence level helps prioritize which rule to match during inference.

When users try to create a general rule of common sense, they can become bogged down if they try to consider all the possible situations and exceptions to the rule. For example, if a user creates a rule to explain why a boy walks to a table, it is not productive to note rare exceptions that would keep the boy from using the normal means of travel – the room is under water, the boy is wearing skates, etc. Therefore, HXP users are instructed to employ only what they consider to be “normal” situations when they formulate rules. Of course, this is highly subjective and relies on a shared cultural context. However, as stated before, the system is intended to be wiki-based where users share and discuss each other’s work, so there is a means to reach consensus on what is normal. Furthermore, HXP rules are defeasible, which means a more specific rule has priority over a more general one. This allows exceptional situations to be covered by more specific rules.

3.3.2 Partial Order on Statements

During forward chaining, the statements in the current situation are paired with the antecedent statements of each rule to see if there is match. I define a partial order on statements that allows specific statements of the narrative to be matched with more general statements of a rule. That is, given two statements s1 and s2, we want to determine if s1 is a more specific version of s2. The HXP partial order on statements is based on the hypernym relationship among concepts. Formally, for concepts c and d:

$c \preccurlyeq d$ IFF $c = d$ or $d \in \{h\text{: h is a hypernym of c}\}$

Building from the partial order of concepts, there is partial order among clauses. Recall that an HXP clause contains a subject, predicate, and optional predicate arguments, each linked to a WordNet synset or an HXP concept. Therefore, clause m is more specific than clause n if each of the elements in m are generalized in n. More formally:

$$
\begin{aligned}
m \preccurlyeq n \text{ IFF } \quad & \text{m.subject} \preccurlyeq \text{n.subject} \quad \text{AND} \\
& \text{m.predicate} \preccurlyeq \text{n.predicate} \quad \text{AND} \\
& \text{m.predicate.arg[i]} \preccurlyeq \text{n.predicate.arg[i]}, \ (i \leq \text{m.predicate.arg.Count})
\end{aligned}
$$

Having defined the partial order on clauses, we can do so for statements, which consist of at least one clause. Statement s is more specific than statement t if all of the clauses in s are generalized in t:

$$s \preccurlyeq t \text{ IFF } \quad \text{s.clause[i]} \preccurlyeq \text{t.clause[i]}, \ i \leq \text{s.clause.Count}$$

3.3.3 Partial Order vs. Entailment

It might seem that the partial order of statements implies entailment. For example, in the statements below (1) $\preccurlyeq$ (2), and also (1) entails (2), or (1) $\supset$ (2).

1) A boy is next to a chair

2) A boy is next to an object

However, statement ordering implies entailment only in the case of upward entailment. The negation of the statement causes downward entailment, which means that if a statement is true, its more specific statement must be true. Thus, (3) $\supset$ (4) because 'easy chair' is more specific than 'chair'. But (3) $\not\supset$ (5). Just because a boy is not next to a chair does not mean that the boy cannot be next to some other object.

3) A boy is not next to a chair

4) A boy is not next to an easy chair

5) A boy is not next to an object

However (3) $\nprec$ (4). Furthermore, (3) $\prec$ (5) even though the entailment relation does not hold. Nevertheless, it is the partial order that best serves the purpose for inference. Consider Rule (1) below:

1. IF a person is not next to an object
   AND the person desires to touch the object
   THEN Probably the person goes to the object

This rule correctly matches situations where a person is not next to a vase, not next to a table, etc., and wishes to touch the object. If entailment were used instead of statement order, the rule would not match.

3.3.4 Rule Prioritization

A system that implements defeasible logic must have a way to prioritize the rules. HXP defines the priority of a rule based on its confidence and specificity: The higher specificity takes priority. In case of a tie, the higher confidence takes priority. The intuitive notion of specificity is similar to (Nute, 1993), but rather than using rule derivations, HXP makes use of existing hypernym relationships to determine specificity. The intuitive idea is that a statement that uses a broad hypernym like *object* is less specific than a statement that uses more narrow hypernyms like *person* and *piece of furniture*. This specificity is based on the Hypernym Distance (HDist), defined as the number of intervening concepts between the base concept and the hypernym.

Given a hypernym chain like *boy* → *male* → *person* → *living thing* → *object*, the distance from *boy* to *person* is 2; the distance from *boy* to *object* is 4. The total distance for a rule is simply the sum of all the HDist values in the rule, where the number of HDist values is computed from the unique concepts contained in the rule's set of statements.

$$\text{Tot Dist} = \sum_{s \in S} \sum_{c \in C} \text{HDist}(c) \qquad (1)$$

where S are the statements of the rule's antecedent and consequence, and C are the unique concepts in S.

A large total distance value from Eq. 1 represents a very general rule; therefore, as a first approximation, we take the inverse of the total distance to represent specificity.

The total distance in Eq. 1 does not take into account the fact that a higher number of statements in a rule decrease its generality, which makes it more specific. For example, a rule with three antecedents is usually more specific than a rule with two. $(A \wedge B \wedge C) \rightarrow D$ is more specific than $(A \wedge E) \rightarrow F$. HXP approximates this relationship by modifying Eq. 1 to compute the average distance, which generally decreases with more statements. The decrease is due to the fact that there will be fewer unique concepts introduced with each additional antecedent.

$$\text{Avg Dist} = \frac{1}{|S|} \sum_{s \in S} \sum_{c \in C} \text{HDist}(c) \qquad (2)$$

Specificity is defined as the inverse of average distance. Thus, rules that use fewer hypernyms or more antecedent statements will have a higher specificity value.[18]

3.4 Summary

This chapter discussed several issues raised by using non-experts to collect narrative statements and commonsense rules, data that currently is available only by using expert labor. The overriding requirement of the HXP data structures is that they be easy to read, write, update, and share by non-experts, as well as be suitable for automated NLP. With the goal of making the task as concrete as possible, HXP divides the task into key frames, each visually depicting one specific action. During annotation, the user describes the effects on the objects and characters in the key frame. Rather than asking users to annotate objective data such as specific locations and detailed positions, HXP prompts users for subjective data that has an impact on the subsequent actions. Users describe the experiences with a reduced vocabulary, with single subject-predicate statements, reminiscent of books for beginning readers. The goal is to capture a coarse approximation of the action and effects rather than to describe fine details with nuanced meanings. There is no “correct” interpretation of an experience. Any interpretation is valuable as long as it can be justified with an acceptable commonsense rule. Redundant data are also acceptable because they help strengthen the association between words and situations that occur most frequently.

[18] The computation of specificity has not yet been implemented in the current HXP system.

The HXP rules are Horn clauses. The rules are defeasible to make it easier for non-experts to focus on typical situations and to ignore exceptional cases until they arise naturally in a less typical experience. Users assign confidence levels of *possibly*, *probably*, and *definitely* to indicate a rules priority. As a demonstration of the future application of HXP data to real NLP tasks, the HXP single-step forward-chaining inference process predicts what else may be true from a given statement. Using the hypernymy relation, already present in WordNet, we can define a partial order on statements. Basically, when matching narrative statements against the more general statements in rules, HXP compares the distance between their corresponding hypernyms, also taking account the rules confidence level. This demonstration addresses a key issue in commonsense modeling, particularly for CBR systems: how to compare one situation with another in order to find the closest match.

CHAPTER 4: DESIGN OF THE USER INTERFACE

This chapter discusses the design of the HXP user interface software. The goal is to interact with users via a subset of English that is both unambiguous and easy to understand. To do so, HXP defines a controlled natural language (CNL) that is specifically geared towards simple physical states, simple actions, and intentional states. As discussed in Chapter 2, HXP shares many ideas from Elson's Scheherezade system (Elson, 2010). They both use a WYSIWYM interface, which means that the system restricts the user's input to simple fill-in-the-blank or drop-down selections but displays the data as natural language statements. Nevertheless, Elson's CNL is designed for paraphrasing of descriptions and dialog in existing stories, whereas HXP's goal is to capture experiential details as simply as possible. Also discussed in Chapter 2, HXP and ACE (Fuchs, 1996) both use simple predicate argument structures, but ACE is geared toward representing logic problems, whereas HXP is geared toward representing simple experiential narratives.

As with designing a programming language or communication protocol, a significant challenge to designing the HXP input language is choosing vocabulary that will provide the most expressivity for the smallest number of choices, giving them intuitive names, and organizing them for easy access. Another challenge is designing the software such that a small number of components can be combined to support a large number of predicates.

This chapter describes the predicate organization and corresponding software data structures. Dividing predicates into *statives*[19] and *actions*, this chapter describes several issues with the stative predicates, followed by a short section on action verbs. Then it describes several

[19] Stative predicates express static physical properties such as color, size, location, as well as mental states.

HXP *value templates*, which are HXP's reusable software structures that specify when and how to prompt for a predicate's arguments.

4.1 Stative Predicates

There are currently about 150 stative predicates usable in HXP, categorized according to *mental state* or *physical state*, with many subcategories.

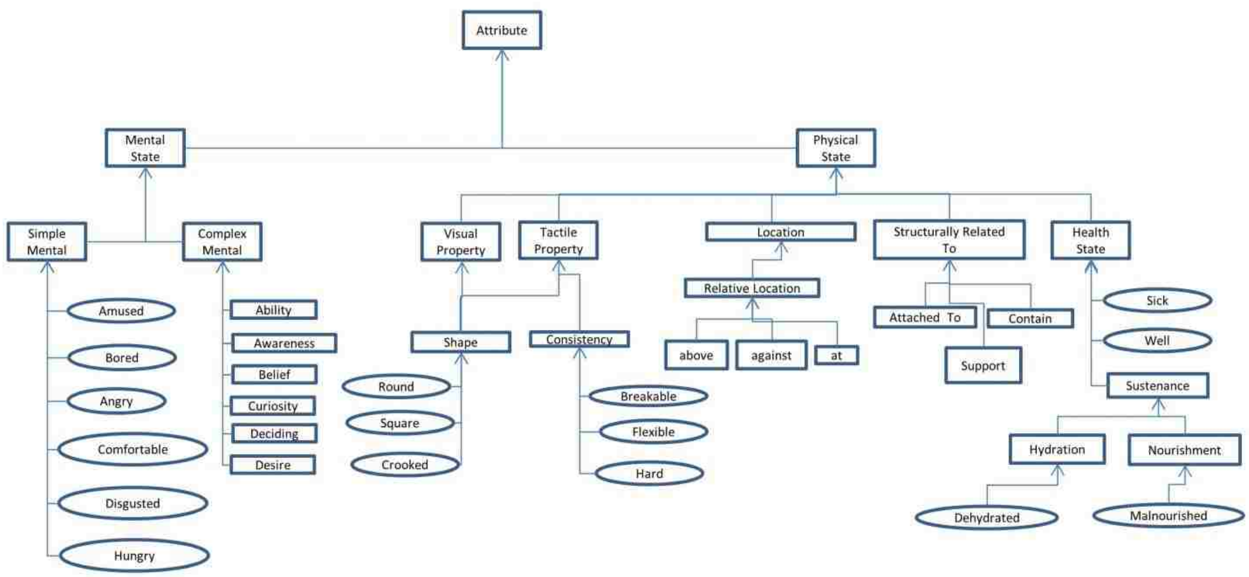


Figure 11: Partial hierarchy of HXP stative predicates

Figure 11 shows a partial hierarchy of HXP stative predicates. Words in square boxes are predicates, and words in ovals are values of *semantic fields*, which are discussed below. The top predicate on this diagram is simply *attribute*, which is divided into *mental state* and *physical state*. Under *mental state* are simple mental states like *amused* and *bored*, as well as complex mental states like *ability* and *awareness.* (The difference between simple and complex is also discussed below.) On the other side of the diagram are the physical states, which include *visual property*, *tactile property*, *location, structurally related to, health state,* and many others not shown, but which are discussed below. A specific predicate can be classified in multiple categories. As the figure shows, *shape* is both a visual property and tactile property.

In the subsections that follow, I describe several design issues with the stative predicates. The first describes predicates that have special user interface prompts to articulate important commonsense data. The second subsection describes special handling of predicate adjectives like *green* and predicate nouns like *daytime*. It treats them as one of several values of a semantic field.[20] The third subsection describes my categorization of physical states into various subcategories to make it easier for users to find predicates. The fourth describes several issues with the HXP treatment of mental states.

4.1.1 Predicate Categories with Special Prompts

There are four types of key frame objects in HXP: key frame, background, character, and prop. Depending on the object, the HXP user interface outputs different prompts. Figure 12 shows the prompt for key frames. The user must enter a *time period* (e.g. daytime) and must choose a *background* (e.g. living room).

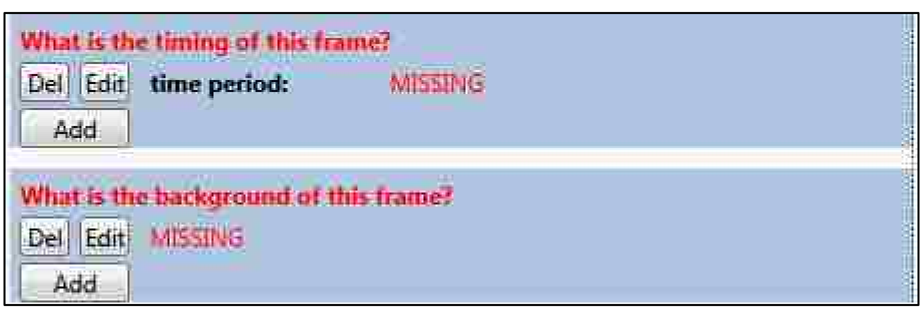


Figure 12: The key frame requires a time period and a background

Figure 13 shows the prompt for a background object such as a living room. As the figure shows, the living room has some parts (floor, walls) and props (sofa, coffee table, etc.) but no characters.

[20] In linguistics, a semantic field describes a set of words related to a specific domain. For instance, blue, green, and red are in the color semantic field.

Although it might seem odd to have an experience without a character, characters are not mandatory so that users could create a scene where, for example, a rock rolls down a hill. Note that in the figure, the parts and props for the living room are part of the stock living room in the system. However, the user has the option of adding more parts or props. For example, in the "Max Breaks the Vase" experience, the user will be adding a vase.

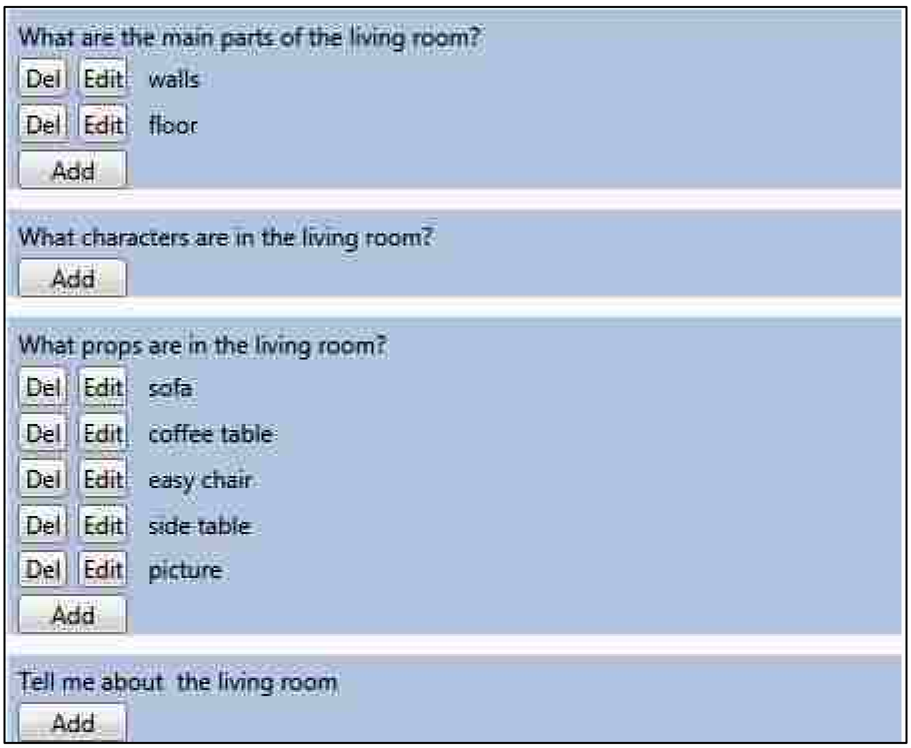


Figure 13: A background has prompts to remind users to fill in the parts, characters, and props. However, there is no mandatory data. For example, this background has no character in it.

Figure 14 shows the prompts for objects like *Max* that are the characters of the narrative. Since the user has not yet specified Max's *relative location*, *position or state*,[21] and *mental state*, the user interface highlights these predicates as having missing values. As the user creates a value for one of these special predicate categories, the user interface changes. Figure 15 shows

[21] The *position or state* predicate is a composite consisting of body positions or functional states; the goal is for users to specify the position of a character or object, for example *sitting* or *standing*. However some objects, like a lamp, might be better described with a functional state like *broken* or *off*.

that the user has provided a value for a *relative location* and *position or state*, but not a mental state.

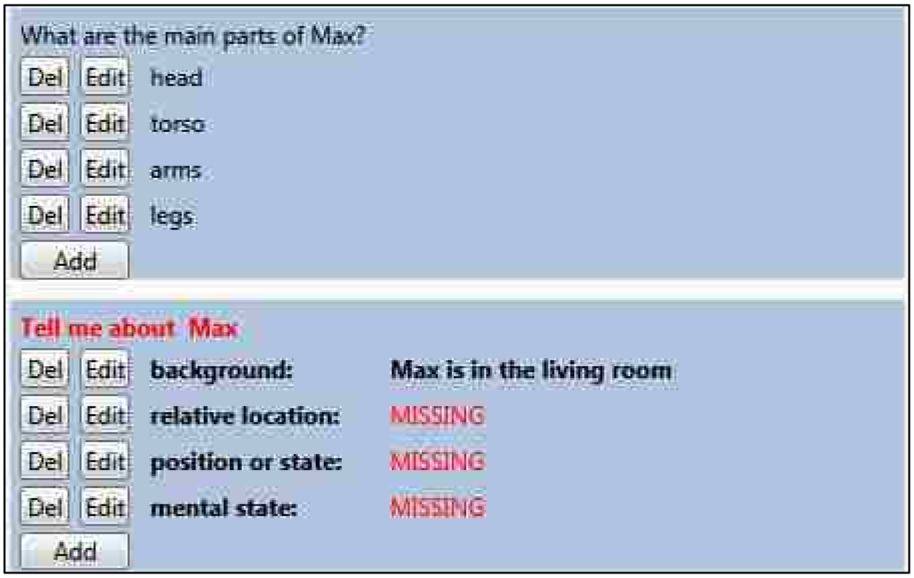


Figure 14: Because *Max* is a character, the user interface prompts for *relative location*, *position or state*, and *mental state.*

Figure 15: The user has entered a value for *relative location* and *position or state*, but not yet for mental state.

The prompts for props are similar to Figure 14 except that there is no prompt for a prop's mental state.

4.1.2 Semantic Field Values

HXP treats single-argument static predicates like *green*, *cold,* or *good* as continuous values in semantic field, similar to Jackendoff's focal values in a continuous domain

(Jackendoff, 1990, p. 34). For example, *green* is a possible argument value for the semantic field predicate *color*. Similarly, *cold* is a possible value for *temperature*. *Good* is a possible value for several predicates such as *visual property* and *sound property*. Appendix C lists the HXP semantic field predicates and their possible values.

One of the advantages of organizing single-argument predicates by semantic field is that the user interface can present a range of related values and let users select an intermediate value. For example, the interface could show a temperature scale with *cold* at one end and *hot* at the other. Another advantage to this organization is that qualitative terms like *good* are explicitly disambiguated with respect to the semantic field they are modifying. In "Max listens to the violin; Max thinks the violin is pretty," *pretty* probably refers to the sound of the violin, but it is a source of ambiguity. By requiring users to choose the semantic field (*visual property* or *sound property*) when choosing *pretty*, the meaning is disambiguated. Furthermore, most of these single-argument predicates are adjectives, and unlike verbs and nouns in WordNet, adjectives have no hypernym links. Thus, it makes sense to mark them for special handling.[22]

4.1.3 Physical States

Appendix D lists all of the approximately 100 current HXP physical states and their categories. I created many non-traditional categories and subcategories because I believed it would help users find the concepts to describe an object. For example, I created *consistency* to group predicates like *breakable* and *hard*, even though WordNet does not categorize them in this way. I created the *structurally related to* category to group predicates of structure and composition such as *part of*, *attached to*, *contain*, *provide access to*, *support*, and *touch*. I

[22] Despite the advantages listed here, there may still not be enough to warrant the separate treatment of single-argument stative predicates as semantic field values. As the HXP system evolves, I may change this organization.

created the *health state* category to group values like *sick*, *well*, and *injured*, as well as the *sustenance* subgroup which relates predicates such as *nourishment* and *hydration*.

However from the user test, it was clear that users do not care about categories, and they much prefer to simply type in a word like *breakable,* or its synonym, rather than start at the top by selecting *physical state* and drill down to the desired predicate. Thus, the categories are of minimal value for helping users choose predicates.

On the other hand, categories are important for generalizing situations in which a character attends to an attribute of something. In one situation, a little boy could be curious about the color of a vase. In another, little girl could be curious about the shape of a cake. In both cases, the situation can be generalized to *a child* being curious about a *visual property* of an *object*. For this reason, there are clear advantages to keeping a hierarchy of categories. But rather than rely on my intuition or even the judgment of group of experts, perhaps the categories should grow organically as users need to make generalizations. That is, when a user wishes to make a generalization, the input screen could allow him or her to type a new category.

The states in Appendix D are just a starting set of physical state predicates. Users need to be able to add new ones when there is a situation that requires it. In the next version of HXP, users will have an input screen for adding a new physical state either by choosing a word from WordNet that has not yet been included in HXP or by typing it directly.

4.1.4 Mental States

Representing and reasoning about mental states is an important aspect of AI (Bacon, 1995). For example, the BORIS system (Dyer, 1983) created a small number of fundamental mental states which could be combined to form more complex states. BDI frameworks (Rao and Georgeff, 1995) model belief, desire, and intention to build agents that formulate plans based on dynamic conditions. Unlike these projects, which seek to formalize the components of mental state and build intelligent agents, HXP does not create an agent model or define mental state primitives. The primary goal for HXP is to collect annotated information, not to build a simulation or to create a theory of meaning. I believe that complex mental states like *curiosity* will eventually take on meaning based on how they are used in a corpus of digitized scenes.

While the precise meaning of a mental state is not an issue with HXP, deciding which predicates map to mental states and designing the structure of the arguments *is* a necessary part of the user interface. It is not always clear in WordNet which concepts can be values for mental states and what their predicate arguments should be. As an illustration, this subsection describes an analysis of *hungry*, which many researchers might say is a physical state. It then presents an analysis for *feel*, which is one of the few cases where WordNet's definition is too broad for HXP. Finally, it discusses *simple* versus *complex* mental states.

Analysis of *Hungry*

WordNet defines two senses of *hunger*. One is a physiological need for food, a *physical state*. The other is a metaphorical non-food desire, which would be a *mental state*, perhaps with a connotation that it is a desire so strong that feels like a physical need. However, WorldNet's definition of the adjective *hungry,* as it applies to food*,* is somewhat inconsistent with the nominal form because it is defined as a desire <u>and</u> a need to eat, that is, both a mental and physical state. But it should be possible to annotate experiences where a character does not have a physical need for food but nevertheless feels hungry, and vice versa, where a character is focusing on something else and forgets to eat even though he needs food. To deal with this separation of concerns, I categorize *hungry* as a mental state, not physical. And I provide the physical state of *need food* to indicate that a character is physically in need of food. The same separation of physical and mental applies to many states: *thirst*, *tiredness*, *pain*, etc. It is important to note that this classification matters only because a state like *hungry* is allowed to stand in as a *mental state*, which is important because at least one mental state is required of characters.

Analysis of *Feel*

WordNet's *feel* synset is one of the few cases where a WordNet sense is too coarse for HXP. WordNet has this definition for *feel*: *(perceive by a physical sensation, e.g., coming from the skin or muscles) "He felt the wind"; "She felt an object brushing her arm"; "He felt his flesh crawl"; "She felt the heat when she got out of the car."* This definition does not distinguish between feeling an internal sensation, like a pain, from an external object, like the wind. However, HXP makes this distinction because the two senses require different argument types.

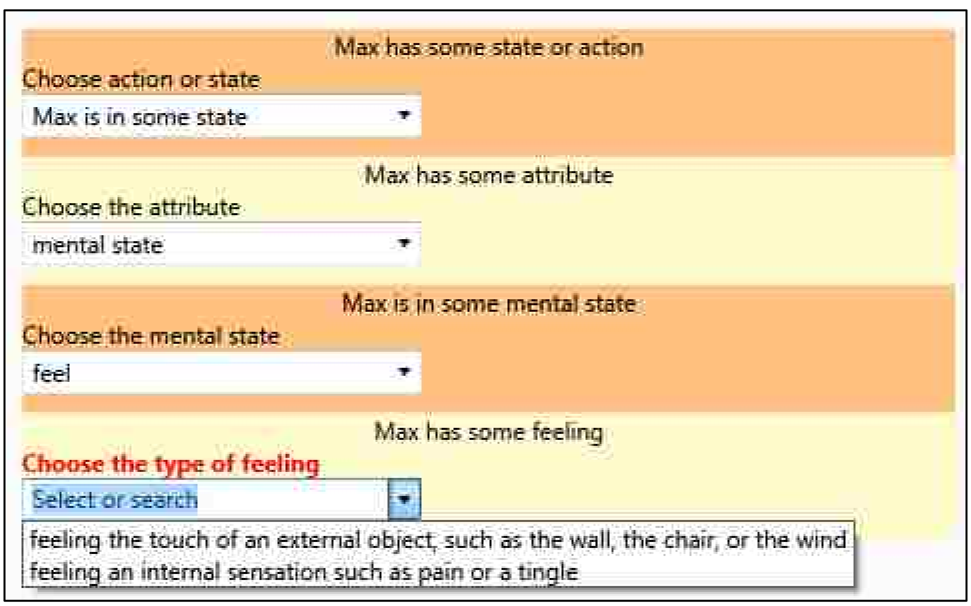


Figure 16: Users select between feeling an external object or feeling an internal sensation.

Figure 16 shows that when users select *feel*, they have two choices: 1) feeling the touch of an external object, such as the wall, the chair, or the wind; and 2) feeling an internal sensation such as pain or a tingle. Figure 17 is an example of the first choice, where Max feels *pain*. Figure 18 shows an example of the second choice, where Max feels the coffee table, perhaps after bumping into it.

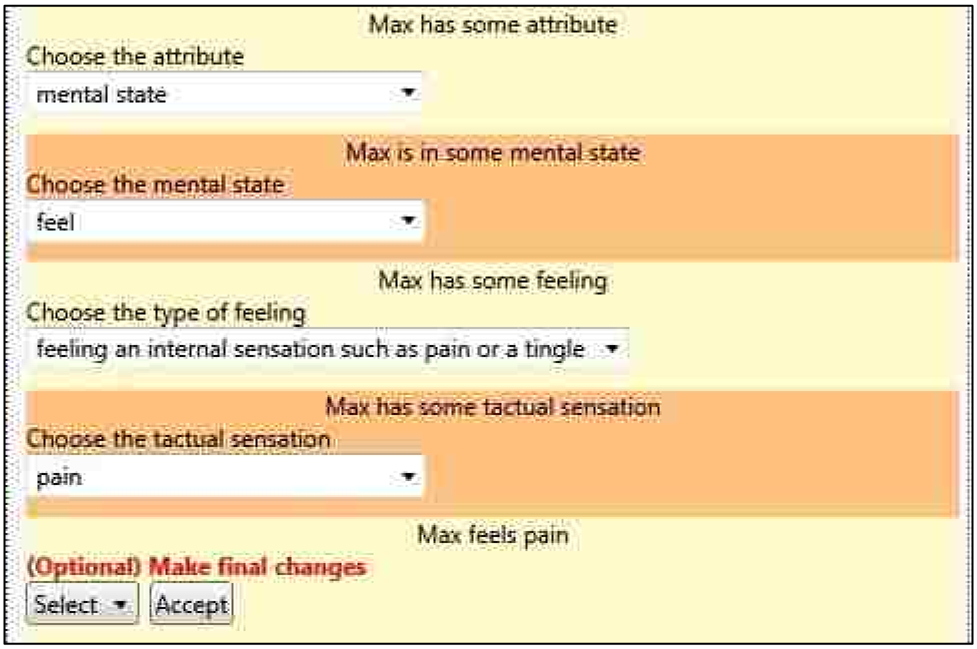


Figure 17: To show that Max feels pain, the user would select "feeling an internal sensation such as pain or a tingle."

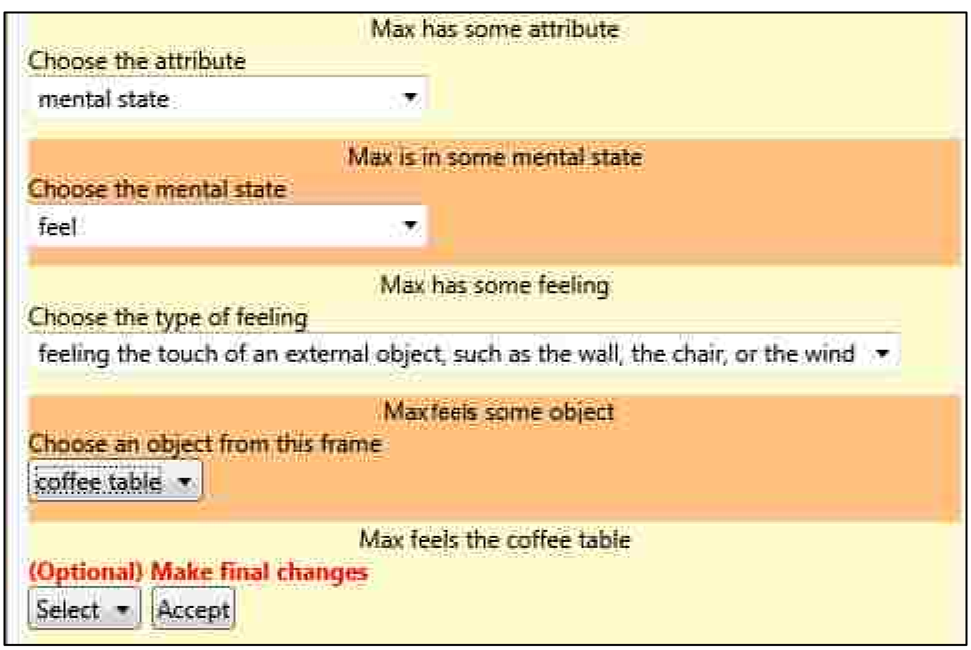


Figure 18: To show that Max brushes up against the coffee table, the user selects "feeling the touch of an external object."

Mental States: Simple versus Complex

Appendix E lists the roughly 50 simple and complex HXP mental states. The value for a *simple mental state* is a single term like *bored* or *amused*, with a subject and no other arguments.[23] That is, "Max is bored" and "Max is amused" have no further description. In contrast, the value for a *complex mental state* like *curiosity* has several arguments. For example, in "Max is curious about the weight of the vase," the argument to *curiosity* includes *weight* and *vase*. In "Max desires to break the vase," *desire* has a clausal complement "to break the vase." Of course, we could have examples of supposed simple mental states with arguments, such as the following:

8) Max is angry that the vase broke.
9) Max is bored with kicking the ball.

We could also use the complex predicate *curious* without an argument:

10) Max is curious.

However, my goal is to reduce the vocabulary and predicate structure as much as possible and still provide users a way to express themselves. A user could express (8) with two statements:

11) The vase breaks
12) Max is angry

Similarly for (9):

13) Max kicks the ball
14) Max is bored

While not perfect paraphrases, they are sufficiently descriptive for modeling simple experiences. On the other hand, it would be difficult to express the "Max is curious about the weight of the

[23] Simple mental states are a subset of affect states in psychology. However, some affect states such as love and hope are complex mental states because they take other arguments.

vase" with a (10) and another sentence. Likewise, it would be difficult express a desire with a separate simple sentence.

4.2 Action Verbs

Whereas I tried to add to the system basic mental states and important physical states needed to capture an experience, actions are much more open ended. Ideally, users would be able to describe virtually any action; therefore, almost all of the WordNet verbs are available to be used for actions. The exceptions are verbs which WordNet marks with the lexical domain *cognition*, *emotion*, *perception*, and *stative*. In these domains, the verbs are mostly stative, not actions. In the remaining WordNet lexical domains, there are currently more than 11,000 action verbs. Many of these verbs have fine shades of meaning that make it difficult for users to choose an action. For example, one sense of *walk* means to "use one's feet to advance." Another sense means "traverse or cover by walking."[24] Since HXP does not seek such nuances, a future version of HXP will have fewer choices of verbs, as discussed in Chapter 6.

4.3 Value Templates

HXP's uses a small number of structures, called *value templates* (VTs), which can be combined in various ways to prompt users for the appropriate argument at the appropriate time. For example, there are many semantic field predicates such as *color*, *texture*, and *shape* where the user selects from a collection of values. For such predicates, a VT called VtEnum enumerates the possible choices. For more complex predicates such as *desire* or *belief*, which require several input values, there are complex VT hierarchies. The parent VT (i.e. top level VT)

[24] Of course there are several other walk verbs, such as "accompany or escort" and "obtain a base on balls," but these are easy for users to distinguish.

contains a sequence of child VTs that walk the user through the choices. In turn, a child VT can have its own children, and so on.

The HXP database associates one top-level VT with each predicate. For example, the predicate *color* is associated with *VtEnum*, which enumerates all of the semantic field values of *color* and presents them to the user in a drop-down list. The predicate *next to* is associated with a VT called *VtSpecificObject*, which gets a list of every object in the key frame and lets users choose one of them.

VTs are somewhat related to FrameNet (Fillmore, 2003) because they encode the type of arguments that a predicate can have. However, VTs are simpler than Fillmore's frames. They do not mark semantic roles like *buyer* or *currency*. Rather, they mark the type of HXP item that the user can select (e.g. an object in the room, a semantic field value, etc). Their purpose is simply to make user input easier. Furthermore, with the goal of a reduced vocabulary and sentence structure, a VT allows a very limited choice for a predicate's argument structure. The next section presents the different types of VTs and how they are used.

4.3.1 VtEnum

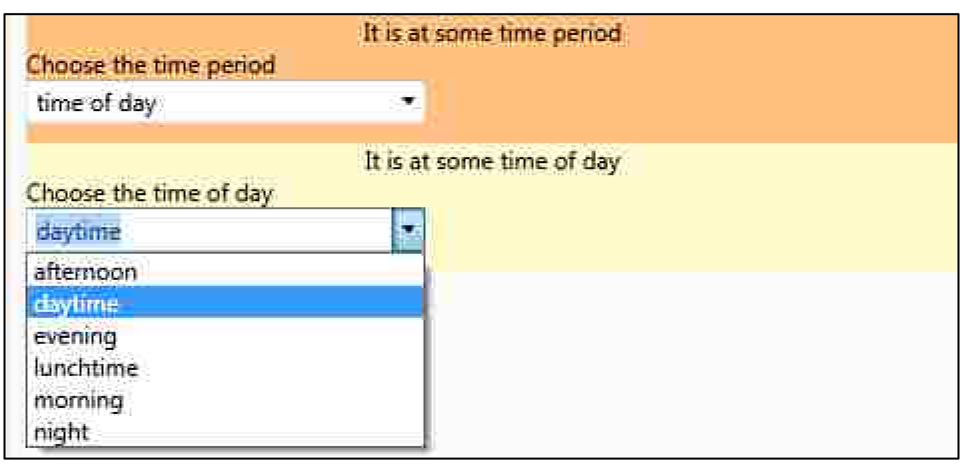


Figure 19: VtEnum displays a drop-down list of items to select, based on an *enum* semantic link.

VtEnum is the top-level VT associated with the semantic field predicates: *color*, *shape*, *time of day*, *simple mental state*, etc. It is implemented by reusing WordNet semantic links. These links have predefined relationships among concepts such as hypernyms and meronyms. In the HXP database, I defined a new semantic link, which I called *enum*. I manually populated the HXP database with appropriate enum links for the various semantic fields. For example, the concepts *afternoon*, *daytime*, *evening*, *lunchtime*, *morning*, and *night* are all linked to the predicate *time of day*. Figure 19 shows an input screen to choose a value for *time of day*. To populate the drop-down list, VtEnum simply retrieves all of the enum links from the HXP data for the given predicate and displays them to the user. A future version of HXP will

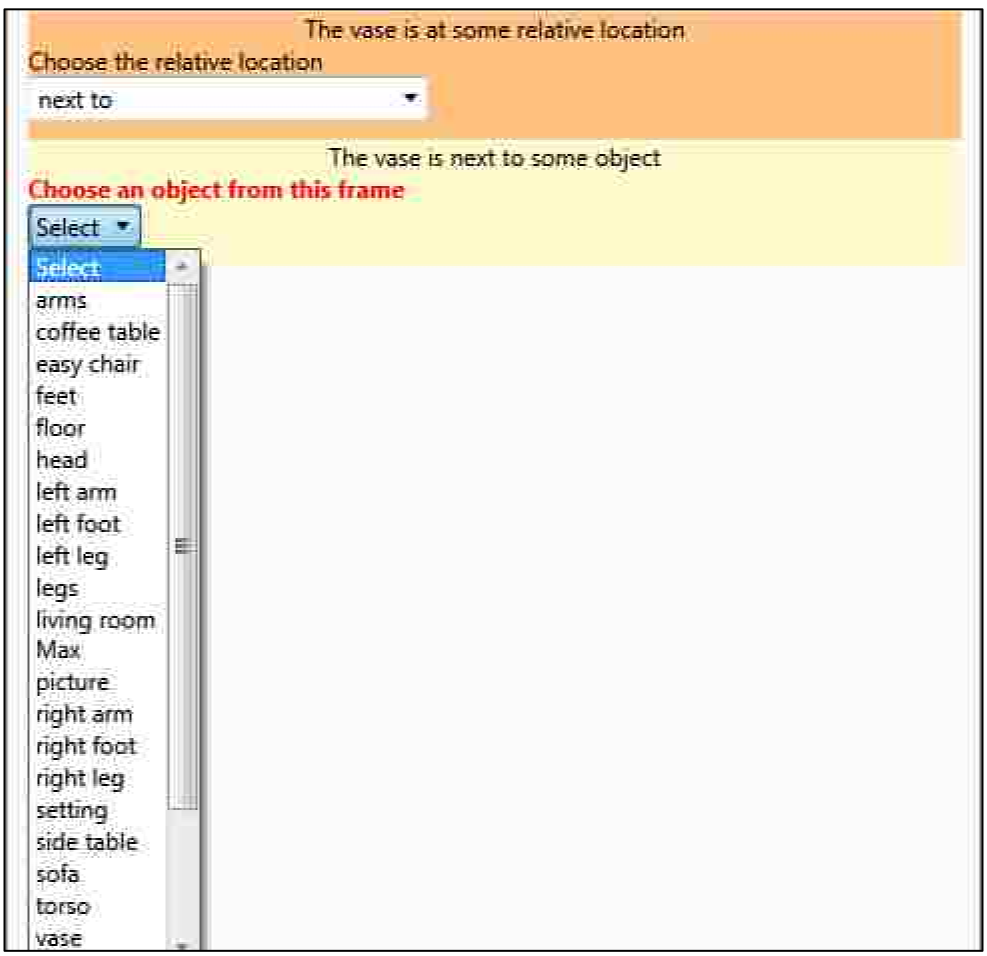


Figure 20: VtSpecificObject displays a list of objects from the current key frame.

have an input screen that allows users to add new enum semantic links so that more choices can be offered.

4.3.2 VtSpecificObject

VtSpecificObject is used whenever the user should choose an object from the current key frame. Figure 20 shows an input screen to select the object of a relative location, and the user can select any object in the key frame, such as *arms*, *coffee table*, and *easy chair*. VtSpecificObject is the top-level VT associated with all relative location predicates. It is also used as a lower-level VT in any situation where an object must be selected.

4.3.3 VtActionVerb

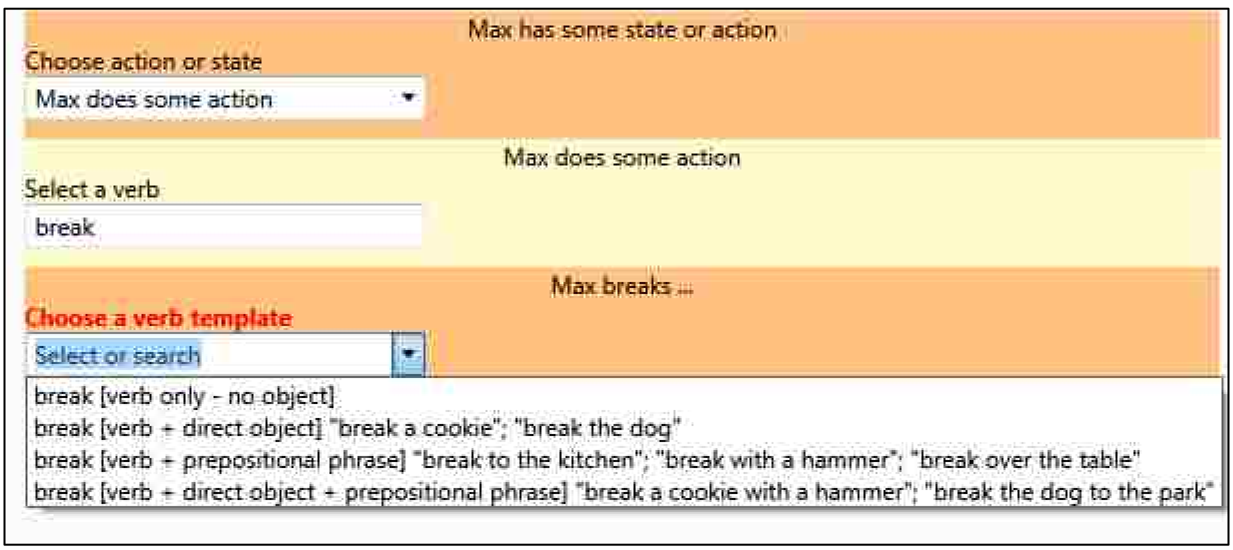


Figure 21: The action verb template allows the user to choose from four verb argument structures.

VtActionVerb is used to fill in the arguments for action verbs. Users can choose from four argument structures:

1. *verb only – no object* (intransitive)
2. *verb + direct object* (transitive)
3. *verb + prepositional phrase* (one adjunct)
4. *verb + direct object + prepositional phrase (*ditransitive*)*

These choices seem to be adequate for most verbs. Depending on which form is selected, VtActionVerb creates the corresponding child VT, either VtIntran, VtTrans, VtAdjunct, and VtDitrans. Figure 21 shows the screen where the user selects choice 2, *verb + direct object*. To handle selection of an object, VtActionVerb transfer control to its child VtTrans, which allows the user to select the direct object.

Figure 22: The user has selected verb + direct object, so now the user is prompted to select an object. The user is going to select *vase*, which will result in “Max breaks the vase.”

As Figure 22 shows, if the user selects *vase*, the resulting sentence that combines the verb and direct object would be "Max breaks the vase." VtTrans reuses VtSpecificObject to handle this logic. To illustrate this reuse, we can see that the drop-down list in Figure 22 above is identical to the list displayed above when selecting a relative location in Figure 20 above.

4.3.4 VtSubordinate

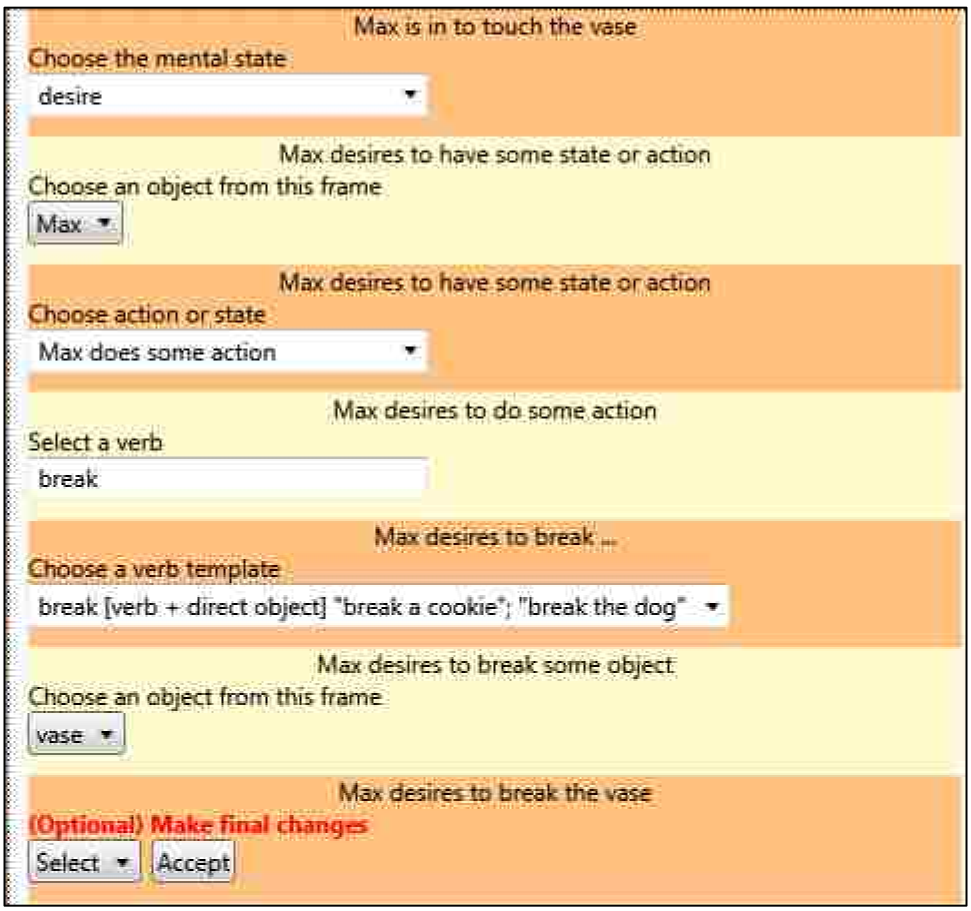


Figure 23: VtSubordinate allows users to specify the arguments of a subordinate clause. It is used with mental states like *desire* and *believe*.

VtSubordinate is used for mental state predicates that require a subordinate clause, such as *able to*, *desire*, *believe*, *decide to*, and *expect*. As shown in Figure 23, VtSubordinate first prompts for the subject of the subordinate clause. (It reuses the previously selected

VtSpecificObject for this task.) The user has selected *Max*, which is the same as the matrix subject. Rather than displaying "Max desires that Max do some action," the software removes the redundant subject in the subordinate clause and outputs "Max desires to do some action." After prompting for the subject of the clause, VtSubordinate then prompts the user to choose an action or state. If the user chooses *action*, VtSubordinate switches control to a VtActionStatement, which prompts for a verb and its arguments, reusing VtActionVerb, described above. If the user chooses state, VtSubordinate switches control toVtStateSentence, which prompts for the state predicate and its arguments.

4.3.5 VtAttention

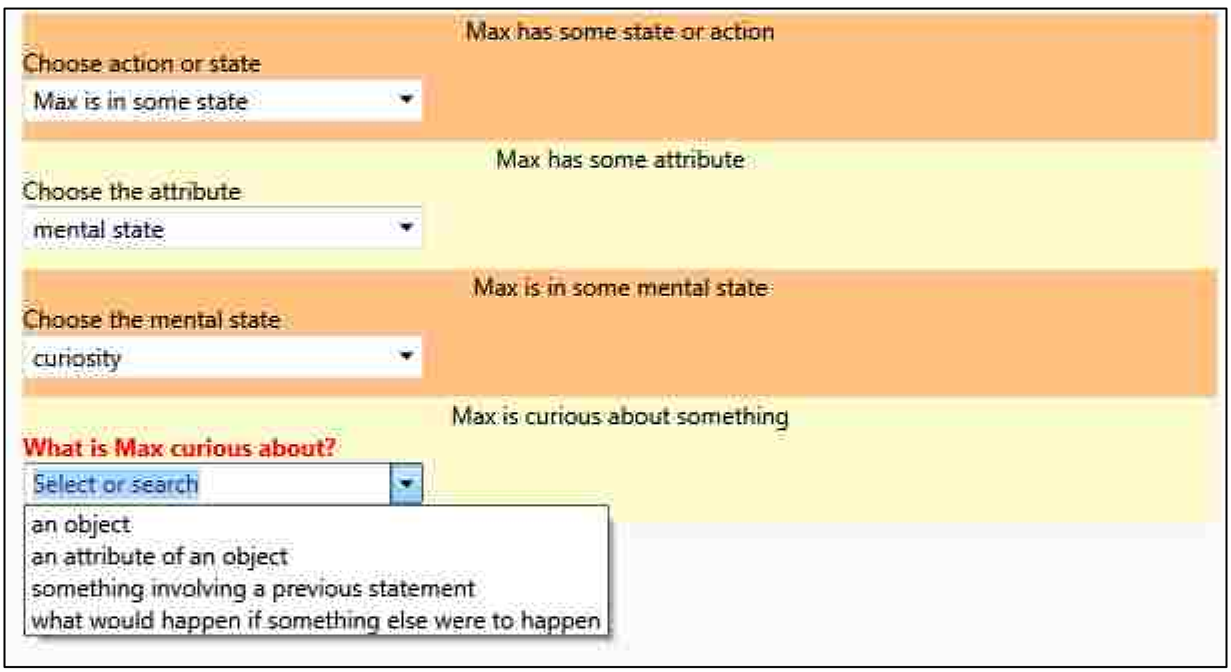


Figure 24: VtAttention is used for predicates that express propositional attitudes and attention. The user first chooses the target of the attitude/attention.

VtAttention is used for predicates that express propositional attitudes and attention like *curiosity*, *know*, *see*, and *love*. The user is prompted to choose the target of the propositional attitude. As shown in Figure 24, the target can be one of four types:

1. *object*: any object in the key frame
2. *attribute of an object*: any stative predicate
3. *something involving a previous statement*: the user will construct a question about a previous statement;
4. *what would happen if something else were to happen*: the user will construct a what-if statement

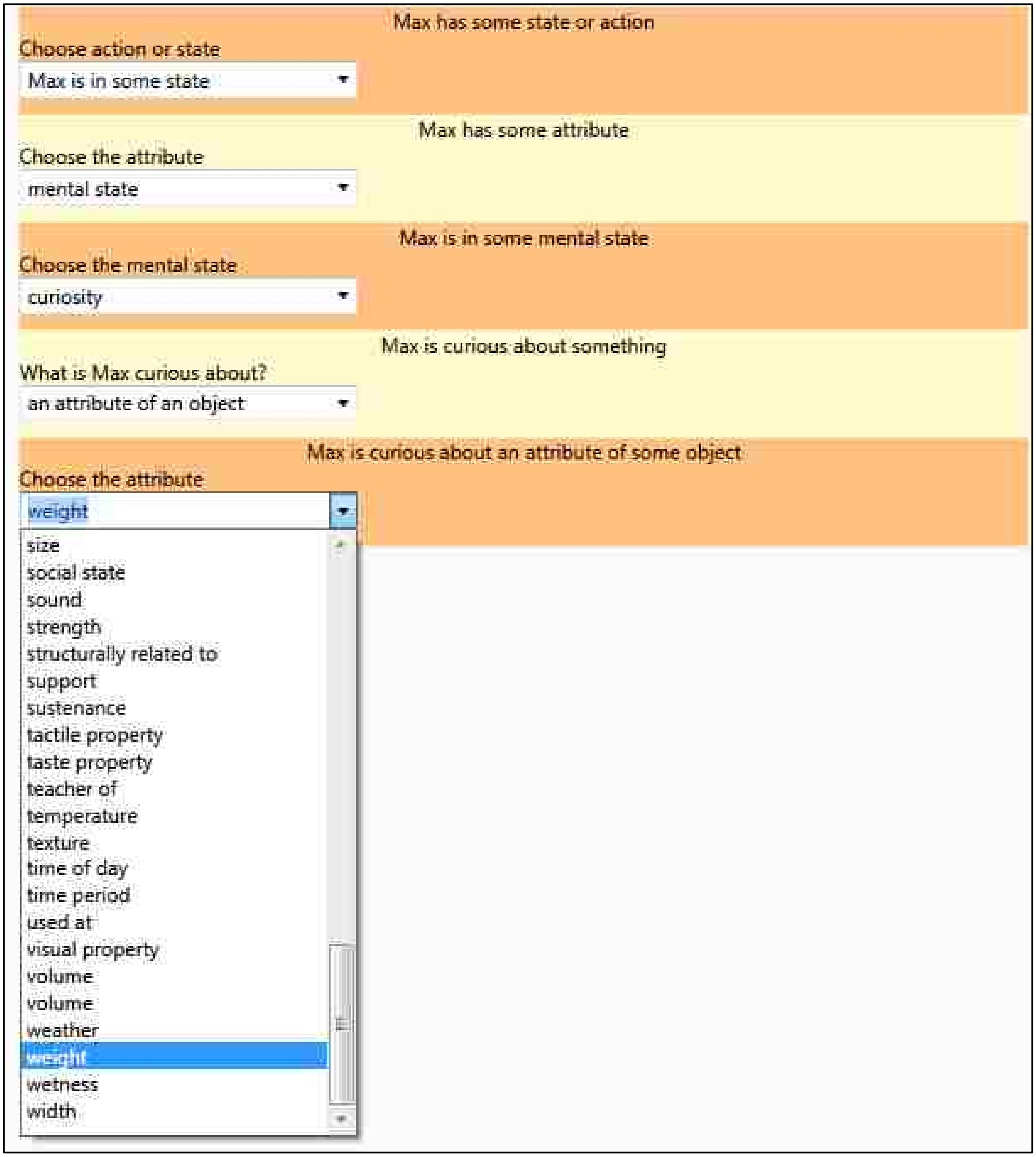


Figure 25: The user is curious about an attribute of an object. In this case, it is *weight*.

These choices allow the user to say things like "Max is curious about the vase," "Max is curious about the texture of the vase," "Max is curious about why the vase is on the side table," and "Max is curious about what would happen if Max drops the vase."

Figure 25 shows a screen where the user has selected *attribute of the object*. In the figure the user is prompted to choose the attribute. In this case, it is *weight*. After selecting the attribute, the user is prompted to select an object, as shown in Figure 26. In this case the user selects *vase*, arriving at the statement "Max is curious about the weight of the vase." As another example of VtAttention, Figure 27 shows how a user constructs "Max is curious about why the vase is on top of the side table." In the middle of the figure, the user has responded to "What is Max curious about?" with *something involving a previous statement*. Following this selection, HXP prompts the user to select an item from the previous statements (not shown). The user selected "The vase is on top of the side table." After this selection, HXP constructs a series of choices from the statement, prepending it with *the fact that*, *why*, *how*, *where*, *what*, and *what happened as a result of*. HXP also creates a filler-gap construction for each argument of the statement's predicate (except for the subject). In this case, the predicate *on top of* has one argument: *side table*. Therefore, HXP creates one filler-gap construction: "what the vase is on top of ___." As another example, suppose, the previous statement had been "Max brings the vase to the side table," which has two arguments. HXP would generate "Max is curious about what Max brings to the side table" as well as "Max is curious about what Max brings the vase to ___."

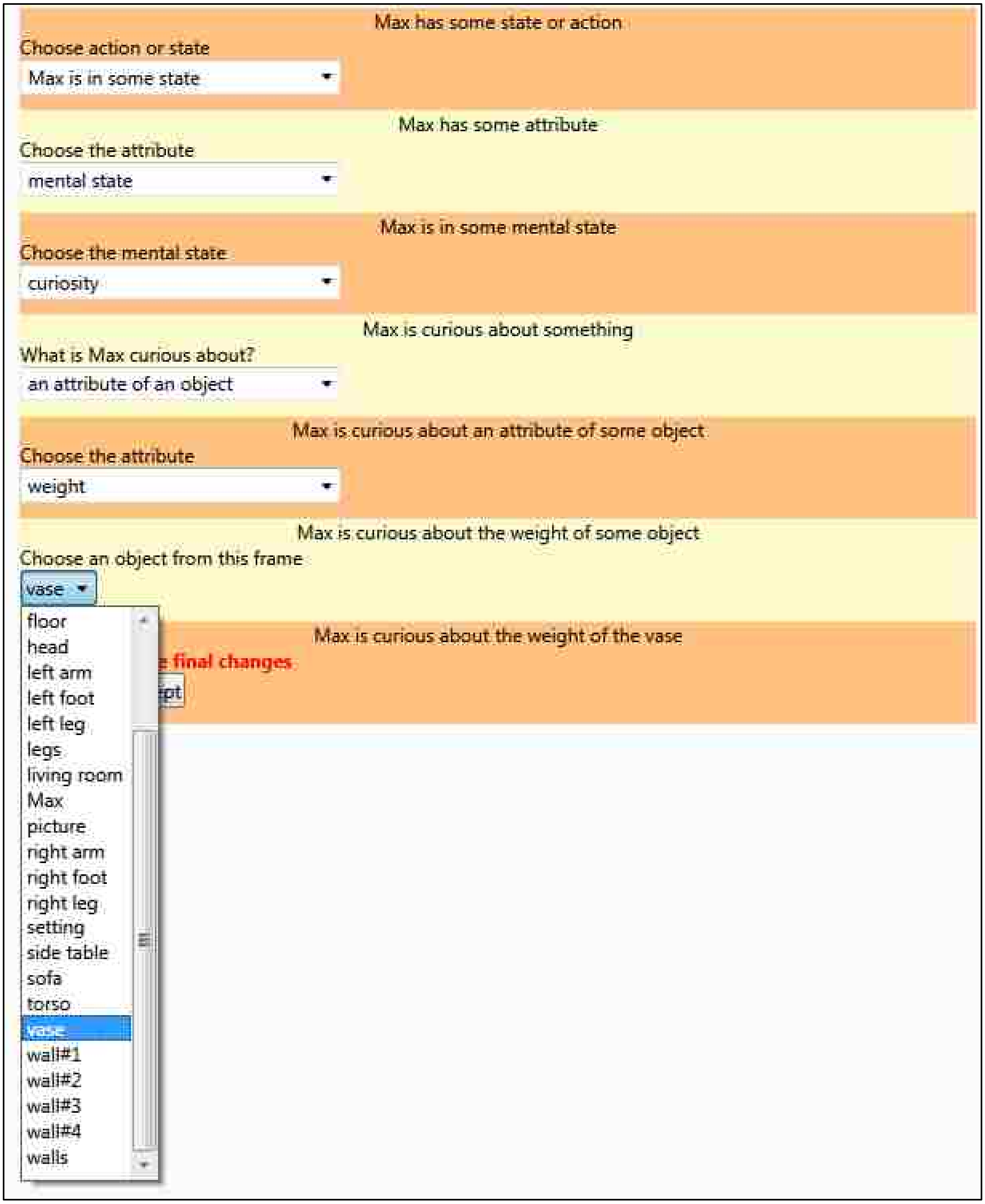


Figure 26: After selecting the *weight* attribute, the user now selects the object.

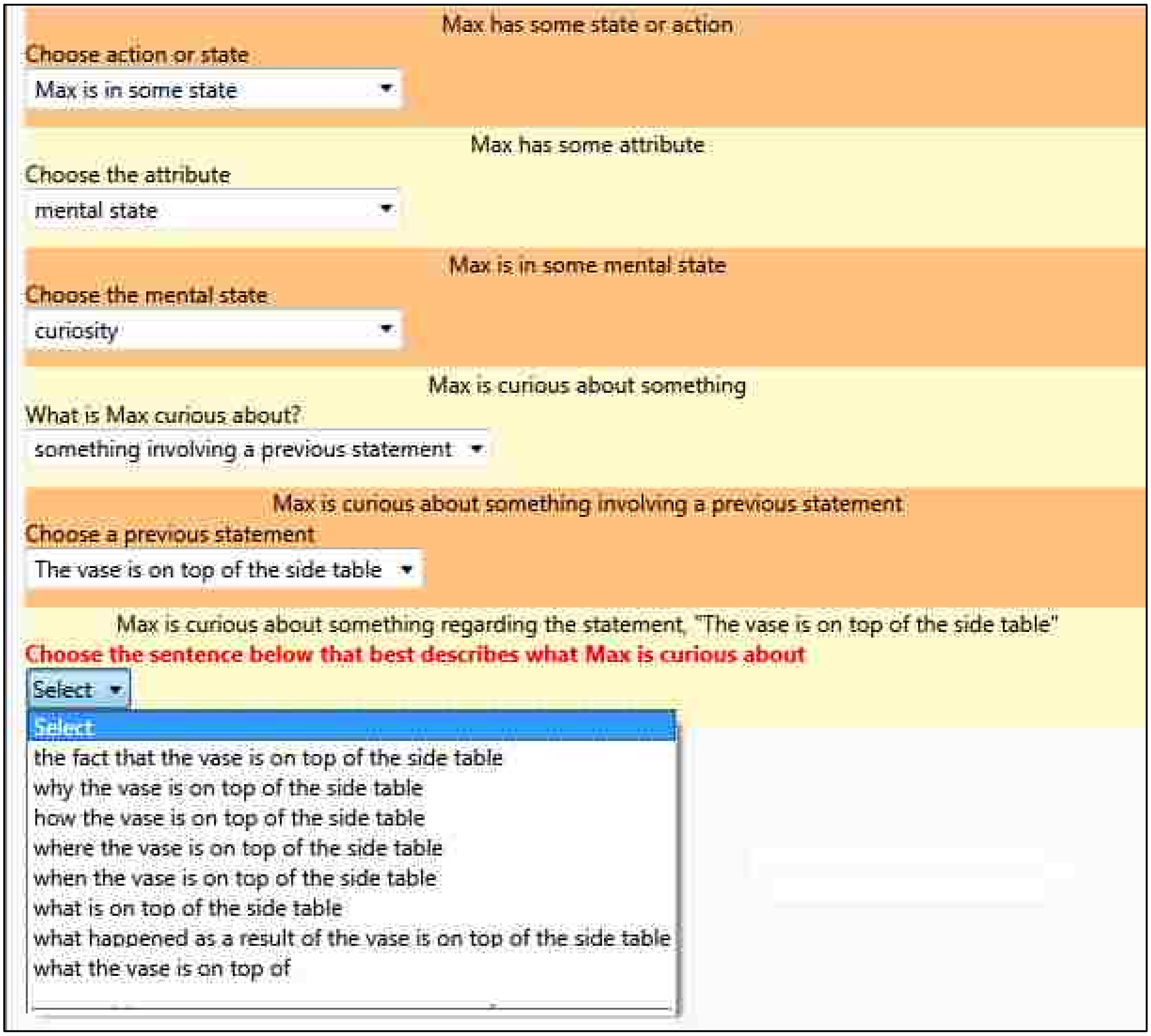


Figure 27: The user is constructing "Max is curious about why the vase is on top of the side table." First the user chooses the previous statement, "The vase is on top of the side table." Then HXP constructs different subordinate clauses from which the user may choose.

4.4 Summary

HXP uses a controlled natural language and a structured user interface to help non-experts enter unambiguous natural language statements. HXP divides predicates into two main classes: *statives* and *actions*. Stative predicates are further divided into *mental states* and *physical states*, and *physical states* are further organized into several subcategories such as health states, locations, and structural relations. The user test revealed that the various subcategories are not very useful to help non-experts find the vocabulary they want. Users prefer simply to type in a word and have the system search for a synonym.

After presenting the various categories of predicates, this chapter describes modular data structures called *value templates*. There are several different types of these structures, and the HXP database associates each predicate with a top-level value template. The value templates provide a general mechanism for guiding the user through a multi-step process to choose input values.

CHAPTER 5: EVALUATION OF THE METHODOLOGY

This chapter presents an evaluation of the HXP methodology via a ten person user test. First, it describes the data collection process for this test. Then it gives an overview of the collected data. After the overview, it presents the data analysis, which includes an evaluation of the data quality as well as the data coverage with respect to what I had previously proposed to collect. Finally, it describes some points of confusion encountered during the user test and their resolution.

5.1 Data Collection Process

There were ten participants for the user test. The test was approved by the Institutional Review Board of LSU, as shown in Appendix F. All of the test subjects were acquaintances, or associated with acquaintances. There were five LSU undergraduates, with one majoring in music, one in English literature, one undecided, and two from biology. There were four professionals with at least a bachelor's degree in accounting, psychology, business computing, and linguistics. Finally, there was one high school student in ninth grade. Using the system requires an understanding of eighth-grade grammar, so the minimum age to participate was 13. For this test, ages ranged from 15 to 54, with half of the subjects below age 25. Half were female, half were male. Although native English ability is not necessary, all of the subjects were native English speakers. Those of student age were paid minimum wage for two hours' work. The evaluation consisted of a one-on-one training session for about two hours, followed by a second session where I asked subjects to annotate on their own a minimum of two key frames.

In the training session, I introduced the subjects to automated NLP and the HXP goal of collecting simple life experiences. Then I went through the first three frames of "Max breaks the vase," creating statements and explaining them with general rules. In each case, the subjects breezed through the first two frames, but as soon as they encountered "Why does Max stand up?" in the third frame, they were stuck. I spent the majority of the training talking them through the process of filling in missing information and generating relevant rules. Consequently, all of the subjects were interactively guided to produce a rule similar to Rule (7) from Chapter 1 during the training.

In the second session, I asked subjects to drive the input, specifying the statements and rules. These sessions were always with individual subjects with no collaboration. Since this was an evaluation of the methodology, not the software per se, I handled the mouse and keyboard to relieve the subjects of worrying about screen navigation. However, one subject preferred to do the navigation.

Two of the ten subjects were able to create statements, but were confused about how to create general rules and did not complete the test. The remaining eight created an average of 12 statements and explained them with general rules. It took about 1-1.5 hours to produce these statements, as subjects were still getting used to the methodology. Four of these eight found the process doable, but tedious and difficult. They were happy to do the minimum and finish, each creating about five statements. However, the other four subjects said that the program was cool and "nerdy fun." They produced an average of 20 statements, with the maximum of 25. All four voluntarily continued until all frames were annotated. Two of them were motivated by the points awarded to each statement, and one in particular asked what others had done and made sure to double it.

5.2 Overview of the Collected Data

Figure 28 shows the statements from the first six frames created by User#4, one of the more prolific subjects. The statements are rich in detail, capturing intention, emotion, location,

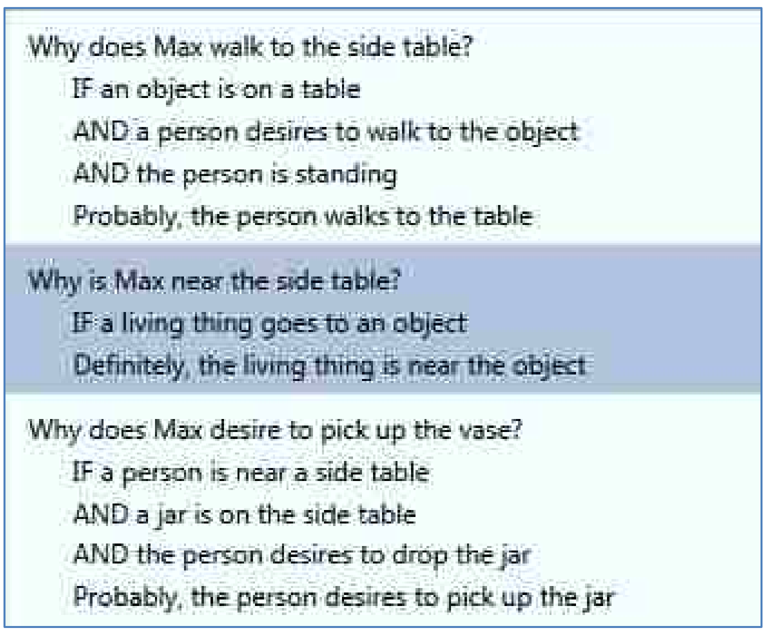
1. Opening setting
    - It is daytime
    - Max is in the living room
    - Max is on the sofa
    - Max is sitting
    - Max is inactive
    - Max is bored
    - The vase is in the living room
    - The vase is on the side table
    - Max is naughty
    - The side table is on the floor
2. Max looks around the living room
    - Max sees the vase
    - Max desires to be having fun
    - Max desires to break the vase
    - Max desires to drop the vase
    - Max desires to walk to the vase
3. Max stands up
    - Max is standing
    - Max is next to the sofa
4. Max walks to the side table
    - Max is near the side table
    - Max desires to pick up the vase
5. Max grasps the vase
    - Max is in contact with the vase
    - The arms are in contact with the vase
6. Max picks up the vase
    - The vase is not on the side table
    - The vase is over the side table
    - Max desires to turn with the vase
    - Max desires to turn far from the side table

Figure 28: User annotations from the first six frames. The numbered statements are key frame captions, representing actions. The statements below each caption are the stative effects of the action.

and movement. Each of the numbered statements is a key frame caption. Except for the opening setting, the captions are actions and the statements below them are states that result from those actions.

All of the captions and statements were created by the user. During training on the first three key frames, I gave hints on what to do. Figure 29, which shows the explanations for the statements in key frame 4, represents the type of rules that users created on their own.

The first line in Figure 29 begins an explanation of the caption of key frame 4: "Why does Max walk to the side table?" The explanation comprises three previous statements: The vase is on the side table, Max desires to walk to the vase, and Max is standing. In the rule, these three statements are generalized. That is, *object* is a generalization of *vase*, *person* stands for *Max*, and *table* stands for *side table*.


Why does Max walk to the side table?
IF an object is on a table
AND a person desires to walk to the object
AND the person is standing
Probably, the person walks to the table

Why is Max near the side table?
IF a living thing goes to an object
Definitely, the living thing is near the object

Why does Max desire to pick up the vase?
IF a person is near a side table
AND a jar is on the side table
AND the person desires to drop the jar
Probably, the person desires to pick up the jar

Figure 29: Rules created to explain the statements in key frame 4

The second block in Figure 29 answers the question, "Why is Max is near the side table?" The explanation comprises just one previous statement, the caption itself: Max walks to the side table. It is generalized so that *living thing* now stands for *Max*, *goes* stands for *walks*, and *object* stands for *side table*.

The third block in Figure 29 answers the question "Why does Max desire to pick up the vase?" The explanation comprises generalizations of three previous statements: Max is near the side table, the vase (generalized to *jar*) is on the side table, and Max desires to pick up the vase.

Each of the three rules in Figure 29 is excellent; they adequately explain the corresponding frame statement, and they are nicely generalized so that they may apply to many situations. The first and third rules are justified in having a confidence of *probably*, because they assert what usually one would expect, given the statements in the antecedent. In contrast, the second rule has a confidence of *definitely*, which is appropriate because in the context of a typical narrative directly after a living thing goes to some object, then an NLP application would almost always be correct in subsequently placing the living thing near that object.

5.3 Data Analysis

Appendix B lists all the data for the 8 test subjects that completed the evaluation. The subjects are labeled User#1 thru User#8. The two subjects that did not complete the test have no data and are not listed. The data shows the user label followed by data for each of the key frames, including those used for training, which are marked as such. Except for User#7, the first three key frames were for training purposes. User#7 requested two additional key frames of individual training. Although the data produced during training are not included in the overall statistics, these data are nevertheless listed in Appendix B because users were allowed to create their own interpretations as they trained, and these interpretations affected the subsequent data.

5.3.1 Differences in Interpretations

HXP welcomes different interpretations of the actions. For this test, most of the users' interpretations assume that Max simply wants to have fun when he drops the vase. However, in the interpretation of User#5, Max drops the vase because it is too heavy and uncomfortable to hold. In contrast, according to User#3 and User#4, Max's actions are because he is a naughty boy. User#3 even added additional details that Mommy is standing in the back yard and angry; Max is also angry and desires to break the vase because he thinks Mommy likes the vase. None of these different interpretations should be considered the most accurate. As long as there are good explanations, it is all valuable data.

5.3.2 Data Quality (Precision)

The users had been asked to produce rules with a confidence of at least *probably* if they could come up with a justified explanation, and all the subjects tried hard to do so. Of the 106 total rules, 96 (90%) have a confidence of either *probably* or *definitely*.[25]

To test the quality of the rules, I used a panel of three judges, consisting of myself and two of the more enthusiastic test subjects (User#2 and User#4). Each judge independently rated all the rules as either *Acceptable* or *Unacceptable*. An acceptable rule was one that judges felt was a generally true statement, independent of a specific story. An unacceptable rule had the confidence level too high or was missing at least one critical explanation. Rule(15) below is unacceptable because the object could be something that does not normally break when dropped. The rule would be acceptable with an additional condition that the object is breakable.[26] Rule (16) is unacceptable because, just because a person desires to do something, we cannot say that the person will <u>definitely</u> do it.[27]The rule would be acceptable if the confidence was changed to *probably*.

[25] There are many rules with *possibly* in the first key frame to establish the opening setting, but these occurred during the training session and are not counted here.
[26] The rule would also be acceptable if the confidence was changed to *possibly*, but the goal is to create rules with a confidence of at least *probably*.
[27] One of the test subjects suggested this sensible guideline: an agent's intentions or desires are never enough to warrant a confidence level of *definitely* because goals can be blocked in many ways.

15) IF a person picks up an object
AND the object falls
THEN probably the object breaks

16) IF a person desires to walk to an object
AND the object is on top of a table
THEN definitely the person walks to the table

The judges found 83 (86%) of the 96 rules to be acceptable. The opinions were unanimous on 75% of the rules, and we took the majority opinion on the remaining 25%.

In a few cases, there was a strong dissent to the majority opinion. However, the judges discussed the different opinions and came to a decision. For example, in Rule 7b of User#10, there was a difference of opinion about what it means to "turn oneself." Paraphrasing the rule, a person who turns himself *definitely* ends up facing a different direction. Two judges felt that *definitely* was too strong because a person could turn his torso and not turn his head; therefore, he could continue facing the same direction. The other judge felt that turning oneself always means turning to face a different direction. After some discussion, the majority ruled. As another example of dissent, rule 7b of User#2 says that if an object is made of ceramic, it is *probably* heavy. One judge disagreed, bringing up the possibility of a tiny ceramic figurine. Nevertheless, the other two judges overruled on the grounds that, in the normal case a ceramic object is *probably* relatively heavy. The decision turns on whether tiny ceramic figurines are "normal" when one thinks of ceramic objects, and this is highly subjective. As mentioned earlier, HXP is intended to be a wiki project, which means there is a good way to arrive at a consensus on such matters. Since users will be viewing and discussing the rules, it is important that the rules be easy to read and understand. The fact that non-expert judges were able to discuss the rules without having to decipher them first suggests that the rules are easy enough to understand.

In the field of information extraction, *precision* measures the relevancy of extracted data.[28] The 86% acceptability rating here could be viewed as analogous to precision. The companion measurement of *recall* is considered in the next section.

5.3.3 Data Coverage (Recall)

Parallel to information data precision is *recall,*[29] which in this case, would evaluate whether I captured all the data that I intended to capture. One way to measure this objectively is to compare the collected data with the sample annotations that I proposed to collect before implementing this project. Appendix A shows the data that I original proposed to collect. It shows my interpretation of “Max breaks the vase” annotated with details about intention, sensory perception, location, and movement. A significant change between this proposed data and the data actually collected is the presentation of key frames. Originally, I had proposed to show a hierarchy of panels instead of the flat, non-hierarchical presentation of key frames. The next subsection describes this change in presentation. Having explained the correspondence between the panels of Appendix A and key frames, I will compute a recall score for each user by comparing the user’s data against the data in Appendix A.

Flat Organization of Key Frames

Figure 1 in Chapter 1 displays the key frames of the sample scene laid out like a film strip. This flat representation is excellent for showing how the action unfolds, and HXP users can “animate” an experience simply by quickly clicking between key frames. Unfortunately, the flat view does not capture commonsense intuitions about how people linguistically conceptualize events. For example, if a narrative has an action, there is often an implied hierarchy of actions

[28] The meaning of *precision* used here should not be confused with the scientific or engineering sense that applies to accuracy of a measurement.
[29] The meaning of *recall* used here should not be confused with the psychological sense that involves memory retrieval.

rather than just the single action. For example as we will see, people could use the same statement "Max picks up the vase" to refer to different levels of the hierarchy.

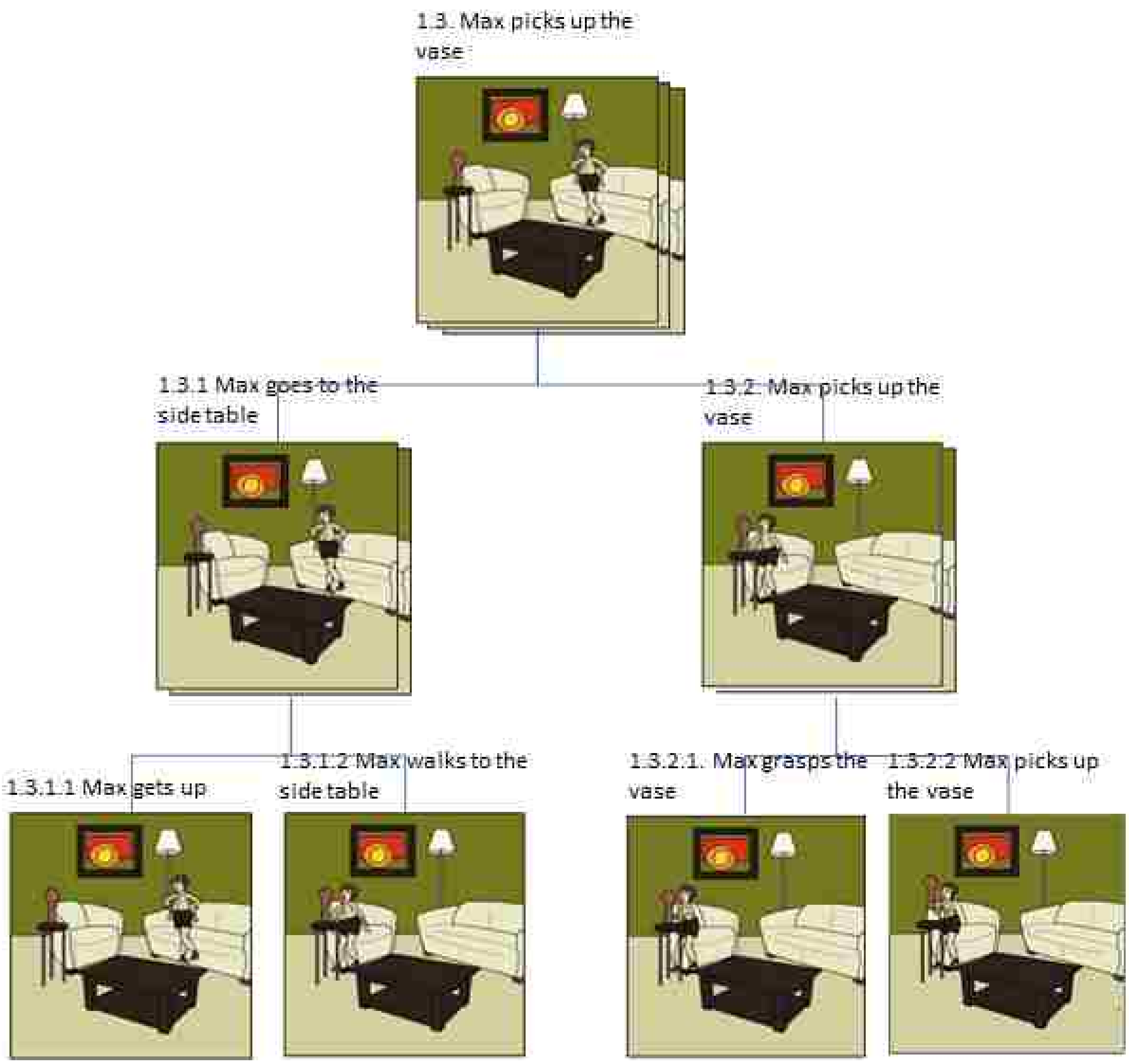


Figure 30: Hierarchical view of "Max Breaks the Vase." Originally, users were to create this hierarchy, but they found this task too difficult and abstract.

Figure 30 shows an example of this type of hierarchy. At the top of this hierarchy is "1.3 Max picks up the vase," which has two implied subactions. The first is "1.3.1 Max goes to the side table," and the second is "1.3.2 Max picks up the vase." In other words, if a person is next to a couch and desires to pick up a vase, we assume that the person will walk to the vase first as a precondition for picking it up. In a normal narrative or day-to-day conversation, this assumption is not articulated; nevertheless, it is an important precondition to the action.

At the second level of Figure 30, we see “1.3.2 Max picks up the vase.” In this conception of the action, there are two very granular actions, one of which is also captioned “Max picks up the vase” but marked with “1.3.2.2.” It is this most granular conception of the action that matches the semantics of the verb phrase *pick up – “to take and lift upward.”* Of course there are probably infinite ways to conceptualize an action into a hierarchy, and the goal was to collect as many reasonable commonsense intuitions as possible about implied action/subaction relationships.

However, early user tests indicated that the hierarchical view was too abstract and difficult to work with. The multiple levels of key frames shown in Appendix B were confusing to users. They did not understand that only “leaf” key frames, the ones at the lowest level, were to have state annotations. For example, only 1.3.2.1 Max picks up the vase has annotations other than a caption, but users had a difficult time understanding what type of annotations belong at which level. To remove this understanding barrier, I removed the hierarchical view altogether. However, I believe it may be possible to infer hierarchies of actions from the data, as discussed in 6.4 Action/Subaction Relationships and Event Semantics.

Measurement of Recall

To measure how well users performed against the statements proposed in Appendix A, the statements were first grouped according to whether they represent intention, location, movement or position, or sensory perception. The Appendix A statements, arranged by key frames and statement categories, are shown in Table 1. Next, the data provided by each user were compared against the data in this table. Since the first three key frames of “Max breaks the vase" were for training, only key frames 4-10 are considered. These seven key frames correspond to Appendix A, panels 1.3.1.2, 1.3.2.1, 1.3.2.2, 1.4.1, 1.4.2, 1.4.3, and 1.4.4.

Table 1: Appendix A statements, arranged by key frame and statement category. Since the first three key frames were for training, only key frames 4-10 are considered. Each column lists the key frame number with its corresponding hierarchical panel number from Appendix A. Each row shows the category of the key frame's statement.

	Key Frame 4 Panel 1.3.1.2 **Max walks to the side table**	Key Frame 5 Panel 1.3.2.1 **Max grasps the vase**	Key Frame 6 Panel 1.3.2.2 **Max picks up the vase**	Key Frame 7 Panel 1.4.1 **Max turns is body**	Key Frame 8 Panel 1.4.2 **Max drops the vase**	Key Frame 9 Panel 1.4.3 **The vase falls**	Key Frame 4 Panel 1.4.4 **The vase breaks**
Intention	Max intends to pick up the vase		Max decides to break the vase	Max expects the vase to fall, Max is excited			Max is having fun
Location	Max is at the side table; The side table is in front of Max		The vase is above the table	The vase is above the floor	Then vase is not supported	The vase is lower	The vase is on the floor; The pieces are on the floor
Movement or Position	Max is standing	Max is holding the vase	Max's arms are raised.	Max is facing away from the side table; The vase is above the floor; Max's torso is twisted	Max's hands are open		The vase is broken
Sensory		Max thinks the vase feels smooth. Max thinks the vase feels cool	Max thinks the vase is heavy				The vase makes a loud crash

In Table 1, the key frame headings are shown with the hierarchical numbering from Appendix A, as well as the key frame caption. For example, the first column is labeled as key frame 4, which corresponds to panel 1.3.1.2 in Appendix A. The caption is "Max walks to the side table." The Intention category has the statement "Max intends to pick up the vase."[30] The Location category has two statements, "Max is at the side table" and "The side table is in front of Max." The Movement or Position category has "Max is standing."[31] The final category of Sensory has no statements for this key frame.

The next column of Table 1 is from key frame 5, which corresponds to Appendix A panel 1.3.2.1 "Max grasps the vase." Appendix A shows no changes in the categories of Intention or

[30] This statement actually occurs in a previous key frame of Appendix A, but the statement still holds.
[31] *Ibid*.

Location, so these are left blank in the table. However, in the Movement or Position category, it has a statement about Max holding the vase. Also, in the Sensory category, it captures Max's perception of feeling the smooth, cool ceramic. It is important to note that the relevance of Max's sensory perception depends on the Appendix A interpretation of why Max desires to pick up the vase. Max is curious about its texture and weight, and so the smoothness and weight are noted. However, for other interpretations, it would not be necessary to capture this information. Therefore, when comparing the user data to the data in Table 1, we must take into account whether the data is necessary for the user's interpretation. On the other hand, I believe some sensory perceptions, such as seeing the vase or hearing the crash, are more critical than others because they are part of the general experience of dropping a breakable vase. The user must see the vase to notice it; the word "CRASH" is specifically shown in the key frame's image. Also, changes in location or position are almost always considered necessary since these are critical to commonsense understanding.

Table 2: Statements taken from User #1 in Appendix B. Key frame 4 has an intentional statement and positional statement, but it has no annotation about Max's location. Key frame 5 has statement about movement/position, but it is missing a statement about sensory perception. However, the user's interpretation did not need to annotate this perception, so the data is marked *Not necessary*. Key frame 6 is missing statements in all four categories, but two are not necessary. Since the user stopped after key frame 6, the rest of the key frame comparisons are NA – Not applicable.

	Key Frame 4 Panel 1.3.1.2 **Max walks to the side table**	Key Frame 5 Panel 1.3.2.1 **Max grasps the vase**	Key Frame 6 Panel 1.3.2.2 **Max picks up the vase**	Key Frame 7 Panel 1.4.1 **Max turns is body**	Key Frame 8 Panel 1.4.2 **Max drops the vase**	Key Frame 9 Panel 1.4.3 **The vase falls**	Key Frame 4 Panel 1.4.4 **The vase breaks**
Intention	Max desires to disturb the vase		Not necessary	NA			NA
Location	MISSING		MISSING	NA	NA	NA	NA
Movement or Position	Max is standing	Max is in contact with the vase	MISSING	NA	NA		NA
Sensory		Not necessary	Not necessary				NA

We will consider the data in Table 1 to be the gold standard by which user performance will be measured. We compute a recall score by categorizing user statements in the same way and then comparing them against Table 1. The categorized user data are shown in the tables at the end of Appendix B. As an example of how recall was measured, let us examine the data for User#1, taken from Appendix B and shown above as Table 2 for convenience.

In key frame 4 of Table 2, the user has statements in two of the three expected categories. That is, for Intention the statement is "Max desires to disturb the vase." For Location the statement is MISSING. For Movement or Position, the statement is "Max is standing."[32] However, as Table 1 shows, the user was expected to annotate Max's location after having walked to the vase. Since the user did not annotate a new location, this information is marked as MISSING for the recall measurement.

In key frame 5 of Table 2, the user has a statement in the Movement or Position category, but the user is missing a statement in the Sensory category that would describe what Max feels when he grasps the vase. However, since the user had no special interpretation involving the texture of the vase, this was marked *Not necessary*, which means the category will not be counted for this key frame.

In key frame 6, the user is missing statements in all four categories. That is, the user has failed to capture Max's excitement or any intentional data after having picked up the vase. Also, the user has not captured the fact that the vase's location has changed, nor the fact that Max's position has changed, nor the sensory data about the weight of the vase. However, in this case, Max was not experiencing any new intentional data, and the weight of the vase is irrelevant. Therefore the Intention and Sensory categories are marked *Not necessary*. In contrast, the

[32] The data for User#1 in Appendix B actually shows that "Max desires to disturb the vase" is in a previous key frame. Nevertheless, this intention continues to hold in key frame 4, so it was not necessary to repeat it.

Location and Movement or Position categories are considered critical, and they are marked *MISSING*. Finally, in key frames 7 – 10, all of the data is marked *NA* – Not Applicable because the user did not attempt to annotate these key frames.

Having compared Table 2 against Table 1, we can now measure the recall for User#1. User#1 provided 3 out of the 6 critical types of statements expected in key frames 4-6, for a recall score of 50%. The data for key frames 7-10 are not counted since the user did not attempt to annotate these frames. Thus, User#1's recall score is 50%.

The recall data for the rest of the users are shown in the tables at the end of Appendix B. The average recall score is 60/90 (67%), with half the users clustering at 50% and the other half clustering at 80%.

5.4 Confusion With Actions Vs. States

The HXP interface allows users to choose either actions or states when creating a caption. This leads to confusion about whether a caption describes an action or state. For example, the user may choose *vase* for the object and then choose the verb *fall.* HXP displays this as an action "The vase falls." But the user may understand it as a state, equivalent to "The vase is falling." It was difficult to explain the difference between states and actions to users. Indeed, during the user test, key frame 9 "The vase falls" was ambiguous because it was not clear whether the vase completes the falling action during the key frame or whether it was in a state where it was continuing the fall. The sample in Appendix A shows that the vase is continuing to fall; User #2 assumed this to be the case. However, other users were confused, and thought the vase should have completed the action.

To clarify the interpretation of actions and states, I will change the structure of key frames so that HXP assumes the statement in the caption is an action that begins and ends in the key frame. Furthermore, it will assume all other statements are somehow related to the action. So if the user is creating the caption, *fall* would be displayed as “The vase falls.” If the user is creating a statement in the body of the key frame to describe the vase, it would be displayed as “The vase is falling.” During the user test, I proposed that users interpret the statements in this way (even though I had not implemented the change in HXP), and they were able to interpret statements more easily.

Another point of confusion for users was how to break up a narrative into key frames. In the original version of “Max breaks the vase” as presented in Appendix A, there is a key frame for “Max looks around for something to do” and a separate key frame for “Max notices the vase.” However, it is not clear whether noticing the vase is a state or action. By adding the constraint that the key frame represents an outwardly observable action, we eliminate confusion about when to create key frames, and we provide a clear relationship between a frame image and its caption. [33]

The well-defined structure that places actions only in the key frame’s caption and effects only in the key frame’s body is less confusing to users, but it also has other benefits. First, whenever an object is part of a caption, the system can automatically remind the user to check if the object has changed. Second, when a user provides an explanation for a statement, the system can verify that the statement is an effect, either directly or indirectly, of the key frame’s caption. These reminders and constraints will help the user decide what to do next and make it easier to choose sensible explanations.

[33] If the action must be observable, then we will have to figure out how to represent actions that are normally not observable, like swallowing food.

5.6 Summary

In a ten person user evaluation, eight people were able to contribute high quality, detailed data using this methodology. Of those eight, half found the process tedious and difficult, but half found it to be challenging and fun. Within the limitations of such a small number of subjects, and of the inherent biases associated with using acquaintances, I am encouraged by the results that 86% of the user-built rules were acceptable, particularly because I purposely did not choose computer science majors or AI fans. I am also very pleased that, of the eight that contributed, four were able to provide more than 80% of the types of data that I had proposed to collect, especially since my proposed data is arguably an upper limit on what a non-expert could be expected to provide. Finally and crucially, users could easily understand the annotations of others and found it enjoyable to identify others' problems. Thus in the future, users should be able to build on each other's ideas.

It took from one to two hours to train the users in the first three key frames, and about 1.5 hours to complete the annotations of the remaining 7 key frames. However, the users got faster as they gained more experience. Clearly it is unrealistic to think that we can recruit armies of casual volunteers to use this framework for collecting experiential data. But with improvements in training and a wiki format where annotators view and discuss each other's work, I believe we may be able to tap into that small percentage of the vast web population that would enjoy collaborating on this important, but difficult, task.

CHAPTER 6: FUTURE WORK

HXP can be extended in many directions for NLP and other related research areas. This chapter first outlines a series of tasks, starting with those with the highest priority, that will take the current work from a proof-of-concept to fully functioning wiki collaboration. Then it discusses several possible follow-up research projects. These involve implementing a question-answering module, integrating HXP with other NLP projects, exploring action/subaction relationships with event semantics, and applying the HXP data and methodology to other areas, including models of cognition.

6.1 Planned Tasks

6.1.1 Improvements to the User Interface

As a result of the user test, several places where the user interface could be improved were noted. The system should explicitly prompt for the intentions or goals of each character. The system should also remind the user to check what states have changed from one key frame to the next. Finally, the system should prompt for sensory states such as seeing the vase, touching the surface of the vase, and hearing the loud crash. Finally, as discussed in Chapter 4, the system should allow users to add new predicates and predicate categories.

Also discussed in Chapter 4, to make it easier for users to choose action words and to identify patterns in the data, we want to reduce the verb vocabulary. One way to reduce the verbs is to use Elson's strategy (2012) of selecting from the pool of verbs in VerbNet. Another would simply be to manually filter out verb meanings in WordNet that are so similar that users would be expected to have difficulty choosing between them. I estimate this filtering effort would take me about two weeks.

6.1.2 More Expressive Input

I have successfully implemented a proof-of-concept system to represent simple experiences. However, many more structures are needed to represent more complex scenarios. At minimum, HXP needs adverb representation because even simple situations need to express the degree of an action. Likewise, it needs simple dialog so that there can be some communication among characters. Later, it needs predicates about social relationships, comparisons and time durations. It also needs to be able represent abstract concepts, group behavior and simple quantification. Furthermore, to work with larger narratives, there needs to be a representation for repetitive actions.

6.1.3 More User Experiments

Once the software is more fully implemented with the changes discussed above, there should be more extensive user tests taking into account user background, education level, age, and interests. The goal is to predict the population of users that would most likely use the system. Non-native English speakers should be included in the test since English fluency is not a requirement to use the system.

6.1.4 Cultural Biases

Since we want to be able to annotate children's stories, at minimum HXP should add more annotation categories to identify objects and situations that are magical. Thus, the general rules can have qualifications like: "If a snowman is magical, possibly it can fly" or "If a rabbit is magical, possibly it can talk." To address inherent cultural biases, we might want to include predicates for culture or geographic location.

6.1.5 Inference

As discussed in Chapter 3, I have implemented a simple single-step forward chaining inference mechanism to predict what else might be true from a given statement. The next logical addition would be to add backward chaining to fill in the gap between what is stated and what might have been the reason behind the statement. This would make the system less tedious to use since it could fill in explanations automatically if there is already a relevant rule.

In the longer term, more work is needed, even with single step inference, to make rule handling match user intuitions. For example, consider the following two rules about eating cake:

17) IF a person eats cake
    AND the person is full
    THEN Probably the person is happy

18) IF a person is full
    AND the person eats cake
    THEN Probably the person is unhappy

The natural language interpretation of these statements has an implicit order. In (1), the person first eats cake and then is full. In (2), the person is already full before eating the cake, causing unhappiness. Currently, the inference procedure does not take into account the implicit order of the statements, so both rules would match if Max eats cake on a full stomach. Neither does the inference procedure take into account intervening steps. We can imagine a rule about playing with the lights:

19) IF a child turns on a light
    And the child turns off the light
    And the child turns on the light
    Possibly the child is playing

However, in a longer scene where Max turns on the light, reads for a while, turns off the light to go to bed, and then turns it on when he gets up in the middle of the night, he is not playing. When users create rules with multiple actions like (19), there may be an implicit natural language interpretation that the actions occur in sequence. Once we have longer scenes, we will have to address this issue and probably others that involve unintended inferences.

6.1.6 Data Format

HXP narratives are currently exportable to a standard extensible mark-up language (XML) file. Nevertheless, the data format is not portable because HXP elements refer to internal database entries which are not standard. There needs to be a portable format for HXP-defined concepts and a versioning system so that XML narrative files either have self-contained data or references to standard data. It will also be desirable to provide a mapping between the statements of HXP and the XML elements in a standard language like ISO-Space (Pustejovsky et al., 2011).

6.1.7 Web Implementation

To take the project from proof-of-concept to fully functioning wiki collaboration, the software should be implemented as a thin client (i.e., accessible via an Internet server) with a host of features to make it more like a social networking or education web site where people comment on each other's work and make suggestions. Similar to a software development environment, there are issues with keeping track of modifications. That is, users may refer to rules in their narratives, and other users may make modifications to those rules, and the system needs to be able to verify that the modified rules still apply to existing references in other narratives.

6.1.8 Motivating Users

A significant future challenge will be to motivate workers to participate in this project. In the small user test, some users were definitely motivated by the competition to get more points by adding more detail. They also had fun looking at other people's work and making comments and refinements.

I suspect users would be motivated by a series of small, reachable goals. We could start with an annotated corpus of experiences that correspond to the children's stories from the initiative of (McCarthy et al., 2002). We could also create a challenge to annotate a set of experiences related to a common theme like getting something off a high shelf, similar to the sample commonsense scenario suggested by (Minsky et al., 2004).

6.1.9 Integration with a 3D system

In the more distant future, I envision integration with a 3D environment like Story-Understanding Alice (Kelleher & Pausch, 2007) or WordsEye (Coyne & Sproat, 2001). As discussed in (Weltman, 2009), 3D environments implicitly represent commonsense knowledge about objective locations and movement. Integrating HXP with a 3D system would offer a valuable mapping between objective data grounded in a virtual environment and subjective intentional descriptions.

6.2 A Question-Answering Module

The HXP methodology produces detailed, coherent narratives, but it captures only commonsense detail that is relevant to the specific experience. It does not capture detail that is important for understanding the experience but not needed to explain an action. For instance, users may not be inclined to annotate the fact that Max faces forward, not backward, when he walks to the vase; that Max is supported by the floor; that the vase is immobile, not spinning, etc. These states are true by default, but an NLP system would not necessarily know the default rules. One way to get at default information would be to add a Question Answering module to HXP that allows users to ask off-topic, arbitrary questions about an experience: Is Max floating in the room? Is Max wearing goggles? If the answer is wrong or unavailable, users could answer the question and then create a rule to explain the answer.

Fortunately, once rules are in place to answer arbitrary questions about one experience, they can be used to answer similar questions in the future. Thus, gradually, HXP would be able to provide reasonable answers to many arbitrary questions about an experience, as we would hope from a story-understanding program.

Related to the problem of default rules is the problem of dealing with what-if questions: What if Max drops the vase on the couch instead of the floor? What if Max drops a wooden ashtray rather than the vase? What if Mommy comes into the room before Max is able to pick up the vase? These alternative scenarios would provide a lot of commonsense information.

6.3 HXP and Other NLP projects

I believe the HXP methodology would integrate with deep semantic architectures such as Scone (Fahlman, 2011). The Scone project models how cognition grows with gradually available knowledge. It represents mental states as descriptions within contexts, somewhat similar to how HXP represents states within a context of a key frame. One of the open problems in Scone is how to make it easier for users to add new knowledge, so the HXP methodology could feed directly into Scone's knowledge bank. (Schubert, 2006) outlines several architectures and strategies to achieve human-level reasoning. All of these systems would benefit from the type of experiential data proposed for HXP. Finally, once HXP has collected a very large number of experiences, perhaps on the order of ten thousand, statistical NLP applications and tools could make use of the associations between actions, intentions and locations.

6.4 Action/Subaction Relationships and Event Semantics

As described in Chapter 5, my original proposal was to collect information on actions and subactions by having users arrange key frames into a hierarchy of actions, but users found it too difficult to understand and interpret the different levels of annotations.

My plan now is for the system to infer action/subaction relationships from key frames whose captions have similar explanations. For example, suppose key frame 3 “Max stands up” and key frame 4 “Max walks to the side table” both include the explanation, “Max desires to touch the vase.” HXP could infer that the two key frames are subactions of a larger goal to touch the vase. Thus, the system could automatically construct a key frame hierarchy, with “Max touches the vase” at the top, and with two sub items, “Max stands up” and “Max walks to the side table.”

This type of hierarchical analysis of events may shed new light on event semantics. For example, according to the classic event typology of (Vendler, 1957), an *accomplishment* is an event that unfolds gradually before it ends (e.g. bake a cake, pick up a vase), while an *achievement* is an event that happens instantaneously (e.g. notice, explode). In a situation where Max is not at a vase’s location and has to walk to it, the event described by “Max touches the vase” is clearly an *accomplishment*. We could refer to the ongoing process of Max’s getting up to go to the vase and reaching out to touch it as “Max is touching the vase.” However, in the situation where Max’s finger is right next to the vase and then touches it, Max makes contact in an instance, which is an achievement. As is common with achievements, the instantaneous act of touching, as a process, evokes repeated touching movements,[34] as in “Stop touching that vase!” Thus, the same action statement “Max touches the vase” could denote either the gradually

[34] *Touch* could also be interpreted statively, as in “The bookcase touches the ceiling,” but we are referring here to the touching action here.

unfolding event or the instantaneous event, even though the action *touch* means “make contact with” in both cases. In other words, the accomplishment lexicalized as “Max touches the vase” has a subevent which is an achievement that can be expressed lexically with the very same phrase, “Max touches the vase.” It may be helpful to consider a contrasting example. An accomplishment like “Max bakes the cake” does not seem to have a culminating achievement subevent that could be lexicalized as “Max bakes a cake.” Rather, subevents that indicate the culmination of the baking event might be “Max closes the oven door” or “Max removes the cake from the oven.” In fact, it is quite difficult to pin down a single culminating endpoint for the baking accomplishment. These differences suggest that the two events, although both accomplishments, could be in different classes. In sum, using the HXP methodology, we could examine the lexicalization of events with respect to their implied actions and subactions.

6.5 Specialized Knowledge Domains

HXP currently aims to collect common sense that we learn from simple childhood experiences, so it does not target the type of knowledge found in specialized knowledge domains. However, I believe that the same software and methodology could have special topics. For example, there could be a special section aimed at health care professionals that focuses on gathering knowledge about simple health care experiences. For specialized domains, users would need to create new predicates, similar to the Disciple’s domain-specific implementations (Tecuci et al. 2005; Tecuci et al., 2008), as discussed in the case-based reasoning systems of Chapter 2. It would be interesting to see if the HXP methodology could be applied to a complex domain like military center of gravity analysis, and if so, would the resulting data be different from the type produced by the Disciple process (Tecuci et al., 2008). Clearly, to work with global warfare, the scale of the experiential model would have to be adapted. Instead of a living room for a background, perhaps it would be a map. Instead of individual characters, perhaps there would be army tokens, like a board game.

6.6 Models of Cognition

Hxp scenes consist of agents, actions, objects, mental states, and background setting – the same properties that cognitive scientists use to model cognition in the human brain (Krueger et al., 2009). In these models, the brain abstracts from concrete experiences as it performs essential cognitive tasks such as planning and interpreting actions. If concrete experiences and abstractions of experiences are critical to cognitive processing, then HXP could be a fundamentally new type of resource for general models of human cognition.

CONCLUSION

The goal of the Human Experience Project is to collect highly structured life experiences in order to help programs associate words and phrases with a larger situational context. AI researchers have long recognized the importance of using narrative structures for natural language processing. However, attempts to narrow the problem to artificial worlds or specific domains (e.g. eating at a restaurant) do not lead to more general AI capabilities. Furthermore, attempts to use non-experts to provide simple stories from which commonsense can be extracted have also failed because it is difficult for non-experts to articulate knowledge that is obvious to people but not to machines. HXP structures scenes into small time slices, guiding annotators to describe each frame, with particular focus on intent, location, and movement. Furthermore, it applies an automated Socratic Method to draw out hidden assumptions that humans make about common situations. I implemented a proof of concept and conducted a small user evaluation of this HXP methodology. The results suggest that non-experts are able to create the high quality experiential data that I proposed.

HXP users create two types of narrative data: narrative statements model an experience at a particular time and place; commonsense rules model how people explain the narrative and generalize situations. Other researchers have proposed similar narrative data, but HXP's formalisms open up the possibility of non-experts creating these data, a key first step in creating a large-scale effort to collect experiential narratives for both statistically-oriented applications and deep semantic processing.

REFERENCES

Aamodt, A., & Plaza, E. (1994). Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications, 7*(1), 39-59.

Bacon, W. F. (1995). What everyone knows about attention. Representing Mental States & Mechanisms. In M. T. Cox, & M. Freed (Ed.), *Papers from the 1995 AAAI Symposium.* Technical Report SS-95-05.

Baker, S. (2011). *Final Jeopardy: Man vs. Machine and the Quest to Know Everything.* New York: Houghton Mifflin Harcourt.

Chambers, N., & Jurafsky, D. (2009). Unsupervised Learning of Narrative Schemas and Their Participants. *Proceedings of the Association of Computational Linguistics (ACL).* Columbus, Ohio.

Chambers, N., & Jurafsky, D. (2011). Template-Based Information Extraction without the Templates. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT) - Volume 1* (pp. 976-986). Stroudsburg, PA: ACL.

Charniak, E. (1972). *Toward a model of children's story comprehension.* Technical Report AITR-266. Cambridge, MA: Artificiial Intelligence Laboratory, Massachusetts Institute of Technology.

Coyne, B., & Sproat, R. (2001). WordsEye: An Automatic Text-to-Scene Conversion System. *SIGGRAPH Proceedings.* Los Angeles, CA.

Coyne, B., Bauer, D., & Rambo, O. (2011). VigNet: Grounding Language in Graphics using Frame Semantics. *ACL Workshop on Relational Models of Semantics (RELMS).* Portland, Oregon.

Dretske, F. I. (1999). *Knowledge and the Flow of Information.* Stanford, CA: CSLI.

Dreyfus, H. L. (1992). *What Computers "Still" Can't Do: A Critique of Artificial Reason, Revised edition.* Cambridg: MIT Press.

Dyer, M. G. (1983). *In-Depth Understanding: A Computer Model of Integrated Processing for Narrative Comprehension.* Cambridge, MA.: MIT Press.

Elson, D. K. (2012). *Modeling Narrative Discourse.* PhD Thesis, Columbia University, New York City.

Elson, D., & McKeown, K. R. (2010). Building a Bank of Semantically Encoded Narratives. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010).* Malta.

Fahlman, S. E. (2011). Using Scone's Multiple-Context Mechanism to Emulate Human-Like Reasoning. *Proceedings of the AAAI Fall Symposium on Advances in Cognitive Systems.* Arlington, VA.

Fan, J., Kalyanpur, A., Murdock , J. W., & Boguraev, B. K. (2011). Mining Knowledge from Large Corpora for Type Coercion in Question Answering. *Web Scale Knowledge Extraction (WEKEX) Workshop at International Semantic Web Conference.* Bonn, Germany.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database.* Cambridge, MA.: MIT Press.

Fillmore, C. J., Petruck, M. R., Ruppenhofer, J., & Wright, A. (2003). FrameNet in Action: The Case of Attaching. *International Journal of Lexicography, 16*(3), 297-332.

Fuchs, N. E., & Schwitter, R. (1996). Attempto Controlled English (ACE). *Proceedings of the First International Workshop on Controlled Language Applications (CLAW).* Leuven, Belgium.

Fuchs, N. E., Kaljurand, K., & Kuhn, T. (2010). *Discourse Representation Structures for ACE 6.6.* Technical Report ifi-2010.0010, University of Zurich, Department of Informatics.

Jackendoff, R. (1990). *Semantic Structures.* Cambridge, MA: MIT Press.

Kelleher, C., & Pausch, R. (2007, July). Using storytelling to motivate programming. *Communications of the ACM, 50*(7), 58-64.

Kipper, K., Korhonen, A., Ryant, N., & Palmer, M. (2006). Extensive Classifications of English Verbs. *Proceedings of the 12th EURALEX International.* Turin, Italy.

Krueger, F., Barbey, A. K., & Grafman, J. (2009, March). The medial prefrontal cortex mediates social event knowledge. *Trends in Cognitive Sciences, 13*(3), 103-109.

Kuhn, T. (2008). AceWiki: Collaborative Ontology Management. *Proceedings of the 3rd Semantic Wiki Workshop (SemWiki).* Tenerife, Spain.

Laird, J. E. (2012). *The Soar Cognitive Architecture.* Cambridge, MA: MIT Press.

Lehnert, W. G. (1978). *The Process of Question Answering: A Computer Simulation of Cognition.* Hillsdale, NJ: Lawrence Erlbaum.

Lehnert, W. G. (1981). Plot Units and Narrative Summarization. *Cognitive Science: A Multidisciplinary Journal, 5*(4), 293–331.

Lenat, D. B., Guha, R., Pittman,, K., Pratt, D., & Shepherd, M. (1990). Cyc: toward programs with common sense. *Communications of the ACM, 33*(8), 30-49.

Li, B., Appling, D. S., Lee-Urban, S., & Riedl, M. O. (2012). Learning Sociocultural Knowledge via Crowdsourced Examples. *Proceedings of the 4th AAAI Workshop on Human Computation.* Toronto, Canada.

Liu, H., & Singh, P. (2004, October). ConceptNet — A Practical Commonsense Reasoning Tool-Kit. *BT Technology Journal, 22*(4), 211-226.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: the Penn Treebank. *Journa of Computational Linguistics - Special issue on using large corpora, 9*(2), 313-330.

McCarthy, J., Minsky, M., Sloman, A., Gong, L., Lau, T., Morgenstern, L., et al. (2002). An architecture of diversity for commonsense reasoning. *IBM Systems Journal, 41*(3), 530-539.

McCloud, S. (1994). *Understanding Comics: The Invisible Art.* New York: Harper Perennial.

Minsky, M. (1975). A Framework for Representing Knowledge. In P. H. Winston (Ed.), *The Psychology of ComputerVision.* New York, NY: McGraw-Hill.

Minsky, M. (2007). *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind* . New York: Simon & Schuster.

Mueller, E. T. (1998). *Natural Language Processing with ThoughtTreasure.* New York: Signiform.

Mueller, E. T. (2000). Prospects for in-depth story understanding by computer. *CoRR*.

Mueller, E. T. (2003). Story understanding through multi-representation model construction. In G. Hirst, & S. Nirenburg (Ed.), *Text Meaning: Proceedings of the HLT-NAACL 2003 Workshop* (pp. 46-53). East Stroudsburg, PA: Association for Computational.

Mueller, E. T. (2007). Modelling Space and Time in Narratives about Restaurants. *Literary and Linguistic Computing, 22*(1), 67-84.

Nute, D. (1993). Defeasible Prolog. *Working Papers of the 1993 AAAI Fall Symposium on Automated Deduction and Nonstandard Logics.* (pp. 105-112). Menlo Park, California: AAAI Press.

Palmer, M., Kingsbury, P., & Gildea, D. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics, 31*(1), 71-106.

Pangburn, B. E. (2002). *Experienced-Based Language Acquisition:A Computational Model of Human Language Acquisition.* PhD Dissertation, Louisiana State University, Computer Science, Baton Rouge.

Power, R., Scott, D., & Evans, R. (1998). What you see is what you meant: direct knowledge editing with natural language feedback. *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI-98)*.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., et al. (2008). The Penn Discourse TreeBank 2.0. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC).* Marrakech, Morocco.

Pustejovsky, J., Hanks, P., Saurí, R., See, A., Day, D., Ferro, L., et al. (2003). The TIMEBANK Corpus. *Proceedings of Corpus Linguistics 2003*, (pp. 647–656).

Pustejovsky, J., Moszkowicz, J., & Verhagen, M. (2011). ISO-Space: The Annotation of Spatial Information in Language. *Proceedings of ISA-6: ACL-ISO International Workshop on Semantic Annotation.* Oxford, England.

Rao, A. S., & Georgeff, M. P. (1995). BDI Agents: From Theory to Practice. *Proceedings of the First International Conference on Multiagent Systems (ICMAS-95).* San Francisco, CA.

Rouhizadeh, M., Bowler, M., Sproat, R., & Coyne, B. (2011). Collecting Semantic Data by Mechanical Turk for the Lexical Knowledge Resource of a Text-to-Picture Generating System. *Proceedings of the Ninth International Conference on Computational Semantics (IWCS).* Oxford, UK.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, Plans, Goals, and Understanding.* Hillsdale, NJ: Lawrence Erlbaum.

Schubert, L. (2006). Turing's Dream and the Knowledge Challenge. *21st Nat. Conf. on Artificial Intelligence (AAAI'2006)* (pp. 1534-8). Menlo Park, CA: AAAI Press.

Singh, P. (2005). *EM-ONE: An Architecture for Reflective Commonsense Thinking.* PhD Thesis, MIT, Cambridge.

Singh, P., & Barry, B. (2003). Collecting Commonsense Experiences. Sanibel Island, FL: Proceedings of the Second International Conference on Knowledge Capture.

Singh, P., Barry, B., & Liu, H. (2004, October). Teaching machines about everyday life. *BT Technology Journal, 22*(4), 227-240.

Singh, P., Lin, T., Mueller, E. T., Lim, G., Perkins, T., & Zhu, W. L. (2002). Open Mind Common Sense: Knowledge Acquisition from the General Public. *Proceedings of the First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems.* Irvine, CA.

Smith, D. A., & Arnold, K. (2009). Learning Hierarchical Plans by Reading Simple English Narratives. *Commonsense Workshop at the International Conference on Intelligent User Interfaces (IUI-09).*

Tecuci, G., & Hieb, M. R. (1996). Teaching Intelligent Agents: the Disciple Approach. *International Journal of Human-Computer Interaction, 8*(3), 259 - 285.

Tecuci, G., Boicu, M., Boicu, C., Marcu, D., Stanescu, B., & Barbulescu, M. (2005, November). The Disciple-RKF Learning and Reasoning Agent. *Computational Intelligence, 21*(4), 462-479.

Tecuci, G., Boicu, M., Marcu, D., Barbulescu, M., Boicu, C., Le, V., et al. (2008, March). Teaching Virtual Experts for Multi-Domain Collaborative Planning. *Journal of Software, 3*(3), 38-59.

Vendler, Z. (1957). Verbs and Times. *The Philosophical Review, 66*(2), 143-60.

Von Ahn, L., Kedia, M., & Blum, M. (2006). Verbosity: A Game for Collecting Common-Sense Facts. *ACM Conference on Human Factors in Computing Systems, CHI Notes (2006)*, (pp. 75-78).

Weltman, J. (2009). Human XP: Using Virtual Worlds to Capture Common Sense. *MG 2009 Proceedings. Paper 13.* http://aisel.aisnet.org/mg2009/13.

Williams, R., Barry, B., & Singh, P. (2005). ComicKit: Acquiring story scripts using common sense feedback. *Proceedings of the 10th International Conference on Intelligent User Interfaces (IUI-05)* (pp. 302-304). San Diego, CA: ACM.

Zarri, G. P. (2010). *Representation and Management of Narrative Information : Theoretical Principles and Implementation* . London: Springer-Verlag.

APPENDIX A: SAMPLE SCENE

These annotations show the type of data I proposed to collect for HXP before implementing the project. It begins with a fragment of non-contextual, general knowledge, mostly taken from WordNet. Then it shows some examples of stock props and stock background settings. Finally, it presents the fully annotated scene "Max breaks the vase." The collected data in the user test is compared to the data in the "Max breaks the vase" scene.

General Knowledge

Red is bright
Ceramic is used for making a vase
A vase is vessel.
A vessel is a container.
A container is a man-made object
A man-made object is an object
…

A Stock Vase

The vase is red
It is made of ceramic
It is smooth
It is hard
It is breakable
It is reflective
It weighs a few pounds
It is medium sized

Stock Prop: vase

Q: Why is the vase red?
A: One of many possibilities
A vase is possibly red

Q: Why is the vase made of ceramic?
A: Ceramic is used for making a vase
IF a material is used for making an object
THEN the object is possibly made of the material

Q: Why is the vase smooth?
A: The vase is made of ceramic
IF an object is on made from ceramic
THEN the object probably is smooth

Q: Why is the vase hard?

A: The vase is made of ceramic
IF an object is made of ceramic
*THEN the object is definitely h*ard

Q: Why is the vase breakable?
A: The vase is made of ceramic
IF an object is made of ceramic
THEN the object is definitely breakable

Q: Why is the vase reflective?
A: The vase is made of ceramic
IF an object is made of ceramic
THEN the object is probably reflective

Q: Why does the vase weigh a few pounds?
A: One of many possibilities
A vase is possibly weighs a few pounds

Q: Why is the vase medium sized?
A: The vase weighs a few pounds
IF a vase weighs a few pounds
THEN the vase is probably medium sized

A Stock Living Room

Stock background: living room

The living room is medium temperature.
Four walls are part of the living room.
Wall#1 is part of the walls.
Wall#2 is part of the walls.
Wall#3 is part of the walls.
Wall#4 is part of the walls.
The living room contains a couch.
The couch is against wall#1
The living room contains a coffee table.
The coffee table is in front of the couch.
The living room contains a chair.
The chair is next to the couch.
The living room contains a side table.
The side table is next to the chair.
The living room contains a picture.
The picture is on wall#2.
It is a few feet above the floor.
It is hanging.
Wall#2 supports the picture.

Q: Why is the living room medium temperature?

A: This is typical
A living room is probably medium temperature

Q: Why are there four walls of the living room?
A: This is typical
A living room probably has four walls

Q: Why does the living room have a couch?
A: This is typical
A living room probably has a couch

Q: Why is the couch against wall#1?
A: One of many possibilities
A couch is probably against a wall

Q: Why does the living room have a coffee table?
A: This is typical
A living room probably has a coffee table

Q: Why is the coffee table in front of the couch?
A: This is typical
A coffee table is probably in front of a couch

Q: Why does the living room have a side table?
A: This is typical
A living room possibly has a side table

Q: Why does the living room have a chair?
A: This is typical
A living room possibly has a chair

Q: Why is the chair beside the couch?
A: The living room contains the chair
    AND the living room contains the couch
IF a living room contains a chair
AND the living room contains a couch
THEN the chair is possibly beside the couch

Q: Why is the side table next to the chair?
A: This is typical
A side table is probably next to a seat

Q: Why does the living room have a picture?
A: This is typical
A living room probably has a picture

Q: Why is the picture on wall#2?
A: One of many possibilities
A picture is probably on a wall

Q: Why is the picture a few feet above the floor?
A: This is typical
A picture is probably a few feet above a floor

Q: Why is the picture hanging?
A: This is typical
A picture is probably hanging

Q: Why does wall#2 support the picture?
A: The picture is on wall#2
IF a picture is on a wall
THEN the wall probably supports the picture

Scene: Max breaks the vase

Panel 1.0 Max has nothing to do


1.0 Max has nothing to do

It is daytime.
The living room is light.
The living room is at medium temperature.
The living room contains Max.
Max is on the center of the couch.
He is sitting.
The living room contains a vase.
The vase is on top of the side table.
Max is bored.
He wants to do something fun.
He does not know what to do.

Q: Why is it daytime?
A: This is assumed in the opening scene
It is possibly daytime

Q: Why is the living room light?
A: It is daytime
IF it is daytime
THEN a room is probably light

Q: Why does the living room have medium temperature?
A: It is typical
A living room probably has medium temperature

Q: Why does the living room contain Max?
A: This is assumed in the opening scene
A person is possibly in a living room

Q: Why is Max on the center of the couch?
A: This is assumed in the opening scene
A person is possibly on the center of a couch

Q: Why is Max sitting?
A: Max is on the couch
IF a person is on a seat
THEN the person is probably sitting

Q: Why does the living room contain the vase?
A: This is typical
A living possibly contains a vase

Q: Why is the vase on top of the side table?
A: This is typical
A vase is probably on top of a table

Q: Why is Max bored?
A: Max has nothing to do
IF a kid has nothing to do
THEN the kid probably is bored

Q: Why does Max want to do something fun?
A: Max bored
IF a kid is bored
THEN the kid probably wants to do something fun

Q: Why does Max not know what to do?
A: Max has nothing to do
IF a kid has nothing to do
THEN the kid probably does not know what to do

Panel 1.1 Max looks around for something to do

Max is facing the vase.
The vase is bright.
The vase is shiny.

Q: Why does look around for something to do?
A: Max wants to do something fun,
AND Max does not know what to do.

1.1 Max looks around for something to do

IF a person wants to do something fun
AND the person does not know what to do
THEN the person possibly looks around for something to do.

Q: Why is Max facing the vase?
A: Max looks around for something to do,
AND the living room contains Max,
AND the living room contains the vase.
IF a person looks around for something to do
AND a background contains the person
AND the back ground contains an object
THEN the person possibly is facing the object

Q: Why is the vase bright?
A: The vase is red,
AND red is bright,
AND the living room contains the vase,
AND the living room is light.
IF an object is a color
AND the color is bright
AND a background contains the object
AND the background is light
THEN the object is probably bright

Q: Why is the vase shiny?
A: The vase is reflective
AND the living room contains the vase
AND the living room is light
IF an object is reflective
AND a background contains the object
AND the background is light
THEN the object is probably shiny

Panel 1.2 Max notices the vase


1.2 Max notices the vase

Max thinks the vase is pretty.
He is curious about the texture of the vase.
He is curious about the weight of the vase.

Q: Why does Max notice the vase?
A: Max is looking around for something to do,
AND Max is facing a vase,
AND the vase is shiny,
AND the vase is bright.
IF a kid is looking around for something to do

AND the kid is facing an object
AND the object is shiny
AND the object is bright
THEN the kid possibly notices the object

Q: Why does Max think the vase is pretty?
A: The vase is shiny
AND the vase is bright
IF an object is shiny
AND the object is bright
THEN a kid might think the object is pretty

Q: Why is Max curious about the texture of the vase?
A: Max thinks the vase is pretty
IF a kid thinks an object is pretty
THEN the kid possibly is curious about the texture of the object

Q: Why is Max curious about the weight of the vase?
A: Max thinks the vase is pretty
IF a kid thinks an object is pretty
THEN the kid possibly is curious about the weight of the object

Panel 1.3 Max picks up the vase

The vase is on the side table.
1.3.1 Max goes to the side table
1.3.2 Max picks up the vase

1.3 Max picks up the vase

Q: Why does Max pick up the vase?
A: Max is curious about the texture of the vase
AND Max is curious about the weight of the vase
IF a kid is curious about the texture of an object
AND the kid is curious about the weight of an object
THEN the kid possibly picks up the object

Q: Why is the vase on the side table?
A: This is a restatement from a previous panel

Panel 1.3.1 Max goes to the side table

Max is a few feet from side table.
1.3.1.1 Max gets up
1.3.1.2 Max walks to the side table

Q: Why does Max go to the side table?
A: Max intends to pick up the vase
AND the vase is on the side table
IF a person intends to do something to an object
AND the object is on top of a second object
THEN the person probably goes to the second object

Q: Why is Max a few feet from the side table?
A: Max is on the center of the couch
AND the side table is next to the couch
IF a person is on the center of a couch
AND an object is next to the couch
THEN the person is probably a few feet from the object

Q: Why does Max walk to the side table?
A: Max intends to go to the side table
AND Max is a few feet from the side table
IF a person intends to go to an object
AND the person is a few feet from the object
THEN the person probably walks to the object

1.3.1.1 Max gets up

Panel 1.3.1.1 Max gets up

Max is standing

Q: Why does Max get up?
A: Max intends to go to the side table
AND Max is sitting
IF a person intends to go to an object
AND the person is sitting
THEN the person probably gets up

Q: Why is Max standing?
A: Max gets up.
IF a person gets up
THEN the person is definitely standing

1.3.1.2 Max walks to the side table

Panel 1.3.1.2 Max walks to the side table

Max is at the side table.
The side table is in front of Max.

Q: Why is Max at the side table?
A: Max walks to the side table.
IF a person walks to an object
THEN the person is probably at the object

Q: Why is the side table in front of Max?
A: Max walks to the side table.
IF a person walks to an object
*THEN the object is probably in front of the perso*n

Panel 1.3.2 Max picks up the vase

1.3.2 Max picks up the vase

1.3.2.1 Max grasps the vase
1.3.2.2 Max picks up the vase

Q: Why does Max pick up the vase?
A: This is the main subaction of the larger action of picking up the vase.

Panel 1.3.2.1 Max grasps the vase

1.3.2.1 Max grasps the vase

Max's hands are in a position of holding a medium-sized object.
Max notices that the vase is smooth.
He thinks the vase feels cool.
He thinks the vase feels nice.

Q: Why does Max grasp the vase?
A: Max intends to pick up the vase.
IF a person intends to pick up an object
THEN the person probably grasps the object

Q: Why are Max's hands in a position of holding a medium-sized object?
A: Max grasps the vase
AND the vase is medium sized
IF a person grasps a vase
AND the vase is medium sized
THEN the person's hands probably are in a position of holding a medium-sized object

Q: Why does Max notice that the vase is smooth?
A: Max is curious about the texture of the vase
AND Max's grasps the vase
AND the vase is smooth
IF a person is curious about what an object feels like
AND the person grasps the object

AND the object is smooth
THEN the person probably notices that the object is smooth

Q: Why does Max think the vase feels cool?
A: Max's grasps the vase
AND the vase is made of ceramic
AND the living room contains the vase
AND the living room is medium temperature
IF a person grasps an object
AND the object is made of ceramic
AND a background contains the object
AND the background is medium temperature
*THEN a person possibly thinks that the object feels co*ol

Q: Why does Max think the vase feels nice?
A: Max grasps the vase
AND Max notices that the vase is smooth
AND Max thinks the vase feels cool
IF a person grasps an object
AND the person notices that the object is smooth
AND the person thinks that the object feels cool
THEN the person possibly thinks that the object feels nice

Panel 1.3.2.2 Max picks up the vase

Max's arms are raised.
The vase is a view inches above the table.
Max's arms support the vase.
Max thinks the vase is heavy.
He thinks the vase is breakable.
He decides to break the vase.

1.3.2.2 Max picks up the vase

Q: Why does Max pick up the vase?
A: This is the main subaction of the larger action of picking up the vase.

Q: Why are Max's arms raised?
A: Max picks up the vase
IF a kid picks up a vase
THEN the kid's arms are probably raised

Q: Why do Max's arms support the vase?
A: Max picks up the vase
AND Max's arms are raised
IF a person picks up an object
AND the person's arms are raised

THEN the person's arms probably support the object

Q: Why does Max think the vase is heavy?
A: Max's arms support the vase
AND the vase weighs a few pounds
IF a kid's arms support an object
AND the object weighs a few pounds
THEN the kid probably thinks the object feels heavy

Q: Why is the vase a few inches above the table?
A: The vase is on the table
AND Max picks up the vase
IF an object is on the table
AND a person picks up the object
THEN the object is possibly a few inches above the table

Q: Why does Max think the vase is breakable?
A: The vase is pretty
AND the vase is heavy
AND the living room contains the object
IF an object pretty
AND the object is heavy
AND a living room contains the object
THEN a kid possibly thinks the object is breakable

Q: Why does Max decide to break the vase?
A: Max wants to do something fun
AND Max thinks the vase is heavy
AND Max thinks the vase is breakable
IF a kid wants to do something fun
AND kid thinks an object is heavy
AND kid thinks an object is breakable
THEN kid possibly decides to break the object

Panel 1.4 Max drops the vase

Max wants the vase to make a loud crash.
Max wants the vase to be high above the floor.
1.4.1 Max turns his body.
1.4.2 Max lets go of the vase.
1.4.3 The vase falls.
1.4.4 The vase breaks.

1.4 Max drops the vase

Q: Why does Max drop the vase?
A: Max decides to break the vase
IF a person decides to break an object

THEN the person possibly decides to drop the object

Q: Why does Max want the vase to make a loud crash?
A: Max wants to do something fun
AND Max intends to drop the vase
IF a kid wants to do something fun
AND the kid intends to drop an object
THEN the kid possibly wants the object to make a loud crash

Q: Why does Max want the vase to be high above the floor?
A: Max intends to drop the vase
AND Max wants the vase to make a loud crash
IF a person wants to drop an object
AND the person wants the object to make a loud crash
THEN the person probably wants the object to be high above the floor

Panel 1.4.1 Max turns his body

Max's torso is twisted.
He is facing away from the side table.
He is facing the floor.
The vase is a few feet above the floor.
Its edge is a few inches from Max's side.
The table is beside Max.
He expects the vase to fall.
He expects to have fun.
He is excited.

1.4.1 Max turns his body

Q: Why does Max turn his body?
A: Max intends to drop the vase
AND Max grasps the vase
AND Max's hands are in a position of holding a medium-sized object
AND the vase is a few inches above the table
AND Max wants the vase to be high above the floor
IF a person intends to drop an object
AND the person grasps the object
AND the person's hands are in a position of holding a medium-sized object
AND the object is a few inches above a table
AND the person desires that the object be high above the floor
THEN the person probably turns the person's body
Q: Why is Max's torso twisted?
A: Max turns his body
IF a person turns the person's body
THEN the person's torso is probably twisted

Q: Why is Max facing away from the table?
A: The vase is a few inches above the table
AND Max is holding the vase with both hands
AND Max twists his body
IF an object is a few inches above a table
AND a kid's hands are in a position of holding a medium-sized object
AND the kid twists the kid's body
THEN the kid is probably facing away from the table

Q: Why is Max facing the floor?
A: Max intends to drop the vase
IF a person intends to drop an object
THEN the person is possibly facing the floor

Q: Why is the vase a few feet above the floor?
A: The vase is a few inches above the table
AND Max's hands are in a position of holding a medium-sized object
AND Max turns his body
IF an object is a few inches above a table
AND a kid's hands are in a position of holding a medium-sized object
AND the kid twists his body
THEN the object is probably a few feet above the floor

Q: Why is the table's edge a few inches from Max's side?
A: The table's edge is a few inches from Max's chest
AND Max turns his body
IF an object's edge is a few inches from a person's chest
AND the person turns the person's body
THEN the object's edge is probably a few inches from a person's side

Q: Why is the table beside Max?
A: The table is in front of Max
AND Max turns his body
IF an object is in front of a person
AND the person turns the person's body
THEN the object is probably beside the person

Q: Why does Max expect the vase to fall?
A: Max intends to drop the vase
IF an person intends to drop an object
THEN the person probably expects the object to fall

Q: Why does Max expect to hear a loud crash?
A: Max intends to drop the vase
AND the vase is high above the floor

AND the vase is breakable
AND the vase is hard
IF a person expects an object to fall
AND the object is high above the floor
AND the object is breakable
AND the object is hard
THEN the person probably expects to hear a loud crash

Q: Why is Max excited?
A: Max expects to hear a loud crash
AND Max wants to hear a loud crash
IF a kid expects to hear a loud crash
AND the kid wants to hear a loud crash
THEN the kid is probably excited

Panel 1.4.2 Max lets go of the vase

Max's hands are open.
The vase is not supported.

1.4.2 Max lets go of the vase

Q: Why does Max let go of the vase
A: This is the main subaction of the larger action of dropping the vase.

Q: Why are Max's hands open
A: Max grasps the vase
AND Max lets go of the vase
IF a person grasps an object
AND the person lets go of the object
THEN the person's hands are probably open.

Q: Why is the vase not supported?
A: Max's hands support the vase
AND Max lets go of the vase
IF a person's hands support an object
AND the person lets go of the object
THEN the object is not supported

Panel 1.4.3 The vase falls

1.4.3 The vase falls

The vase is moving down.
It is lower than before.

Q: Why does the vase fall?

A: The vase is high above the floor
    AND the vase is not supported
IF an object is above the floor
AND the object is not supported
THEN the object probably falls

Q: Why is the vase moving down?
A: The vase falls
IF an object falls
THEN the object is definite moving down.

Q: Why is the vase lower than before?
A: The vase is moving down
IF an object is moving down
THEN the object is lower than before

Panel 1.4.4 The vase breaks

The vase makes a crash.
The crash is very loud
The vase is broken.
The vase's pieces are on the floor.
They are scattered.
Max hears the crash.
Max sees the vase's pieces.
Max thinks the crash is cool.
Max feels powerful (thinks he is powerful)
He is having fun.

1.4.4 The vase breaks

Q: Why does the vase break?
A: The vase falls
    AND the vase is high above the floor
    AND the vase is breakable
IF an object falls
AND the object is high above the floor
AND the object is breakable
THEN the object probably breaks

Q: Why does the vase make a crash?
A: The vase falls
    AND the vase is high above the floor
    AND the vase is breakable
    AND the vase is hard
IF a an object to fall
AND the object is high above the floor
AND the object is breakable

AND the object is hard
THEN the object makes a crash

Q: Why is the crash very loud?
A: This is typical
A crash is probably very loud

Q: Why is the vase broken?
A: The vase breaks
IF an object breaks
THEN the object is definitely broken

Q: Why are the vase's pieces on the floor?
A: The vase is high above the floor
AND the vase falls
AND the vase breaks
IF an object is high above the floor
AND the vase falls
AND the vase breaks
*THEN the object's pieces are probably on the flo*or

Q: Why are the vase's pieces scattered?
A: The vase is high above the floor
AND the vase falls
AND the vase breaks
IF an object is high above the floor
AND the vase falls
AND the vase breaks
THEN the object's pieces are probably scattered

Q: Why does Max hear the crash?
A: Max drops the vase
AND the vase makes a crash
IF a person drops an object
AND the object makes a crash
THEN the person probably hears the crash

Q: Why does Max see the vase's pieces?
A: Max is facing the floor
AND Max drops the vase
AND the vase breaks
AND the vase's pieces are on the floor
IF person is facing the floor
AND the person drops an object
AND the object breaks
AND the object's pieces are on the floor

THEN the person probably sees the object's pieces

Q: Why does Max think the crash is cool?
A: This is typical
A kid possibly thinks a crash is cool

Q: Why does Max feel powerful?
A: Max decides to break the vase
AND Max drops the vase
AND the vase breaks
IF person decides to damage an object
AND the person does something to the object
AND the object is damaged
THEN the person probably feels powerful

Q: Why is Max having fun?
A: Max feels powerful
IF person feels powerful
THEN the person probably is having fun

APPENDIX B: USER TEST RESULTS

As discussed in 5.3 Data Analysis, this appendix lists the data collected in the user test. First we see the statements and rules from the eight test subjects that completed the user evaluation. Then we see a comparison of the expected data from Appendix A against the collected data.

User#1

1. Opening Setting (***Training***)
    a. It is daytime
        Possibly it is daytime
    b. The ashtray is in the living room (*Note, the user put in the ashtray as an experiment in adding objects*)
        Possibly an ashtray is in the living room (*NA – Training*)
    c. The vase is in the living room
        Possibly a vase is in a living room (*NA – Training*)
    d. The ashtray is on top of the coffee table
        Possibly an ashtray is on top of a coffee table (*NA – Training*)
    e. The vase is on top of the side table
        Possibly a vase is on top of a side table (*NA – Training*)
    f. Max is in the living room
        Possibly a boy is in a living room (*NA – Training*)
    g. Max is on the sofa
        Possibly a boy is on a sofa (*NA – Training*)
    h. Max is sitting
        If a person is on a seat
        Probably the person is sitting (*NA – Training*)
    i. Max is bored
        Possibly a boy is bored (*NA – Training*)
2. Max looks around for something to do (***Training***)
    a. Caption: Max looks around
        If a person is bored
        Probably the person looks around (*NA – Training*)
    b. Max sees the vase
        If an object is in a living room
        AND a person is in the living room
        AND the person looks around
        Possibly the person sees the object (*NA – Training*)
    c. Max desires to disturb the vase
        If a boy is bored
        AND the boy sees a vase
        Possibly the boy desires to disturb the vase (*NA – Training*)
    d. Max desires to be near the vase
        If a person desires to do an action involving a vase
        Probably the person desires to be near the vase (*NA – Training*)
    e. Max is not close to the vase

Possibly a boy is not close to a vase (*NA – Training*)

3. Max gets up (***Training***)
    a. Caption: Max stand up
        If a person is sitting
        AND the person desires to be near an object
        AND the person is not close to the object
        Probably the person stands up (*NA – Training*)
4. Max walks to the side table
    a. Caption: Max walks to the side table
        If an object is on top of a second object
        AND a boy desires to disturb the first object
        AND the boy is not close to the first object
        Probably the boy walks to the second object (*Acceptable*)
    b. Max is standing
        If a person walks to an object
        Definitely the person is standing (*Acceptable*)
5. Max grasps the vase
    a. Caption: Max grasps the vase
        If a person desires to do an action involving an object
        Probably the person grasps the object (*NOT Acceptable – person could do other things besides grasp*)
    b. Max is in contact with the vase
        If a person desires to disturb an object
        AND the person grasps the object
        Definitely the person is in contact with the object (*Acceptable but the first antecedent is not necessary*)
6. Max picks up the vase
    a. Caption: Max picks up the vase
        If a person desires to disturb an object
        Probably the person picks up the object (*NOT Acceptable – person could do things besides pick up the object*)

User#2

1. Opening Setting (***Training***)
    (*Same as User #1 except no ashtray, which is inconsequential*)
2. Max looks around for something to do (***Training***)
    a. Caption: Max looks around
        If a boy is bored
        Probably the boy looks around (*NA – Training*)
    b. Max is curious about the texture of the vase
        If an object looks reflective
        And it is daytime
        And a juvenile person is in a living room
        And the juvenile person is bored

And the juvenile person looks around
Probably the juvenile person is curious about the texture of the object (*NA – Training*)

c. Max desires to play with the vase
If a child is bored
And the child is curious about the physical state of an object
Probably the child desires to play with the object (*NA – Training*)

d. Max desires to be holding the vase
If person is curious about the texture of an object
Probably the person desires to be holding the object (*NA – Training*)

e. Max is not near the vase
Possibly a boy is not near a vase (*NA – Training*)

f. Max desires to walk to the vase
If a person desires to be holding an object
And the person is not near the object
Probably the person desires to walk to the object (*NA – Training*)

3. Max gets up (***Training***)
   a. Caption: Max stand up
      If a person desires to walk to an object
      Definitely the person stands up (*Not an acceptable rule, but this occurred during training and was left unchanged*)
4. Max walks to the side table
   a. Caption: Max comes to the side table
      If an object is on top of a second object
      And a person desires to walk to the first object
      Probably the person comes to the second object (*Acceptable*)
   b. Max is near the vase
      If an object is on top of a second object
      And a person comes to the second object
      Definitely the person is near the first object (*Acceptable*)
5. Max grasps the vase
   a. Caption: Max grasps the vase
      If a person is curious about the texture of an object
      And the person desires to be holding the object
      Probably the person grasps the object (*Acceptable, but first antecedent is unnecessary*)
   b. Max is in curious about the weight of the vase
      If a boy grasps a vase
      Possibly the boy is curious about the weight of the vase (*NA – Possibly*)
6. Max picks up the vase
   a. Caption: Max lifts the vase
      If a person is bored
      And the person is curious about the texture of an object
      And the person grasps the object
      Probably the person lifts the object (*Acceptable*)
   b. The vase is in the right arm

If a person grasps an object
Definitely the vase is in an arm (*Acceptable*)

7. Max turns his body
    a. Caption: Max turns
        If a boy desires to play with an object
        And the boy lifts the object
        Probably the boy moves (*Acceptable*)
    b. The vase is heavy
        If an object is made of ceramic
        Probably the object is heavy (*Acceptable, with dissent. One judge brought up the case of a tiny ceramic object. Nevertheless, two judges felt the "normal" case should be Acceptable*)
    c. Max does not know the weight of the vase
        Probably a boy does not know the weight of a vase (*Acceptable but barely*)
8. Max lets go of the vase
    a. Caption: Max drops the vase
        If a child is curious about the physical state of an object
        And the child lifts the object
        And the object is heavy
        And the child does not know the weight of the object
        Probably the child drops the object (*Acceptable*)
9. The vase falls (starts to fall)
    a. Caption: The vase falls (starts to fall)
        If a boy is standing
        And an object is in an arm
        And the boy drops the object
        Definitely the object falls (*Acceptable*)
    b. The vase is above the floor
        If an object starts to fall
        Definitely the object is above the floor (*Acceptable, with dissent. One judge interpreted "floor" as a floor of a man-made structure. The other two judges interpreted "floor" as any bottom-most level*)
    c. The vase is not on the floor
        If a vase starts to fall
        Definitely the vase is not on a floor (*Acceptable*)
10. The vase breaks
    a. Caption: The vase hits the floor
        If an object falls
        Probably the object hits a floor (*Acceptable, with dissent – see 9b*)
    b. The vase breaks
        If an object if breakable
        And a person drops the object

And the object hits the floor
Definitely the object breaks (*Acceptable*)

c. The vase is on the floor
   If an object falls
   And the object hits a floor
   And the object breaks
   Definitely the object is on the floor (*Acceptable*)

User#3

1. Opening Setting (***Training***)
   (Same as User #1 with the addition of the following)
   a. Mommy is in the backyard
      Possibly a woman is in a backyard (*NA – Training*)
   b. Mommy is near the tree
      Possibly a woman is near a tree (*NA – Training*)
   c. Mommy is standing
      Possibly a woman is standing (*NA – Training*)
   d. Mommy is angry
      Possibly a woman is angry (*NA – Training*)
   e. Max is in punishment
      Possibly a boy is in punishment (*NA – Training*)
   f. Max does not desire to be bored
      Definitely a boy does not desire to be bored (*NA – Training*)
   g. Max desires to be naughty
      If a boy is in punishment
      Probably the boy desires to be naughty (*NA – Training*)
   h. The floor is hard
      Possibly a floor is hard (*NA – Training*)
2. Max looks around for something to do (***Training***)
   a. Caption: Max looks around
      If a person is bored
      And the person is in punishment
      And the person desires to be naughty
      Definitely the person looks around (*NA – Training*)
   b. Max is angry
      Possibly a boy is angry (*NA – Training*)
   c. Max sees the vase
      If a person is in a room
      And the object is in the room
      And the person looks around
      Possibly the person sees the object (*NA – Training*)
   d. Max is not near the vase
      If an object is next to a piece of furniture
      And a second piece of furniture is next to the object
      And a person is on the first piece of furniture
      And a second object is on top of the second piece of furniture

Probably the person is not near the second object (*NA – Training*)

e. Max thinks Mommy likes the vase

Possibly a boy thinks a woman likes a vase (*NA – Training*)

f. Max is curious about what would happen if the vase were to smash

If a boy is angry
And the boy sees a vase
And the boy thinks a woman likes the vase
Probably the boy is curious about what would happen if the vase were to smash (*NA – Training*)

g. Max intends to walk to the vase

If a person is not near an object
And the person is curious about what would happen if the object were to smash
Probably the person intends to walk to the object (*NA – Training*)

3. Max gets up (***Training***)
   a. Caption: Max stands up

      If a person is sitting
      And the person intends to walk to an object
      Probably the person stands up (*NA – Training*)

   b. Why is Max on the floor

      If a person is on a seat
      And the person stands up
      Definitely the person is on a floor (*NA – Training*)

   c. Why is Max standing

      If a living thing stands up
      Definitely the living thing is standing (*NA – Training*)

4. Max walks to the side table
   a. Caption: Max walks to the side table

      If an object is on top of a second object
      And living thing intends to walk to the first object
      Probably the living thing walks to the second object (*Acceptable*)

   b. Max is near the vase

      If an object is on top of a second object
      And a living thing walks to the second object
      Definitely the living thing is near the first object (*Acceptable*)

5. Max grasps the vase
   a. Caption: Max grasps the vase

      If a person desires to be naughty
      And the person is curious about what would happen if a vase were to smash
      Probably the person grasps the object (*Acceptable*)

   b. Max is not bored

      If a person grasps an object
      Possibly the person is not bored (*NA – Possibly*)

6. Max picks up the vase
   a. Caption: Max picks up the vase

If a person desires to be naughty
And the person is curious about what would happen if an object were to smash
And the person grasps the object
Definitely the person picks up the object (*Acceptable*)

b. The vase is not on top of the side table
    If a person picks up an object
    Probably the object is not on top of a second object (*Acceptable*)
c. The vase is in the arms
    If a person grasps an object
    Definitely the object is in some arms (*Acceptable*)

User#4

1. Opening Setting (***Training***)
    (Same as User #1 with the addition of the following)
    a. Max is inactive
        Possibly a boy is inactive (*NA – Training*)
    b. Max is naughty
        Probably a boy is naughty (*NA – Training*)
    c. The side table is on the floor
        Probably a side table is on a floor (*NA – Training*)
2. Max looks around for something to do (***Training***)
    a. Caption: Max looks around the living room
        If a child is in a room
        And the child is bored
        Probably the child looks around the room (*NA – Training*)
    b. Max sees the vase
        If a vase is red
        And a person is in a room
        And the vase is in the room
        And the person looks around the room
        Probably the person sees the vase (*NA – Training*)
    c. Max desires to be having fun
        If a child is bored
        Probably the child desires to be having fun (*NA – Training*)
    d. Max desires to break the vase
        If a child is naughty
        And the child sees a jar
        And the child desires to be having fun
        Probably the child desires to break the jar
    e. Max desires to walk to the vase
        If a person desires to break a jar
        Probably the person desires to walk to the jar (*Not a good rule but this is training*)
3. Max gets up (***Training***)
    a. Caption: Max stands up

If a person is sitting
And the person desires to walk to an object
Probably the person stands up (*NA – Training*)

b. Max is standing
If a living thing stands up
Definitely the living thing is standing (*NA – Training*)

c. Max is next to the sofa
If a person is on a piece of furniture
And the person stands up
Probably the person is next to the piece of furniture (*NA – Training*)

4. Max walks to the side table
    a. Caption: Max walks to the side table
    If an object is on a table
    And a person desires to walk to the object
    And the person is standing
    Probably the person walks to the table (*Acceptable*)
    b. Max is near the side table
    If a living thing goes to an object
    Definitely the living thing is near the object (*Acceptable*)
    c. Max is near the easy chair
    If an object is next to a second object
    And a living thing walks to the first object
    Definitely the living thing is near the second object (*Acceptable*)
    d. Max desires to pick up the vase
    If a person is near a side table
    And a jar is on the side table
    And a person desires to drop the jar
    Probably the person desires to pick up the jar (*Acceptable*)
5. Max grasps the vase
    a. Caption: Max grasps the vase
    If a person desires to pick up an object
    Probably the person grasps the object (*Acceptable*)
    b. Max is in contact with the vase
    If a person grasps an object
    Definitely the person is in contact with the object (*Acceptable*)
    c. The arms are in contact with the vase
    If some arms are part of a person
    And the person grasps an object
    Probably the arms are in contact with the object (*Acceptable*)
6. Max picks up the vase
    a. Caption: Max picks up the vase
    If a person desires to pick up an object
    And the person grasps the object
    Probably the person picks up the object (*Acceptable*)
    b. The vase is not on the side table
    If an object is on a piece of furniture

And a person picks up the object
Definitely the object is not on the piece of furniture (*Acceptable*)

c. The vase is over the side table
   If an object is on a piece of furniture
   And a person picks up the object
   Definitely the object is over the piece of furniture (*Acceptable*)
d. Max desires to turn with the vase
   If an object is breakable
   And the person desires to break the object
   And the person desires to drop the object
   And the object is over a piece of furniture
   Probably the person desires to turn with the object
e. Max desires to turn far (away) from the side table (*Acceptable*)
   If an object is breakable
   And a person desires to break the object
   And the person desires to drop the object
   And the object is over a piece of furniture
   Probably the person desires to turn far (away) from the piece of furniture (*Acceptable*)

7. Max turns his body
   a. Caption: Max turns far (away) from the side table
      If a person desires to turn far (away) from the side table
      Probably the person turns far (away) from the side table (*Acceptable*)
   b. Max turns with the vase
      If person picks up an object
      And the person turns far away from the second object
      Definitely the person turns with the first object (*Acceptable*)
   c. The vase is not over the side table
      If a person picks up an object
      And the object is over a second object
      And the person turns far (away) from the first object
      Definitely the first object is not over the second object (*Acceptable*)
   d. Max is facing the coffee table
      If a person turns far from an object
      Possibly the person is facing the second object (*NA – Possibly*)
   e. The vase is over the floor
      If an object is on a floor
      And a second object is over the first object
      And a person turns with the second object
      Definitely the second object is over the floor (*Acceptable*)
8. Max lets go of the vase
   a. Caption: Max releases the vase
      If a person desires to drop an object
      Probably the person releases the object (*Acceptable*)
   b. Max is not in contact with the vase
      If a person releases an object

Definitely the person is not in contact with the object (*Acceptable*)

c. The arms are not in contact with the vase
If some arms are part of a living thing
And the living thing is not in contact with an object
Definitely the arms are not in contact with the object (*Acceptable*)

9. The vase falls
    a. Caption: The vase falls
        If an object is not on a second object
        And a living thing releases the first object
        Definitely the first object falls (*Acceptable*)
10. The vase breaks
    a. Caption: The vase breaks
        If an object is breakable
        And the object is over a floor
        And the object falls
        Probably the object breaks (*Acceptable*)
    b. Max is having fun
        If a person desires to break an object
        And the object breaks
        Definitely the person is having fun (*Not Acceptable – the person may desire to break the object out of a sense of duty*)
    c. The vase is on the floor
        If an object is over a floor
        And the object hits a floor
        And the object falls
        Definitely the object is on the floor (*Acceptable*)
    d. The vase pieces are part of the vase
        If a vase breaks
        Definitely the pieces are part of the vase (*Acceptable*)

User#5

1. Opening Setting (***Training***)
    (Same as User #1 with the addition of the following)
    a. Max is inactive
        Possibly a boy is inactive (*NA – Training*)
    b. Max desires to move
        If a child is inactive
        Probably the child desires to move (*NA – Training*)
2. Max looks around for something to do (***Training***)
    a. Caption: Max explores
        If a person is bored
        Probably the person explores (*NA – Training*)
    b. Max sees the vase
        If an object is in a living room

And a person is in a living room
And the person explores
Probably the person sees the object (*NA – Training*)

c. The vase is rounded
Possibly a vase is rounded (*NA – Training*)

d. Max notes the vase
If an object is in a living room
And the object is on a side table
And a person is in the living room
And the person is bored
Probably person notes the object (*NA – Training*)

e. Max desires to explore the vase
If a person notes a vase
Probably the person desires to explore the vase (*NA – Training*)

3. Max gets up (***Training***)
   a. Caption: Max stands up
      If a person is sitting
      And the person desires to explore an object
      Probably the person stands up (*NA – Training*)
   b. Max is active
      If a person stands up
      Definitely the person is active (*NA – Training*)
   c. Max is next to the sofa
      If a person is on a seat
      And the person is sitting
      And the person stands up
      Probably the person is next to the seat
   d. Why is Max standing
      If a person stands up
      Definitely the person is standing (*NA – Training*)
4. Max walks to the side table
   a. Caption: Max walks to the vase
      If person desires to explore an object
      Probably the person walks to the object (*Acceptable with dissent. Two judges felt "explore" involved walking to an object*)

5. Max grasps the vase
   a. Caption: Max grasps the vase
      If a person desires to explore an object
      Probably the person grasps the object (*Acceptable*)

   b. Max is next to the side table
      If an object is on a side table
      And a person desires to explore the object
      Definitely the person is next to the side table (*Not Acceptable – need an*

action to get to the side table)

6. Max picks up the vase
    a. Caption: Max picks up the vase
        If a person desires to explore an object
        Probably the person picks up the object (*Not Acceptable – not necessarily pick it up*)
    b. The vase is above the side table
        If an object is on a side table
        And a person picks up the object
        Definitely the object is above the side table (*Acceptable*)
    c. The vase heavy
        Possibly a vase is heavy (*NA – Possibly*)
    d. Max is uncomfortable
        If person picks up an object
        And the object is heavy
        Definitely the person is uncomfortable (*Not Acceptable – could be strong person*)
    e. Max desires to sit on the sofa
        If a person is uncomfortable
        Probably the person desires to sit on a seat (*Not Acceptable – need more info to conclude the person sits*)
7. Max turns his body
    a. Caption: Max turns with the vase
        If a person is uncomfortable
        And the person desires to sit on a seat
        And an object is heavy
        Definitely the person turns with the object (*Not Acceptable – need more info*)
    b. The vase is next to the side table
        If person is next to a side table
        And the person picks up an object
        And the person turns with the object
        Definitely the object is next to the side table (*Acceptable*)
8. Max lets go of the vase
    a. Caption: Max releases the vase
        If a person picks up an object
        And the object is heavy
        And the person is uncomfortable
        Probably the person releases the object (*Not Acceptable – not true for adult*)
9. The vase falls
    a. Caption: The vase falls
        If a person picks up an object
        And the object is heavy
        And the person is uncomfortable
        Probably the person releases the object

Definitely the object falls (*Acceptable*)

10. The vase breaks
    a. Caption: The vase breaks
       If an object is next to a side table
       And a person picks up the object
       And the object is heavy
       And the person is uncomfortable
       And the person turns with the object
       And the person releases the object
       And the object falls
       Probably the object breaks (*Acceptable with dissent. One judge felt object should be explicitly breakable. Other two judges thought "normal" case would be to break*)
    b. Max is inactive
       If a person releases an object
       Probably the person is inactive (*Acceptable with dissent – One judge felt that releasing an object could be part of an active game. Others felt it depends on meaning of "inactive"*)

User#6

1. Opening Setting (***Training***)
   (Same as User #1)
2. Max looks around for something to do (***Training***)
   a. Caption: Max looks around the living room
      If a person in an area
      And the person is bored
      Probably the person looks around the area (*NA – Training*)
   b. Max does not desire to be bored
      Definitely a boy does not desire to be bored (*NA – Training*)
   c. Max desires to be having fun
      If a person does not desire to be bored
      Probably a person desires to be having fun (*NA – Training*)
   d. The vase has (is) novel
      Probably a vase has novel (*NA – Training*)
   e. Max desires to examine the vase
      If an object is in a room
      And a juvenile person is in a room
      And the juvenile person desires to be having fun
      And the object has novel
      Probably the juvenile person desires to examine the object (*NA – Training*)
   f. Max is not close to the vase
      Possibly a vase is not close to a boy (*NA – Training*)
3. Max gets up (***Training***)

    a. Caption: Max stands up
        If a person is sitting
        And the person desires to examine an object
        And the object is not close to the person
        Probably the person stands up (*NA – Training*)
    b. Max is in front of the sofa
        If a person is on a seat
        And the person stands up
        Definitely the person is in front of the seat (*NA – Training*)
    c. Why is Max standing
        If a living thing stands up
        Definitely the living thing is standing (*NA – Training*)
4. Max walks to the side table
    a. Caption: Max walks to the vase
        If a piece of furniture is in an area
        And an object is on top of the piece of furniture
        And a person is in the area
        And the person desires to examine the object
        And the object is not close to the person
        Probably the person walks to the piece of furniture (*Acceptable*)
    b. Max is near the side table
        If a living thing goes to an object
        Definitely the living thing is near the object (*Acceptable*)
    c. Max is near the vase
        If an object is on top of a second object
        And a living thing is near the second object
        Definitely the living thing is near the first object (*Acceptable*)
    d. The vase is close to Max
        If a living thing is near an object
        Definitely the object is close to the living thing (*Acceptable*)
5. Max grasps the vase
    a. Caption: Max grasps the vase
        If a child desires to examine an object
        And the child is near the object
        Probably the child grasps the object (*Acceptable*)
6. Max picks up the vase
    a. Caption: Max picks up the vase
        If a child desires to examine an object
        And the child grasps the object
        Probably the child picks up the object (*Not Acceptable, should be possibly*)
    b. The vase is inside the arms
        If some arms are part of a living thing
        And the living thing picks up an object
        Definitely the object is inside the arms (*Acceptable*)
    c. Max desires to break the vase
        If an object is breakable

And a child desires to be having fun
And the child is near the object
Possibly the child desires to break the object (*NA – Possibly*)

User#7

1. Opening Setting (***Training***)
   (Same as User #6; Note, Users #6 and #7 were trained together)
2. Max looks around for something to do (***Training***)
   (Same as User #6)
3. Max gets up (***Training***)
   (Same as User #6)
4. Max walks to the side table (***Training***)
   a. Caption: Max walks to the vase
      If a first object is on top of a second object
      And a person desires to examine the first object
      And the first object is not close to the person
      Probably the person walks to the second object (*NA – Training*)
   b. Max is next to the side table
      If a person walks to an object
      Definitely the person is next to the object (*NA – Training*)
   c. The vase is close to Max
      If a first object is on top of a second object
      And a person is next to the second object
      Definitely the first object is close to the person (*NA – Training*)
5. Max grasps the vase (***Training***)
   a. Caption: Max grasps the vase
      If a first object is on top of a second object
      And a person is next to the second object
      And the person desires to examine the first object
      Definitely the person grasps the first object (*NA – Training*)
6. Max picks up the vase
   a. Caption: Max picks up the vase
      If a person desires to examine an object
      Possibly the person picks up the object (*NA – Possibly*)
   b. The vase is in the arms
      If a person holds a vase
      And the person picks up the vase
      Definitely the vase is inside the arms (*Acceptable*)
   c. Max is not bored
      If a person does an action involving an object
      Possibly the person is not bored (*NA – Possibly*)
   d. Max desires to break the vase
      If a person desires to be having fun
      Possibly the person desires to break an object (*NA – Possibly*)
   e. Max desires to turn away far from the side table
      If a person is next to a first object

And the person desires to break a second object
Possibly the person desires to turn away far from the first object (*NA – Possibly*)

7. Max turns his torso
   a. Caption: Max turns the torso
      If a boy desires to break a vase
      Probably the boy turns a torso (Not Acceptable)

User#8

1. Opening Setting (***Training***)
   (Same as User #1 with the addition of the following)
   a. Max is inactive
      Possibly a boy is inactive (*NA – Training*)
   b. The vase is next to the easy chair
      If a first object is next to a second object
      And a third object is on top of the first object
      Definitely the third object is next to the second object (*NA – Training*)
2. Max looks around for something to do (***Training***)
   a. Caption: Max looks around the living room
      If a person is in a structure
      And the person is bored
      Probably the child looks around the structure (*NA – Training*)
   b. Max sees the vase
      If a person is in a room
      And an object is in a room
      And the person looks around the room
      Probably the person sees the vase (*NA – Training*)
   c. Max desires to play
      If a child desires to be having fun
      Probably the person desires to play (*NA – Training*)
   d. Max desires to play with the vase
      If a child sees a vase
      And the child desires to play
      Possibly the child desires to play with the vase (*NA – Training*)
   e. Max desires to get the vase
      If a person is on a seat
      And the person desires to play with an object
      And the seat is not by the object
      Probably the person desires to get the object (*NA – Training*)
3. Max gets up (***Training***)
   a. Caption: Max stands up
      If a person is on a seat
      And the person desires to get an object
      And the seat is not by the object

Probably the person stands up (*NA – Training*)

b. Max is in front of the sofa

If a person is on a seat
And the person stands up
Probably the person is in front of the seat (*NA – Training*)

c. Max is standing

If a object stands up
Definitely the object is standing (*NA – Training*)

d. Max desires to walk to the vase

If a person is in front of a seat
And the seat is not by the object
And the person desires to get the object
Probably the person desires to walk to the object (*NA – Training*)

4. Max walks to the side table
   a. Caption: Max walks to the vase

      If a person desires to walk to an object
      Probably the person walks to the object (*Acceptable*)

   b. Max is active

      If a person walks to an object
      Definitely the person is active (*Acceptable*)

   c. Max is by the side table

      If a first object is on top of a second object
      And a person walks to the first object
      Definitely the person is by the second object (*Acceptable*)

   d. Max is facing the vase

      If a person walks to an object
      Probably the person is facing the object (*Acceptable*)

5. Max grasps the vase
   a. Caption: Max grasps the vase

      If a person desires to get a vase
      Probably the person grasps the vase (*Acceptable*)

6. Max picks up the vase
   a. Caption: Max picks up the vase

      If a person desires to play with an object
      And the person grasps the object
      Probably the person picks up the object (*Acceptable*)

   b. Max desires to throw the vase on the floor

      If a child desires to play with an object
      Possibly the child desires to throw the object on the floor (*NA – Possibly*)

   c. Max does not desire to throw the vase on the side table

      If a person does not desire to throw a first object on a floor
      Probably the person does not desire to throw the first object on a second object(*Acceptable*)

   d. Max is facing the side table

      If a first object is on top of a second object
      And a person is facing the first object

Definitely the is facing the second object (*Acceptable*)

e. Max desires to turn far (away) from the side table (*Acceptable*)
   If a person desires to throw a first object on the floor
   And the person does not desire to throw the first object on a second object
   And the person is facing the second object
   Probably the person desires to turn far (away) from the second object (Acceptable)

f. The vase is in the arms
   If a person picks up an object
   Definitely the object is in some arms (*Acceptable*)

g. Max is holding the vase
   If a person picks up an object
   Probably the person is holding the object (*Acceptable*)

7. Max turns his body
   a. Caption: Max turns Max
      If a person is facing an object
      And the person desires to turn far (away) from the object
      Probably the person turns the person (*Acceptable*)
   b. Max is not facing the side table
      If person is facing an object
      And the person turns the person
      Definitely the person is not facing the object (*Not Acceptable, with dissent. Two judges felt that head does not definitely turn when a person turns*.)

8. Max lets go of the vase
   a. Caption: Max releases the vase
      If some arms are part of a person
      And an object is in the arms
      And the person desires to throw the object on a floor
      Probably the person releases the object (*Acceptable*)
   b. The arms are above the floor
      Definitely some arms are above a floor (*Acceptable*)
   c. The vase is not in the arms
      If some arms are part of a person
      And an object is in the arms
      And the person releases the object
      Definitely the object is not in the arms (*Acceptable*)
   d. The vase is over the floor
      If some arms are part of a person
      And an object is in the arms
      And the person releases the object
      And the arms are above the a floor
      Definitely the object is above the floor (*Acceptable*)
   e. Max is not holding the vase

If a person releases an object
Definitely the person is not holding the object (*Acceptable*)

9. The vase falls
    a. Caption: The vase falls
        If a person is holding an object
        And a person releases the object
        Probably the object falls (*Acceptable*)
10. The vase breaks
    a. Caption: The vase breaks
        If an object is breakable
        And the object is over a floor
        And the object falls
        Probably the object breaks (*Acceptable*)
    b. The vase is on the floor
        If an object is over a floor
        And the object falls
        And the object hits a floor
        Definitely the object is on the floor (*Not Acceptable – could bounce*)
    c. Max plays (is playing) with the vase
        If a person picks up an object
        And the person releases the object
        Probably the person plays with the object (*Acceptable*)
    d. Max is having fun
        If a person desires to play with an object
        And a person plays (is playing) with the object
        Probably the person is having fun (*Acceptable*)

Comparison with Expected Data

This section shows how the recall measurement was computed. First we see the data that I proposed to collect. These data are taken from Appendix A. Then, for each user, we see the corresponding collected data and a measurement of how many user statements matched the proposed data.

Originally Proposed Data

	Key Frame 4 Panel 1.3.1.2 **Max walks to the side table**	Key Frame 5 Panel 1.3.2.1 **Max grasps the vase**	Key Frame 6 Panel 1.3.2.2 **Max picks up the vase**	Key Frame 7 Panel 1.4.1 **Max turns is body**	Key Frame 8 Panel 1.4.2 **Max drops the vase**	Key Frame 9 Panel 1.4.3 **The vase falls**	Key Frame 4 Panel 1.4.4 **The vase breaks**
Intention	Max intends to pick up the vase		Max decides to break the vase	Max expects the vase to fall, Max is excited			Max is having fun
Location	Max is at the side table		The vase is above the table	The vase is above the floor	Then vase is not supported	The vase is lower	The vase is on the floor; The pieces are on the floor
Movement or Position	Max is standing	Max's in a position of holding or in contact with the vase	Max's arms are raised.	Max is facing away from the side table; The vase is above the floor; Max's torso is twisted	Max's hands are open		The vase is broken
Sensory		Max thinks the vase feels smooth. Max thinks the vase feels cool	Max thinks the vase is heavy				The vase makes a loud crash

Results: No results. This table is for comparing with the user tables below.

User#1 vs. Proposed

	Key Frame 4 Panel 1.3.1.2 **Max walks to the side table**	Key Frame 5 Panel 1.3.2.1 **Max grasps the vase**	Key Frame 6 Panel 1.3.2.2 **Max picks up the vase**	Key Frame 7 Panel 1.4.1 **Max turns is body**	Key Frame 8 Panel 1.4.2 **Max drops the vase**	Key Frame 9 Panel 1.4.3 **The vase falls**	Key Frame 4 Panel 1.4.4 **The vase breaks**
Intention	Max desires to disturb the vase		Not necessary	NA			NA
Location	MISSING		MISSING	NA	NA	NA	NA
Movement or Position	Max is standing	Max is in contact with the vase	MISSING	NA	NA		NA
Sensory		Not necessary	Not necessary				NA

Results: 3/6 (50%) of the data types were collected.

User#2 vs. Proposed

	Key Frame 4 Panel 1.3.1.2 **Max walks to the side table**	Key Frame 5 Panel 1.3.2.1 **Max grasps the vase**	Key Frame 6 Panel 1.3.2.2 **Max picks up the vase**	Key Frame 7 Panel 1.4.1 **Max turns is body**	Key Frame 8 Panel 1.4.2 **Max drops the vase**	Key Frame 9 Panel 1.4.3 **The vase falls**	Key Frame 4 Panel 1.4.4 **The vase hits the floor**
Intention	Max desires to play with the vase	Max is curious about the weight of the vase	(see previous key frame)	Max does not know the weight of the vase			MISSING
Location	Max is near the vase		The vase is in the right arm	MISSING	MISSING	The vase is above the floor; The vase is not on the floor	The vase is on the floor
Movement or Position	MISSING	Max is in contact with the vase	MISSING	MISSING	MISSING		The vase breaks (is broken)
Sensory		Not necessary	(see next key frame)	The vase is heavy			MISSING

Results: 9/17 (53%) of the data types were collected. Note, this user had 17 expected statements instead of 16 because the weight of the vase was relevant to this interpretation.

User#3 vs. Proposed

	Key Frame 4 Panel 1.3.1.2 **Max walks to the side table**	Key Frame 5 Panel 1.3.2.1 **Max grasps the vase**	Key Frame 6 Panel 1.3.2.2 **Max picks up the vase**	Key Frame 7 Panel 1.4.1 **Max turns is body**	Key Frame 8 Panel 1.4.2 **Max drops the vase**	Key Frame 9 Panel 1.4.3 **The vase falls**	Key Frame 4 Panel 1.4.4 **The vase breaks**
Intention	Max is curious; Max intends to walk to the vase	Max is not bored	Not necessary	NA			NA
Location	Max is near the vase		The vase is on top of the side table	NA	NA	NA	NA
Movement or Position	Max is standing	MISSING	The vase is in the arms	NA	NA		NA
Sensory		Not necessary	Not necessary				NA

Results: 5/6 (83%) of the data types were collected

User#4 vs. Proposed

	Key Frame 4 Panel 1.3.1.2 **Max walks to the side table**	Key Frame 5 Panel 1.3.2.1 **Max grasps the vase**	Key Frame 6 Panel 1.3.2.2 **Max picks up the vase**	Key Frame 7 Panel 1.4.1 **Max turns is body**	Key Frame 8 Panel 1.4.2 **Max drops the vase**	Key Frame 9 Panel 1.4.3 **The vase falls**	Key Frame 4 Panel 1.4.4 **The vase breaks**
Intention	Max desires to break the vase; Max desires to pick up the vase		Max desires to turn with the vase	(see previous key frame)			Max is having fun;
Location	Max is near the side table		The vase is not on the side table; The vase is over the side table	The vase is not over the side table; The vase is above the floor	MISSING	MISSING	The vase is on the floor;
Movement or Position	Max is standing	Max is in contact with the vase; The arms are in contact	(see previous key frame)	Max is facing the coffee table	MISSING		The pieces of the vase are on the floor
Sensory		Not necessary	Not necessary	Not necessary			MISSING

Results: 13/16 (81%) of the data types were collected

User#5 vs. Proposed

	Key Frame 4 Panel 1.3.1.2 **Max walks to the side table**	Key Frame 5 Panel 1.3.2.1 **Max grasps the vase**	Key Frame 6 Panel 1.3.2.2 **Max picks up the vase**	Key Frame 7 Panel 1.4.1 **Max turns is body**	Key Frame 8 Panel 1.4.2 **Max drops the vase**	Key Frame 9 Panel 1.4.3 **The vase falls**	Key Frame 4 Panel 1.4.4 **The vase breaks**
Intention	Max desires to explore the vase		Max is uncomfortable; Max desires to sit on the sofa	(see previous key frame)			Max is inactive
Location	(see next key frame)	Max is next to the side table	The vase is above the side table	The vase is not next to the side table	MISSING	MISSING	MISSING
Movement or Position	Max is standing	MISSING	MISSING	MISSING	MISSING		MISSING
Sensory		Not necessary	The vase is heavy				MISSING

Results: 9/17 (53%) of the data types were collected. Note, this user had 17 expected statements instead of 16 because the weight of the vase was relevant to this interpretation.

User#6 vs. Proposed

	Key Frame 4 Panel 1.3.1.2 **Max walks to the side table**	Key Frame 5 Panel 1.3.2.1 **Max grasps the vase**	Key Frame 6 Panel 1.3.2.2 **Max picks up the vase**	Key Frame 7 Panel 1.4.1 **Max turns is body**	Key Frame 8 Panel 1.4.2 **Max drops the vase**	Key Frame 9 Panel 1.4.3 **The vase falls**	Key Frame 4 Panel 1.4.4 **The vase breaks**
Intention	Max desires to examine the vase		Max desires to break the vase	NA			NA
Location	Max is near the side table; Max is near the vase		The vase is in the arms	NA	NA	NA	NA
Movement or Position	Max is standing	MISSING	(The vase is in the arms)	NA	NA		NA
Sensory		Not necessary	Not necessary				NA

Results: 5/6 (83%) of the data types were collected. Note, the Intention in Key Frame 6 is not counted because it was not expected.

User#7 vs. Proposed

	Key Frame 4 Panel 1.3.1.2 **Max walks to the side table**	Key Frame 5 Panel 1.3.2.1 **Max grasps the vase**	Key Frame 6 Panel 1.3.2.2 **Max picks up the vase**	Key Frame 7 Panel 1.4.1 **Max turns is body**	Key Frame 8 Panel 1.4.2 **Max drops the vase**	Key Frame 9 Panel 1.4.3 **The vase falls**	Key Frame 4 Panel 1.4.4 **The vase breaks**
Intention	NA		Max is not bored; Max desires to turn	MISSING			NA
Location	NA		The vase is in the arms	MISSING	NA	NA	NA
Movement or Position	NA	NA	(The vase is in the arms)	MISSING	NA		NA
Sensory			Not necessary				NA

Results: 3/6 (50%) of the data types were collected. Note, this user had training in key frames 4 and 5, and attempted to annotate only key frames 6 and 7.

User#8 vs. Proposed

	Key Frame 4 Panel 1.3.1.2 **Max walks to the side table**	Key Frame 5 Panel 1.3.2.1 **Max grasps the vase**	Key Frame 6 Panel 1.3.2.2 **Max picks up the vase**	Key Frame 7 Panel 1.4.1 **Max turns is body**	Key Frame 8 Panel 1.4.2 **Max drops the vase**	Key Frame 9 Panel 1.4.3 **The vase falls**	Key Frame 4 Panel 1.4.4 **The vase breaks**
Intention	Max desires to walk to the vase		Max desires to throw the vase on the floor; Max does not desire to throw the vase on the side table	(see previous key frame)			Max is having fun
Location	Max is by the side table		The vase is in the arms	MISSING	The vase is over the floor	MISSING	The vase is on the floor
Movement or Position	Max is standing; Max is facing the vase	MISSING	Max is facing the side table; Max is holding the vase	Max is not facing the side table	The arms are above the floor; Max is not hot holding the vase		NA
Sensory		Not necessary	Not necessary				MISSING

Results:13/16 (81%) of the data types were collected. Note, this user had training in key frames 4 and 5, and attempted to annotate only key frames 6 and 7.

APPENDIX C: SEMANTIC FIELD PREDICATES

As discussed in 4.1 Stative Predicates, HXP treats single-argument static predicates like *green*, *cold,* or *good* as continuous values in semantic field. Here are the semantic field predicates and their possible values.

Semantic Field Predicate	Values
absorbency	absorptive, nonabsorbent
activity	active, exhausted, inactive, rested
age	immature, mature, new, old (having lived long time), old (of long duration), young
animation	alive, dead
area	narrow, wide
cleanness	clean, dirty
clothing state	clothed, naked
color	blue, green, orange, red, yellow
comfort	comfortable, uncomfortable
consistency	breakable, dull, flexible, hard, hollow, sharp, soft, solid, unbreakable
depth	deep, shallow
domestication	tame, wild
extraordinariness	extraordinary, ordinary
freshness	fresh, stale
fullness	empty, full
functional state	broken, off, on, repaired, unbroken
gender	female, male
health state	injured, sick, uninjured, well
height	high, low, short, tall
hydration	dehydrated, hydrated, needs water
illumination	dark, light
length	long, short
made of	ceramic, fabric
novelty	novel, old
nutrition	malnourished, nourished, needs food
olfactory property	bad, good, salty, sour, sweet
personality	disobedient, friendly, loyal, mean, naughty, playful, sober, timid, unfriendly
pitch	high, low
position	backward, bent, closed, crouching, curved, flat, forward, hanging, horizontal, loose, moving, open, raised, shut, sitting, standing, stationary, straight, tense, tilted, upside-down, vertical
power	powerful, powerless
safety	harmful, harmless, in danger, safe
season	fall, spring, summer, winter
shape	crooked, round, rounded, square, straight
size	large, medium, small
social state	accompanied, alone, in punishment, supervised, unsupervised

(table cont.)

sound	bad, good
strength	strong, weak
tactile property	bad, good
tactual sensation	bad, cool, good, hot, itchy, pain, painful, pleasing, pleasure, tickling, tingling
taste property	bad, good, salty, sour, sweet
temperature	cold, cool, hot, warm
texture	rough, smooth
time of day	afternoon, daytime, evening, lunchtime, morning, night
visual property	bad, bright, colorless, colorful, dull, good, nonreflective, pretty, reflective, shiny, ugly, age
volume	loud, soft
weather	clear, cloudy, dry, rainy
weight	fat, heavy, light, thin
wetness	dry, wet
width	narrow, thick, thin, wide

APPENDIX D: PHYSICAL STATES

Here are the current HXP physical states. If the concept is found in WordNet 3.0, the WordNet sense key is provided. If it is not available in WordNet, then it is marked as *HXP*. As discussed in 4.1 Stative Predicates, HXP physical states are organized into subcategories according to the hypernyms relation. Furthermore, each state is associated with an HXP value template that helps the user interface prompt for the correct inputs. Each concept here is listed with its definition, list of hypernyms, and top-level value template.

Physical State	WordNet SenseKey or HXP	Definition	Hypernyms	Value Template
above	HXP	in a higher place than; over	relative location	VtSpecificObject
absorbency	1:07:00::	the property of being absorbent	consistency	VtEnum
activity	1:07:00::	the trait of being active; moving or acting rapidly and energetically	physical state	VtEnum
against	HXP	in contact with	relative location	VtSpecificObject
age	1:07:00::	how long something has existed	visual property, measure	VtEnum
along	HXP	through, on, beside, over, or parallel to the length or direction of	relative location	VtSpecificObject
animation	1:07:00::	the property of being able to survive and grow	physical state	VtEnum
area	1:07:00::	the extent of a 2-dimensional surface enclosed within a boundary	size	VtEnum
at	HXP	in, on, or near	relative location	VtSpecificObject
attached to	HXP	joined to	structurally related to	VtSpecificObject
authority of	HXP	'authority of' relationship	social state	VtSpecificObject
background location	HXP	the location of an object	location	VtBackground-Location
behind	HXP	at or toward the rear of	relative location	VtSpecificObject
below	HXP	lower down than	relative location	VtSpecificObject
brightness	1:07:00::	the location of a visual perception along a continuum from black to white	visual property	VtEnum
by	HXP	near or next to	relative location	VtSpecificObject
cleanness	1:26:00::	the state of being clean; without dirt	physical state	VtEnum

(table cont.)

close to	HXP	near	relative location	VtSpecificObject
clothing state	HXP	clothing state	physical state	VtEnum
color	1:07:00::	a visual attribute of things that results from the light they emit or transmit or reflect	visual property	VtEnum
comfort	1:26:00::	a state of being relaxed and feeling no pain	physical state	VtEnum
consistency	1:07:00::	the property of holding together and retaining its shape	tactile property	VtEnum
contain	2:42:13::	contain or hold; have within	structurally related to	VtSpecificObject
day of the week	1:28:00::	any one of the seven days in a week	time period	VtEnum
depth	1:07:00::	the extent downward or backward or inward	visual property, measure	VtEnum
distance above	HXP	distance above	relative location	VtDistance
distance below	HXP	distance below	relative location	VtDistance
distance from	HXP	distance from	relative location	VtDistance
domestication	1:07:00::	the attribute of having been domesticated	physical state	VtEnum
environment	1:26:00::	the totality of surrounding conditions	physical state	VtEnum
extraordinary-iness	1:07:00::	the quality of being extraordinary and not commonly encountered	physical state	VtEnum
facing an object	HXP	position of facing something	position	VtSpecificObject
family member of	HXP	'family member of' relationship	social state	VtSpecificObject
far from	HXP	at a distance from	relative location	VtSpecificObject
father of	HXP	'father of' relationship	social state	VtSpecificObject
freshness	1:07:01::	the property of being pure and fresh (as if newly made); not stale or deteriorated	physical state	VtEnum
friend of	HXP	'friend of' relationship	social state	VtSpecificObject
fullness	1:26:00::	the condition of being filled to capacity	physical state	VtEnum
functional state	HXP	functional state	position or state	VtEnum
gender	1:07:00::	the properties that distinguish organisms on the basis of their reproductive roles	physical state	VtEnum

(table cont.)

health state	HXP	health state	physical state	VtEnum
height	1:07:00::	the vertical dimension of extension; distance from the base of something to the top	size	VtEnum
holding	HXP	be in a position of holding something	position	VtSpecificObject
hydration	1:22:00::	the process of combining with water; usually reversible	sustenance	VtEnum
illumination	1:26:01::	a condition of spiritual awareness; divine illumination	environment	VtEnum
in	HXP	used to indicate location	relative location	VtSpecificObject
in front of	HXP	in front of, not behind	relative location	VtSpecificObject
inside	HXP	on the inner side of	relative location	VtSpecificObject
length	1:07:00::	the linear extent in space from one end to the other; the longest dimension of something that is fixed in place	size	VtEnum
location	1:03:00::	a point or extent in space	physical state	VtEnum
made of	HXP	made of	physical state	VtEnum
measure	1:03:00::	how much there is or how many there are of something that you can quantify		VtEnum
mother of	HXP	'mother of' relationship	social state	VtSpecificObject
name	1:10:00::	a language unit by which a person or thing is known	physical state	VtSpecificObject
near	HXP	in close proximity	relative location	VtSpecificObject
next to	HXP	adjoining	relative location	VtSpecificObject
novelty	1:09:00::	originality by virtue of being new and surprising	extraordinariness	VtEnum
olfactory property	1:07:00::	any property detected by the olfactory system	physical state	VtEnum
on	HXP	used to indicate position in contact with and supported by the top or outer surface of	relative location	VtSpecificObject
on top of	HXP	over or upon	relative location	VtSpecificObject
other physical attributes	HXP	other physical attributes	physical state	VtEnum
outside	HXP	toward the exterior of	relative location	VtSpecificObject

(table cont.)

over	HXP	above	relative location	VtSpecificObject
parent of	HXP	'parent of' relationship	social state	VtSpecificObject
part of	HXP	part of	structurally related to	VtSpecificObject
personality	1:07:00::	the complex of all the attributes--behavioral, temperamental, emotional and mental--that characterize a unique individual	physical state	VtEnum
pitch	1:07:00::	the property of sound that varies with variation in the frequency of vibration	sound	VtEnum
position	1:07:01::	the arrangement of the body and its limbs	position or state	VtEnum
position or state	HXP	position or state	physical state	VtEnum
power	1:07:00::	possession of controlling influence	physical state	VtEnum
produced by	HXP	the action that produces this	physical state	VtSpecificObject
provide access to	HXP	provides a passage to get to	structurally related to	VtSpecificObject
relative location	HXP	a point or place in relation to another point or place	location	VtEnum
safety	1:26:00::	the state of being certain that adverse effects will not be caused by some agent under defined conditions	physical state	VtEnum
season	1:28:00::	one of the natural periods into which the year is divided by the equinoxes and solstices or atmospheric conditions	time period	VtEnum
shape	1:07:00::	any spatial attributes (especially as defined by outline)	physical state	VtEnum
shininess	1:07:00::	the visual property of something that shines with reflected light	visual property	VtEnum
sibling of	HXP	'sibling of' relationship	social state	VtSpecificObject
size	1:07:00::	the physical magnitude of something (how big it is)	visual property, tactile property	VtEnum
social state	HXP	relation to society	position or state	VtEnum
sound	1:07:00::	the auditory effect produced by a given cause	physical state	VtEnum

(table cont.)

strangeness	1:07:00::	unusualness as a consequence of not being well known	extraordinariness	VtEnum
strength	1:07:00::	the property of being physically or mentally strong	physical state	VtEnum
structurally related to	HXP	structurally related to	physical state	VtEnum
support	2:35:00::	be the physical support of; carry the weight of	structurally related to	VtSpecificObject
sustenance	1:13:00::	a source of materials to nourish the body	health state	VtEnum
tactile property	1:07:00::	a property perceived by touch	physical state	VtEnum
taste property	1:07:00::	a property appreciated via the sense of taste	physical state	VtEnum
teacher of	HXP	'teacher of' relationship	social state	VtSpecificObject
temperature	1:07:00::	the degree of hotness or coldness of a body or environment (corresponding to its molecular activity)	VtState	temperature
texture	1:07:00::	the feel of a surface or a fabric	tactile property	VtEnum
time of day	1:28:00::	clock time	time period	VtEnum
time period	1:28:00::	an amount of time	physical state	VtEnum
touch	2:35:01::	be in direct physical contact with; make contact	structurally related to, relative location	VtSpecificObject
towards	HXP	in the direction of	relative location	VtSpecificObject
under	HXP	not over	relative location	VtSpecificObject
visual property	1:07:00::	an attribute of vision	physical state	VtEnum
volume	1:07:02::	the magnitude of sound (usually in a specified direction)	sound	VtEnum
volume	1:23:00::	the amount of 3-dimensional space occupied by an object	size	VtEnum
weather	1:19:00::	the atmospheric conditions that comprise the state of the atmosphere in terms of temperature and wind and clouds and precipitation	environment	VtEnum
weight	1:07:00::	the vertical force exerted by a mass as a result of gravity	size	VtEnum
wetness	1:26:00::	the condition of containing or being covered by a liquid (especially water)	physical state	VtEnum

APPENDIX E: MENTAL STATES

As discussed in 4.1 Stative Predicates, HXP divides mental states into simple and complex. We list the simple mental states first, followed by the complex ones.

Simple Mental States

- afraid
- amused
- angry
- bored
- calm
- comfortable
- discomposed
- disgusted
- disloyal
- frustrated
- glad
- humiliated
- hungry
- interested
- pleased
- proud
- sad
- satisfied
- sleepy
- thirsty
- trustful
- unafraid
- wary
- wakeful

Complex Mental States

Since complex states can be ambiguous, they are displayed with definitions.

Word	Definition
able	(usually followed by `to') having the necessary means or skill or know-how or authority to do something
aware	(sometimes followed by `of') having or showing knowledge or understanding or realization or perception
believe	accept as true; take to be true
curiosity	a state in which you want to learn more about something
decide	reach, make, or come to a decision about something
desire	feel or have a desire for; want strongly
examine	observe, check out, and look over carefully or inspect
expect	regard something as probable or likely
feel	perceive by a physical sensation, e.g., coming from the skin or muscles
focus	direct one's attention on something
hear	perceive (sound) via the auditory sense
imagine	form a mental image of something that is not present or that is not the case
intend	have in mind as a purpose
know	be cognizant or aware of a fact or a specific piece of information; possess knowledge or information about
know	know how to do or perform something
know	be familiar or acquainted with a person or an object
like	be fond of
listen to	hear with intention
love	have a great affection or liking for
notice	notice or perceive
remember	recall knowledge from memory; have a recollection
see	perceive by sight or have the power to perceive by sight
smell	inhale the odor of; perceive by the olfactory sense
surprised	surprised by something
tactual sensation	feel the sensation produced by pressure receptors in the skin
taste	perceive by the sense of taste
think	judge or regard; look upon; judge
touch	perceive via the tactile sense
watch	look attentively

APPENDIX F: INSTITUTIONAL REVIEW BOARD APPROVAL

Application for Exemption from Institutional Oversight

LSU

Institutional Review Board
Dr. Robert Mathews, Chair
131 David Boyd Hall
Baton Rouge, LA 70803
P: 225.578.8692
F: 225.578.5983
irb@lsu.edu
lsu.edu/irb

Unless qualified as meeting the specific criteria for exemption from Institutional Review Board (IRB) oversight, ALL LSU research/ projects using living humans as subjects, or samples, or data obtained from humans, directly or indirectly, with or without their consent, must be approved or exempted in advance by the LSU IRB. This Form helps the PI determine if a project may be exempted, and is used to request an exemption.

- Applicant,Please fill out the application in its entirety and include the completed application as well as parts A-F, listed below, when submitting to the IRB. Once the application is completed, please submit two copies of the completed application to the IRB Office or to a member of the Human Subjects Screening Committee. Members of this committee can be found at http://research.lsu.edu/CompliancePoliciesProcedures/InstitutionalReviewBoard%28IRB%29/item24737.html

- A Complete Application Includes All of the Following:
  - (A) Two copies of this completed form and two copies of parts B thru F.
  - (B) A brief project description (adequate to evaluate risks to subjects and to explain your responses to Parts 1&2)
  - (C) Copies of all instruments to be used.
    - *If this proposal is part of a grant proposal, include a copy of the proposal and all recruitment material.
  - (D) The consent form that you will use in the study (see part 3 for more information.)
  - (E) Certificate of Completion of Human Subjects Protection Training for all personnel involved in the project, including students who are involved with testing or handling data, unless already on file with the IRB. Training link: (http://phrp.nihtraining.com/users/login.php)
  - (F) IRB Security of Data Agreement: (http://research.lsu.edu/files/item26774.pdf)

1) Principal Investigator: Jerry Weltman **Rank:** Graduate Student

Dept: Computer Science **Ph:** 225-907-9629 **E-mail:** jerry.weltman@gmail.com

2) Co Investigator(s): please include department, rank, phone and e-mail for each
*If student, please identify and name supervising professor in this space

IRB# E7006 LSU Proposal #
✓ Complete Application
✓ Human Subjects Training

IRB #E7006

3) Project Title: Toward digitizing the human experience

Study Exempted By:
Dr. Robert C. Mathews, Chairman
Institutional Review Board
Louisiana State University
203 B-1 David Boyd Hall
225-578-8692 | www.lsu.edu/irb
Exemption Expires: 9/17/2015

4) Proposal? (yes or no) no **If Yes, LSU Proposal Number**

Also, if YES, either
○ **This application completely matches the scope of work in the grant**
OR
○ **More IRB Applications will be filed later**

5) Subject pool (e.g. Psychology students) adults (> 18) interested in linguistics

*Circle any **"vulnerable populations" to be used:** (children <18; the mentally impaired, pregnant women, the ages, other). Projects with incarcerated persons cannot be exempted.

6) PI Signature [signature] **Date** 9/17/12 (no per signatures)

** **I certify my responses are accurate and complete.** If the project scope or design is later changes, I will resubmit for review. I will obtain written approval from the Authorized Representative of all non-LSU institutions in which the study is conducted. I also understand that it is my responsibility to maintain copies of all consent forms at LSU for three years after completion of the study. If I leave LSU before that time the consent forms should be preserved in the Departmental Office.

Screening Committee Action: Exempted ✓ Not Exempted ___ Category/Paragraph 2

Signed Consent Waived?: Yes / No

Reviewer Mathews **Signature** [signature] **Date** 9/19/12

VITA

Jerry Scott Weltman was born in Fort Worth, Texas. He received his Bachelor of Arts degree in computer science from the University of Texas at Austin in 1984 and received his Master of Science degree in computer science from Stanford University in 1985. Subsequently, he worked as a software developer at several research and development companies, including Bell Labs (where he earned a patent for an telephone signaling protocol) and Innovative Emergency Management in Baton Rouge. He is an award-winning instructor at LSU, having taught C programming and developed the first LSU computer science object-oriented programming course. As a researcher in natural language processing and artificial intelligence, his interests span the fields of computer science, linguistics, philosophy of language, and cognitive psychology. His specific research areas are natural language processing, story understanding, commonsense modeling, and knowledge acquisition, as well as more general linguistics topics in semantics and pragmatics.