Evaluation, Orientation, and Action in Interactive StoryTelling

Elahe Rahimtoroghi, Reid Swanson, Marilyn A. Walker and **Thomas Corcoran**
Natural Language and Dialogue Systems Lab
University of California Santa Cruz, Santa Cruz, Ca. 95064, USA
{elahe, reid, maw}@soe.ucsc.edu, tcorcora@ucsc.edu

Abstract

Interactive storytelling encompasses a wide variety of applications and systems that aim to allow players in a virtual world to affect the narrative progression prescribed by the author. The story world is often modeled by a set of hand-crafted causal and temporal relationships that define the possible narrative sequences based on the user's actions. In our view, a more promising approach involves applying methods from statistical language processing to create data-driven stories. Here, we explore integrating a typology of narrative clauses from Labov & Waletzky's theory of oral narrative (henceforth L&W) into the *Say Anything* data-driven storytelling approach. We show that we can automatically distinguish EVALUATION clauses from ORIENTATION and ACTION clauses with 89% accuracy in fables, suggesting that it will be possible to develop new types of data-driven stories using L&W's typology.

Introduction

Interactive storytelling encompasses a wide variety of applications and systems that aim to allow players in a virtual world to affect the narrative progression prescribed by the author. The story world is often modeled by a set of causal and temporal relationships that define the possible narrative sequences based on the user's actions. These relationships are generally domain specific and authored by hand using a formal representation language, such as predicate calculus, planning operators or graphical models. While these approaches have been used to create compelling experiences for entertainment (Callaway and Lester 2002), education (Mott and Lester 2006), training (Traum et al. 2007) and health care (Bickmore and Schulman 2006) they are difficult to author and have problems scaling beyond small toy domains.

In our view, a more promising approach involves applying methods from statistical language processing to create data-driven stories. For example, there has been a growing interest in learning causal and temporal knowledge about stories from data (Chambers and Jurafsky 2009; Riaz and Girju 2010; Beamer and Girju 2009; Gordon and Swanson 2009; Manshadi, Swanson, and Gordon 2008; Gordon, Bejan, and Sagae 2011; Goyal, Riloff, and Daumé III 2010), including the idea of extracting narrative schemata or scripts from large amounts of text (Chambers and Jurafsky 2008; 2009; Schank and Abelson 1977) and also using linguistic features to encode sentences and create annotated narrative datasets (Elson 2012; Finlayson 2011). *Say Anything* (Swanson and Gordon 2012) was the first interactive storytelling system to use a large-scale data-driven approach to predict what comes next in a player's story without any prior knowledge of the user's input. It does this by mining appropriate sentences from a large corpus of over 1 million personal stories extracted from Internet weblogs. One of the limitations of this type of approach so far, however, is that they do not distinguish between different types of narrative clauses that can be used to structure a story in different ways.

Here, we explore integrating a typology of narrative clauses from Labov & Waletzky's theory of oral narrative (henceforth L&W) into the *Say Anything* data-driven storytelling approach. We adopt L&W's distinction between the ACTION clauses that describe story events, and the ORIENTATION and EVALUATION clauses that structure the narrative by providing its setting and the reason for telling the story.

We chose the narrative categories of Labov & Waletzky for two primary reasons. First, we believe that any narrative discourse typology should have at least these three types to provide useful analysis. Second, personal stories found on the web are often noisy, difficult to interpret and do not always clearly follow well defined narrative conventions. A deep analysis and annotation scheme, such as the one employed by DramaBank (Elson 2012) that is based on and extends Lehnert's plot units (Lehnert 1981), offers many advantages. However, acquiring such a detailed level of analysis on user generated content, such as blog stories, is resource intensive and difficult to achieve high inter-rater agreement. Instead we believe it is more promising to first demonstrate the feasibility of a simplified but more general schema that applies across multiple domains where we can obtain adequate levels of agreement. This will allow us to scale the annotation process to acquire sufficient training data for learning and evaluation.

As a motivating example, consider the story produced by *Say Anything* (Swanson and Gordon 2012) shown in Table 1 where each story contribution is labeled with a L&W cate-

#	Category	Story Clause
1	Action	The car veered off the road.
2	Action	When I came to, I immediately pulled the wheel away from the ditch I was headed towards, but I did so too quickly and began to slide the other directs.
3	Action	I was headed right for a tree.
4	Action	The people around me SCREAMED.
5	Action	I slammed into the tree, totaling my car.
6	Action	After a bit I was able to breathe but it was really hard to difficult...
7	Evaluation	I felt faint, saw stars and was in an incredible amount of pain!
8	Action	The bystanders ran to my rescue.
9	Evaluation	I was just about to throw myself.
10	Action	Luckily they were able to pull me out before the car caught fire.

Table 1: An example story written with *Say Anything* that has been manually annotated with the Labov & Waletzky labels. User authored sentences are highlighted in gray, while the computer generated sentences are highlighted in white.

gory. The selection of the next system utterance in *Say Anything* is based in large part on predicting what is a likely next event in a chain of narrative events. The algorithm is based on the observation that most personal narratives are told in chronological order (as are Aesops Fables and film scripts). Given the user's input, *Say Anything* currently finds the most similar event in the corpus and returns the next sentence. It does this under the assumption that the next sentence will be the next event in the story. However, even though the stories are told in chronological order, the next sentence may not be describing an event. Instead, it could be an evaluation or orientation as in the case of $6 \rightarrow 7$ and $8 \rightarrow 9$.

While evaluations of *Say Anything* suggest that its current techniques work to some degree, our hypothesis is that there are at least two ways that automatic labeling of clauses or sentences according to L&W's categorization could help interactive storytelling, and *Say Anything* in particular. First, it would provide some structure to the stories *Say Anything* produces. Second, it could enable new types of interactive storytelling applications. For example, *Say Anything* is based on the assumption that the user and computer collaborate by writing sentences of an emerging story, and that they are symmetric collaborators performing essentially the same task. However, if we could label the sentences with the L&W categories, we might also be able to build asymmetric collaborative agents that could elicit stories from users by recognizing what should come next or what is missing.

In this paper we show the viability of applying Labov & Waletzky's typology of narrative clauses to written narratives and the ability to automatically label them with high accuracy. In the next section we describe the L&W categories in more detail and discuss how they could be used to improve data-driven interactive storytelling systems. Then we describe our methodology for automatically labeling narrative clauses. Afterwards, we discuss our experimental results, and finally conclude with a discussion of our results and plans for future work.

L&W and Data-Driven Interactive Storytelling

Labov & Waletzky's theory of oral narrative divides narrative clauses into three main sections: temporal types of narrative clauses, structural types of narrative clauses, and evaluation points in narratives (Labov and Waletzky 1997; Labov 1997). In this paper we focus on the structural types and evaluation points of the stories, because they most clearly apply to written, as well as the oral narratives the original theory was based upon. Of all the L&W categories these also provide the most useful structure for enhancing the generative power of *Say Anything* (Swanson and Gordon 2012) and open the possibility for novel types of interactive storytelling applications in this framework.

Orientation and *action* are two structural types of clauses. An orientation clause introduces the time and place of the events of the story, and identifies the participants of the story and their initial behavior. A clause of action reports an event of the story. An *evaluation* clause provides evaluation points of a narrative and information on the consequences of the events as they relate to the goals and desires of the participants. Evaluation clauses also describe the events that did not occur, may have occurred, or would occur in the future in the story. Thus, a narrative clause in *irrealis* mood is an evaluation clause. Irrealis mood is an event or proposition in the narrative that has not actually occurred at, or before, the time the utterance was made as far as the narrator is aware. For example, conditional statements, such as the one in the sentence

Maybe I should call her and make amends.

from Table 5 describes an action in an alternate reality than the one actually taking place.

Additionally, L&W define two other structural types that we do not currently use: *abstract* and *coda*. Abstract is an initial clause in a narrative that reports the entire sequence of events. A coda is defined as a final clause which returns the narrative to the time of speaking, indicating the end of the narrative. The definition for the abstract clauses is ambiguous to some extent; and there is no single clause reporting all the events in either Aesop's Fables or the film scripts. Coda is the last part of an *oral* narrative, when the narrator finishes the story; and our datasets do not consist of oral narratives.

According to these definitions from L&W, we define and extract structural features from clauses and apply a text classification method to distinguish between different types of clauses in the stories. Table 2 shows an example of Aesop's Fables annotated according to L&W categories. Three types of narrative clauses, *orientation*, *action*, and *evaluation* are identified manually in this fable.

Consider again the example story from *Say Anything* (Swanson and Gordon 2012) shown in Table 1. This story was a highly rated story written with *Say Anything*. Alternating rows indicate turns taken by the system. Rows highlighted in gray were written by the human user, while computer generated sentences are highlighted in white. The

Category	Story Clause
Orientation	A hungry Fox saw some fine bunches of Grapes hanging from a vine
Orientation	that was trained along a high trellis,
Action	did his best to reach them by jumping as high as he could into the air.
Evaluation	But it was all in vain,
Evaluation	for they were just out of reach:
Action	so he gave up trying,
Action	and walked away with an air of dignity and unconcern,
Action	remarking,
Evaluation	"I thought those Grapes were ripe,
Evaluation	but I see now they are quite sour."

Table 2: *The Fox and the Grapes.* An example of Aesop's Fables annotated with L&W categories.

sentences were manually segmented into phrases and annotated with the L&W labels to help show the underlying structure of the narrative and illustrate how they could be used to improve the generation process.

This example demonstrates a relatively successful interaction that exhibits some beneficial narrative properties. The story starts with a user contributed sentence that describes the current state of affairs. It progresses through a series of complicating actions to a climax. In nearly all cases, with the exception of the last sentence, the user input describes an active event in this progression. The system is able to respond with an appropriate event or evaluation that continues the story in a coherent manner.

Although the story follows basic narrative structure, much of this success is coincidental or primarily imposed by the user's actions. For example, the system has no knowledge of how to start a story and is always initiated by a user. In this example story very little orientation is provided. The system has no model to understand that background information about the characters and scene is important or that this information generally goes early in the narrative. Similarly, the system has no model of what it means for a story to reach a conclusion. The user must either wait for an appropriate concluding sentence to appear by chance or write one of their own, which they have done here.

During the rising action, the system is able to select events that progress the story in a natural way and provides some evaluation clauses indicating the significance of these events to the narrator and narrative. The ability of *Say Anything* to intersperse evaluations into the story is an attribute that is not shared by many other interactive narrative systems. However, the generation of these clauses is uncontrollable by the system and only appears as a supervenient property due to the structure of the underlying architecture.

To assess more broadly how stories written with *Say Anything* are structured according to these categories, we annotated the 15 most highly rated stories for each model (*retrieval*, *reranking* and *adaptation*) described in Swanson & Gordon (Swanson and Gordon 2012) with these three labels. This resulted in 44 stories (there was one duplicate) and 489 narrative clauses that were annotated. Table 3 shows the distribution of each of the three labels. This table shows the distribution for all clauses and also for each type of author: user and computer. All distributions show that the majority of clauses in these stories are actually *evaluation* with computer generated sentences even more skewed in this direction.

Label	Total	User	Computer
Orientation	24.74	28.57	20.18
Action	20.04	23.68	15.70
Evaluation	55.21	47.74	64.13

Table 3: The distribution of L&W labels as a percentage of the number of clauses for each set.

Label	Retrieval	Reranking	Adaptation
Orientation	23.21	27.45	23.81
Action	24.40	20.92	14.88
Evaluation	52.38	51.63	61.31

Table 4: The distribution of L&W labels as a percentage of the number of clauses in each model.

We also investigated the label distributions for each of the three different model types, which are shown in Table 4. The higher performing models (reranking and adaptation) also show a greater propensity toward evaluation and contain fewer action clauses overall. We find these results interesting, because they highlight a problem with one of the assumptions of *Say Anything* and suggest avenues for improvement in the system. Although *Say Anything* assumes (correctly) that personal narratives are generally told in temporal order, it is not necessarily the case that events are located, or should be located, in consecutive sentences. This incorrect assumption did not seem to affect the high level evaluation concerning the quality and coherence of the generated stories, but it significantly changes the interpretation for why these narratives were judged highly by the raters in the first place.

The ability to control which type of clause the system generates next would allow testing of more specific aspects of the model. For example, only generating action clauses to test the system's ability to generate coherent causal sequences. The ability control when, how much, and what type of evaluation to provide would also allow the stories to be told from a specific, predetermined point of view. For example, including only negative or positive polarity evaluations could be used to give a strong voice to the narrator. Consider the difference in the tone and interpretation of the narrative if the second row of Table 5 had been:

It was entirely the fault of the idiot driver next to me.

Additionally, individual users have different demographic backgrounds, life experiences, and personalities. These characteristics influence what the user finds important in the story and their willingness to engage in sharing their experiences (Thorne 1987). Information about the user's profile could be used to selectively choose evaluations that target an

Category	Story Clause
Evaluation	I can't remember how long it has been since I have seen my sister.
Orientation	So much has happened in that time.
Orientation	I got married and had 6 children.
Evaluation	I have a wonderful life.
Evaluation	I just want my sister to be a part of it.
Action	My memory fails me here...
Action	I can't even remember what we had fought about.
Evaluation	O.o hmm...
Action	I remember now.
Orientation	It was over a stupid boy.
Evaluation	Jeez.
Evaluation	That is not a valid reason to abandon your sister.
Evaluation	Don't know...
Evaluation	Maybe I should call her and make amends.
Evaluation	I don't know if I can.
Orientation	At that very moment my phone rang and it was my sister.
Evaluation	What a wonderful surprise!
Action	We talked for hours as if nothing had ever come between us.
Evaluation	It was great.

Table 5: An example story from *Say Anything*. Sentences written by the user are highlighted in gray and the system generated sentences are shown in white.

individual's interests and help elicit more meaningful narrative content from the user. Leveraging this personal information to tailor which evaluative clauses are selected could greatly improve the enjoyment of the interaction and satisfaction with the completed narrative artifacts.

The style of the second story, in Figure 5, provides a contrast to the first. It also scored highly by raters, but appears to do so for different reasons. In this story, most of the computer generated responses are evaluative. These clauses tend to be more general and are applicable in more contexts than actions, which typically have stronger causal constraints. The system is succeeding in this case, not by good causal knowledge, but by implicitly relying on the broadness of evaluative statements. Having the ability to automatically identify and generate specific types of sentences would allow us to test this hypothesis and perform a more detailed error analysis of the system. The distribution of labels in this story also makes clear the difference in expository style from the first example. This story reads much more like a conversation or journal entry.

Automatically labelling of L&W categories would allow category labels to be provided to the user along with the next utterance selections that *Say Anything* makes in order to provide some high level structure. This could be used for relatively simple improvements, for example, biasing towards orientation sentences at the beginning of a story. However, it would also enable the possibility of learning and reproducing writing styles.

No.	Feature	Example
1	Stative verbs	There **was** once a house,
2	Non-stative verbs	and then **climbed up** the wall,
3	Future	you **will** certainly be found out,
4	Conditional	that she should pay him a high fee **if** he cured her,
5	Quotes	**"**That's awkward,**"**
6	Questions	who is going to bell the cat**?**
7	Indefinite articles	**A** hungry Fox saw some fine bunches of Grapes
8	Time entity	a Goose which laid a Golden Egg **every day**.

Table 6: Structural features used for clause classification.

Classification Method

In this section we describe our methodology for automatically labeling narrative clauses with the three L&W categories described above: *orientation*, *action*, and *evaluation*. To label individual clauses we used a standard text classification (Langley, Iba, and Thompson 1992) approach. Text classification is a general problem of labeling a segment of text into one of several predefined categories. Following other text classification approaches, we used a supervised machine learning algorithm trained and applied on a hand labeled dataset. In this work we used a Naïve Bayes classifier, which is a simple probabilistic model. Similar to other classifiers, it represents a data instance as a mapping $\boldsymbol{x} \mapsto y$. $\boldsymbol{x}$ is a feature vector representation of the clause and y is a scalar value indicating whether the clause belongs to the desired class. We applied binomial model of Naïve Bayes classifier which is common for these types of tasks, because it generally performs well with short text fragments and small numbers of features (McCallum, Nigam, and others 1998).

Text classification systems often use surface lexical tokens, i.e., the words in the document, as features. These types of features generally require a large amount of training data due to the large dimensionality and make it difficult to recognize the underlying properties separating the classes. In this paper we propose a set of structural features that more closely align with the high level analysis of L&W and do not require a large amount of training data. The eight features we propose are shown in Table 6. In addition, an example clause is provided for each feature showing its use in context.

Stative verbs can indicate non-action verbs that generally describe some property of the world over some period of time. More precisely Dowty (Dowty 1979) defines a stative verb as one that adheres to the following 3 criteria:

1. They cannot occur in the progressive form.
2. They cannot occur as the complement to causatives.
3. They cannot form imperatives.

Our algorithm identifies stative verbs using a list of predefined verbs which are exclusively stative, such as *enjoy,* *like*, and *be*. The list is based in part on the stative definition. A Non-stative verb feature is set to 1 when the clause contains a verb not included in the stative list. Future, conditional, quotes and questions are instances of *irrealis* mood which indicate evaluative points of the narrative. Indefinite articles can show an introductory clause which is a part of the orientation. An orientation also describes the time of the story, thus a time entity can indicate an orientation clause. However, we observe that clauses from same category do not share common lexical features, for example consider rows 3 and 5 in table 6. After annotation, both clauses are labeled as evaluation while they are not lexically similar. But they both have structural features associated with evaluative points: *future* and *quotes*. According to these observations, we define and use several structural features for classifying clauses of the stories instead of literal words.

Evaluation and Results

We evaluated our system using a hand labeled gold standard dataset. For this evaluation we used a collection of *Aesop's Fables* (Jones 1912). We chose this dataset, instead of blog stories or other narratives more similar to *Say Anything*, for several reasons. First, the narratives are relatively short and well defined. Nearly every clause in these fables serves a purpose to the narrative being told, which is not true of the blog stories of *Say Anything* or other narrative data sources such as film scripts. Second, we expect a high level of agreement between annotators on this dataset compared with informal narratives found on the web. Third, the language is relatively simple and easy to extract the structural features from.

Aesop's Fables are also used in other previous work on encoding narrative text (Goyal, Riloff, and Daumé III 2010; Elson 2012), which allows a more direct comparison of the reliability and expressive power between our approach these alternatives. We consider these experiments a best case scenario to establish the viability of this approach on simple well structured stories and provides a baseline for further experiments.

We annotated the first 20 fables from the collection cited above. The fables were segmented into individual clauses and each one was annotated into one of the L&W categories. Inter-annotator agreement was measured using percentage of agreement and Kappa statistics (Cohen 1960). We obtained 89.6% of agreement between two annotators and the kappa coefficient was 0.816 which represents almost perfect agreement (Landis and Koch 1977). After annotating, there were 315 total clauses including 20 orientation, 167 action and 128 evaluation clauses. Each clause was processed by the Stanford CoreNLP toolkit[1] (Finkel, Grenager, and Manning 2005) to help identify the structural features of the clauses.

To evaluate the performance of our classifier we performed a 5-fold cross-validation experiment. We use the standard evaluation metrics precision, recall and accuracy to assess the quality of our classifier. The results of our experiments are provided in Table 7. We achieve a high accuracy and precision for all label types. Recall is also near or above 90% for action and evaluation clauses, but is substantially lower (45%) for orientation clauses. It is sometimes difficult to distinguish orientation clauses from evaluation and the position in the story can help distinguish between the categories, since orientation is usually expected toward the beginning of a narrative. Our classifier does not take into account the relative position of clauses, which may be affecting the performance.

Measure	Orientation	Action	Evaluation
Accuracy	0.96	0.90	0.92
Precision	0.90	0.88	0.90
Recall	0.45	0.93	0.89

Table 7: A summary of the 5-fold cross-validation of our classifier on the fables dataset.

To gain some perspective on our results we also compared our system to a *majority class* baseline. This baseline classifier always chooses the category with the maximum number of instances. Over all the categories, this baseline achieves an accuracy of 0.531. Compared with an accuracy of 0.885 with our Naïve Bayes classifier, we achieve a 66% improvement.

Discussion and Future Work

In this paper we have described how the Labov and Waletzky typology of narrative clauses can be used to analyze written narrative and improve the level of narrative understanding in interactive storytelling systems. There are at least two ways that automatic labeling of clauses or sentences into ACTION, ORIENTATION, and EVALUATION categories could improve data-driven interactive storytelling. First, it would provide some semantic structure to the generation process. Second, it could enable new types of interactive storytelling applications, such as building asymmetric collaborative agents. We imagine new agents that could elicit stories from users by recognizing what type of utterance should come next based on particular writing styles or user preferences. Or alternatively writing agents that help identify poorly formed narrative structure, missing events or unexplained consequences that require evaluation.

We demonstrated the viability of automatically labeling textual stories with these categories using a subset of Aesop's fables that were manually annotated, which achieved an overall accuracy of 89%. These results indicate that our structural features are capable of discriminating these categories on simple, well defined narrative clauses. Given the high accuracy on this data, we plan to annotate other sources of narratives including blog stories and film scripts that are closer to our target domain. Identifying additional indicators and applying them in classification, such as *positions of the clauses* in narrative and *verb tenses*, is another direction of our future works.

[1] http://nlp.stanford.edu/software/corenlp.shtml

References

Beamer, B., and Girju, R. 2009. Using a bigram event model to predict causal potential. In *Computational Linguistics and Intelligent Text Processing*. 430–441.

Bickmore, T., and Schulman, D. 2006. The comforting presence of relational agents. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, 550–555.

Callaway, C., and Lester, J. 2002. Narrative prose generation. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, volume 139(2), 213–252.

Chambers, N., and Jurafsky, D. 2008. Unsupervised learning of narrative event chains. *Proceedings of ACL-08* 789–797.

Chambers, N., and Jurafsky, D. 2009. Unsupervised learning of narrative schemas and their participants. In *Proceedings of the Association of Computational Linguistics (ACL)*.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1):37–46.

Dowty, D. R. 1979. *Word meaning and Montague grammar: The semantics of verbs and times in generative semantics and in Montague's PTQ*. Dordrecht Reidel.

Elson, D. 2012. Dramabank: Annotating agency in narrative discourse. In *Proceedings of the English International Conference on Language Resources and Evaluation (LREC 2012)*, 2813–2819.

Finkel, J. R.; Grenager, T.; and Manning, C. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 363–370.

Finlayson, M. A. 2011. Corpus annotation in service of intelligent narrative technologies. In *Proceedings of the 4th Workshop on Intelligent Narrative Technologies*.

Gordon, A., and Swanson, R. 2009. Identifying personal stories in millions of weblog entries. In *Third International Conference on Weblogs and Social Media, Data Challenge Workshop*.

Gordon, A.; Bejan, C.; and Sagae, K. 2011. Commonsense causal reasoning using millions of personal stories. In *Twenty-Fifth Conference on Artificial Intelligence (AAAI-11)*.

Goyal, A.; Riloff, E.; and Daumé III, H. 2010. Automatically producing plot unit representations for narrative text. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 77–86.

Jones, V. S. 1912. *Aesop's Fables: A New Translation*. Avenel Books.

Labov, W., and Waletzky, J. 1997. Narrative analysis: Oral versions of personal experience. *Journal of Narrative and Life History* 3–38.

Labov, W. 1997. Some further steps in narrative analysis. *Journal of narrative and life history* 7:395–415.

Landis, J. R., and Koch, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics* 159–174.

Langley, P.; Iba, W.; and Thompson, K. 1992. An analysis of bayesian classifiers. In *Proceedings of the national conference on artificial intelligence*, 223–223.

Lehnert, W. G. 1981. Plot units and narrative summarization. *Cognitive Science* 5(4):293–331.

Manshadi, M.; Swanson, R.; and Gordon, A. S. 2008. Learning a probabilistic model of event sequences from internet weblog stories. In *Proceedings of the 21st FLAIRS Conference*.

McCallum, A.; Nigam, K.; et al. 1998. A comparison of event models for naive bayes text classification. In *AAAI-98 workshop on learning for text categorization*, volume 752, 41–48.

Mott, B., and Lester, J. 2006. Narrative-centered tutorial planning for inquiry-based learning environments. In *Eighth International Conference on Intelligent Tutoring Systems*, 675–684.

Riaz, M., and Girju, R. 2010. Another look at causality: Discovering scenario-specific contingency relationships with no supervision. In *Semantic Computing (ICSC)*, 361–368.

Schank, R. C., and Abelson, R. 1977. *Scripts Plans Goals*. Erlbaum.

Swanson, R., and Gordon, A. S. 2012. Say anything: Using textual case-based reasoning to enable open-domain interactive storytelling. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2(3):16.

Thorne, A. 1987. The press of personality: A study of conversations between introverts and extraverts. *Journal of Personality and Social Psychology* 53(4):718.

Traum, D.; Roque, A.; Georgiou, A. L. P.; Gerten, J.; Narayanan, B. M. S.; Robinson, S.; and Vaswani, A. 2007. Hassan: A virtual human for tactical questioning. In *Proceedings of the 8th SIGDial Workshop on Discourse and Dialogue*, 71–74.