DERIVING NARRATIVE MORPHOLOGIES VIA ANALOGICAL STORY MERGING

Mark Alan Finlayson
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
32 Vassar St. Room 32-258, Cambridge MA, 02138 USA
markaf@mit.edu

Abstract

It has long been suspected that stories drawn from the same cultural setting share important narrative structure. One example of such structure, first identified by Vladimir Propp in 1928, is the *morphology* of a set of stories, which describes the set of plot elements and their allowed sequences. Until now, the extraction of morphologies has remained a manual task, the purview of anthropological virtuosos (e.g., Propp, Lévi-Strauss, Dundes, Campbell); reproduction or validation of their analyses is a time-consuming, prohibitively difficult endeavor. I demonstrate a technique called *Analogical Story Merging* that derives a morphology given a set of stories. It incorporates Forbus & Gentner's Structure Mapping Engine as well as a method called *Bayesian Model Merging* for inducing a grammar from a corpus of positive examples. I present the output of the basic implementation when applied to a small example story corpus, a set of summaries of Shakespearean plays.

Keywords: Analogical story merging; Analogical mapping; Analogical generalization; Narrative morphology; Folktales; Vladimir Propp; Structure Mapping Engine; Bayesian model merging

Understanding narrative and narrative processes is critical to achieving a complete grasp of human cognition. In particular, narrative's role in transmitting culture-specific information such as values, norms, and beliefs is of special and long-standing interest. It is clear that different cultures maintain different sets of assumptions about the world that affect participants' interpretation and understanding, and the stories that are prevalent in a culture seem to be important to the transmission of such assumptions. Variously called folktales, fairy tales, fables, or myths, these narratives are passed down through generations of retelling and are thought to be subject to a Darwinian-like natural selection process, in which portions of the narratives that are congruent with the culture are retained and amplified, and those that are incongruent are distorted or discarded. Such an effect was famously shown *en miniature* by Bartlett (1920).

One strong candidate for culture-specific information that may be folded into cultural narratives is a common framework for story plots. Readers of even a moderate number of tales from a specific culture will note the habitual recurrence of certain plot elements, such as preparatory sequences, motivating events, main actions, and end-states. Examples of patterns of repeated plot elements include the *Three Brothers* in Slavic cultures, the *Trickster* in North American native cultures, or *Cinderella* in the West. These sequences may reflect information such as the culture's assumptions regarding cause and effect, the proper response to various situations, important life goals, the constituents of a good or happy life, or the qualities to be demonstrated by a heroic individual. Vladimir Propp presented such a framework in his seminal work *The Morphology of the Folktale (1968),* in which he identified what he called the *morphology* of a group of Russian hero tales – a set of narrative pieces (which he called *functions*) and their subtypes along with what was essentially a grammar for combining them into stories.

Following in Propp's footsteps, both Dundes (1964) and Colby (1973) worked out partial morphologies for, respectively, Native American Indian and North Alaskan Eskimo folktales. What they found was that the morphology for each culture was distinct, but similar to other known morphologies in important ways. For example, all morphologies seem to share a common high-level structure (preparation, motivation, main actions, conclusion), and have significant overlap in rough identity of the functions, but vary considerably in specific function sequences and other details.

It would be of wide-ranging interest if the morphology of a culture could reliably be exposed to scientific investigation. If morphologies actually have psychological reality, they potentially would be a powerful window into cultural thought. Until now the extraction of morphologies has remained a manual task, the purview of anthropological virtuosos. Constructing a morphology for a particular culture takes many years of reading and analysis, and once complete, it is unclear how much the morphology owes to the folklorist himself or his familiarity with other extant morphologies, rather than truly reflecting the character of the tales under investigation. Furthermore, reproduction or validation of a morphological analysis is a time-consuming, prohibitively difficult endeavor.

Here I demonstrate a technique called *Analogical Story Merging* that for the first time gives computational purchase on the problem of identifying a morphology from a given set of stories. The algorithm, which has been implemented in a computer program, takes as input a corpus of stories, the semantics of which have been encoded in a computer-readable representation. The algorithm incorporates an analogical mapper – in particular, the Structure Mapping Engine (Falkenhainer, Forbus, & Gentner, 1989) – to determine similarity between portions of different stories. It also makes use of a method called Bayesian model merging (Stolcke & Omohundro, 1994) for measuring the fit of the derived morphology to the story corpus.

The progression of the paper is as follows. First I explain model merging, which serves as the conceptual framework for the algorithm. Following that, I explain the logic behind Analogical Story Merging, and present the output of the basic implementation when applied to a small story corpus, a set of five summaries of Shakespearean plays. Finally I discuss the technique's parameter space, what infrastructure it requires to truly capitalize on its power, and what new types of experiments and investigations it enables.

Model Merging

Model merging forms the conceptual foundation of Analogical Story Merging. Model merging is used to derive a grammar or model from a set of examples. Consider the set of two characters sequences `{ab,abab}`. What is the pattern that explains these two sequences? One plausible guess is `ab` repeated one or more times, or, written as a regular expression, `(ab)+`. Model merging allows us to find this pattern given the sequences. For the purposes of this work, we will only consider the Bayesian variety of model merging, and only as applied to Hidden Markov Models (HMMs). The technique is illustrated in Figure 1, and proceeds as follows.[1] First, we generate an initial HMM, called M_0 in the figure, by incorporating each of our examples explicitly into the model. This HMM has six states, each of which will emit a single character, either `a` or `b`, and will generate with 50% probability either the sequence `ab` (top branch) or the sequence `abab` (bottom branch). Second, we define a prior over models, which is a probability mass function for how probable a model is *a priori*. This example uses a

[1] Attentive readers will note that the models presented in the figure are not written as HMMs, with hidden and observable states, but rather as regular Markov Models. This is purely to promote clarity of the figure – no state emits both symbols, and so the observable states are left implicit. The states shown in the figure are hidden states, and only the emitted symbols are directly observable.

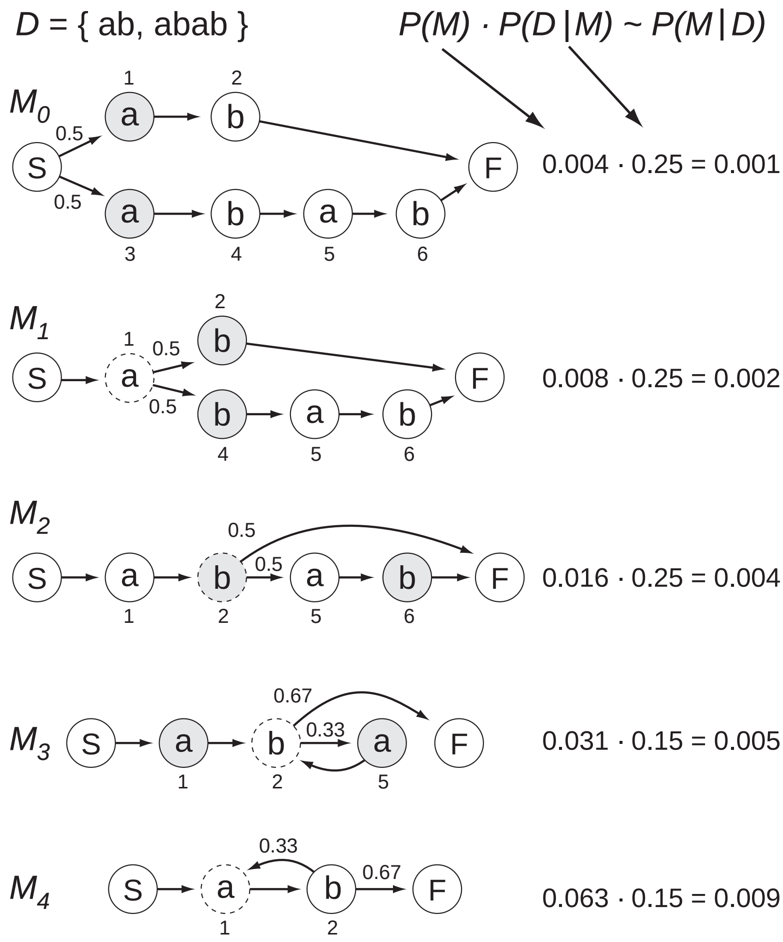

Figure 1: Example of Bayesian model merging over an HMM; after Figure 1 from (Stolcke & Omohundro, 1994). The original character sequences are labeled *D*. States are represented by a circle, and transitions by arrows. The symbols (either a or b) emitted by a state are listed inside its circle. The symbols *S* and *F* indicate start and final states, respectively. States are numbered, and these appear outside each state's circle. States shaded in one step are merged into the dashed state in the next step. The probabilities of the prior (geometric, *p=0.5*) and the data given the model are given on the right, along with their product, which is proportional to the posterior probability. The merging proceeds until the product cannot be further increased.

geometric distribution, which formalizes the intuition that smaller models are more likely. Using the model itself we can calculate the probability of seeing the observed data (it is 25% for M_0). Using Bayes' rule, we can calculate a number proportional to the posterior, which is the probability of the model given the observed data. Finally, we search the space of *state merges*, where we merge two states by removing them from the model and replacing them with a single state that inherits their transitions and emissions. The search is driven by trying to the find the model that maximizes the posterior. In the figure, states shaded in one step are merged together into the dashed state in the next step. The figure shows the step-by-step progression (found by search) from the initial model through a series of merges to the final model that maximizes the posterior. The first merge combines states 1 and 3 generating model M_1. This model still only produces the original two examples, but it is smaller than M_0, so it is more probable. The second merge combines states 2 and 4 to produce M_2. Again, no change in the output, but a smaller model. The third merge combines states 2 and 6 to produce M_3, and in this step generalization occurs, in that the HMM can now produce any string that matches `(ab)+`. This reduces $P(D|M)$, the probability of the data given the model, but by not as much as is gained from the increase in the prior. The final merge produces M_4, a smaller model, and any further merging causes a reduction in the posterior.

Analogical Story Merging

Analogical Story Merging (ASM) is a special form of model merging, where the model in question is a morphology of stories. A morphology will be treated for the purposes of this paper as an HMM-like structure. Propp's morphology consists of 34 states (functions) with a set of allowed transitions. For example, if a story contains the function *interdiction to the protagonist*, it must proceed at some point to the *violation of the interdiction* function. The first occurs before the second, and the second cannot occur without the first. They both come before any states that define the main action and resolution of the story (e.g., *acquisition of the magical item*, *battle with the villain*, *marriage of the hero*, etc.). Analogical Story Merging proceeds along the same line as regular Bayesian model merging over HMMs. The steps are as follows:

1. Construct the Initial Morphology

The initial morphology is constructed by extracting a sequence of states from each story. Each sequence of states is incorporated into the initial morphology as a single, linear branch. For the example presented, I define each state to be a single event in the story, and their order to be the order in which they occur in the story timeline. An initial morphology, labeled M_0, can be seen at the top of Figure 2, where each of the two simple example stories with their four constituent events is transformed into a sequence of four states. Causal and temporal linkages are ignored for the purposes of this example, but it will be explained later how they need not be left from consideration.

2. Define the Merge Operation

After the initial morphology is constructed, the algorithm requires a merge operation over states. Such an operation takes two states and replaces them by a single state. For the example presented, I closely follow Stolcke & Omohundro, where the merged state inherits the weighted sum of the transitions and emissions of its parent. The representations contained in each event are atomic for the purposes of merging, so each state can emit any event that is agglomerated into it. That is, if a state S_1 emits only event *A*, and state S_2 emits only event *B*, then for the purposes of calculating $P(D|M)$, the merged state has a 50% chance of emitting either *A* or *B*.

3. Define the Prior over Morphologies

To perform its search the algorithm requires a prior over the space of morphologies. Usually, with no information to the contrary, morphologies with fewer states should have higher probabilities. We must also modulate the prior probability on the basis of what is inside each state, in that states with similar events inside them should be more probable than states with dissimilar events. Thus, if event *A* is similar to *B*, but not similar to *C*, then a state S_1 containing events *A* and *B*, should have a higher prior probability than another state S_2 containing events *A* and *C*. This way we bias the search toward morphologies that group similar events together into the same state. The simplest approximation to this is to disallow non-similar events in the same state; of course this is theoretically unsatisfying, but serves the purpose of illustration, and allows for an efficient search. For the example presented, the prior is a geometric distribution[2] with parameter $p=0.95$, multiplied by the product of the individual probabilities of each state in the morphology, where the probability of each state is 1 if all the events in the state are pairwise similar, and 0 otherwise:

$$P(M) = p(1-p)^{n-1}\prod_i K(S_i) \qquad \textbf{(1)}$$

$$K(S_i) = \begin{cases} 1 & if\ \forall e_j, e_k \in S_i,\ Sim(e_j, e_k) > 0 \\ 0 & otherwise \end{cases} \qquad \textbf{(2)}$$

In Equation 1, *M* is the model, *p* is the single parameter of the geometric distribution, *n* is the number of states in the morphology, and S_i is the i^{th} state. In Equation 2, e_j and e_k are events in state S_i, and *Sim* is the similarity function. For these examples, the similarity function is implemented by the Structure Mapping Engine, where two events are considered similar if they have a complete mapping (i.e., every item in the event is mapped to some item in the other event), and dissimilar otherwise.[3]

4. Search the Merge Space

The final step is to perform a search of the space of merges. In Bayesian model merging the posterior drives the search, in that we seek the morphology that maximizes the posterior. For the example presented, it was sufficient to perform a search that was exhaustive, aside from ignoring portions of the merge space that generate morphologies with zero probability.

Figure 2 shows the extraction of a simple morphology from two extremely short stories. The first story has to do with a boy and girl playing, a chasing event, a running away event, ending with a thinking event. The second story has to do with a man stalking a woman, followed by a scaring event, a fleeing event, and ending with a decision event. At some level of analysis these two stories are similar. The chasing and stalking events are similar in that they involve the one participant following after another, the running away and fleeing events are similar because they involve movement of one participant away from the other, and the thinking and decided events are both mental events involving an evaluation. If we represent these aspects of the semantics of these events, we can use an analogical mapping algorithm, such as the Structure Mapping Engine (Falkenhainer et al., 1989) to find the semantic and structural similarities. In the set of merges shown in the figure, first the chasing and stalking events are merged, then the running away and fleeing events, and then thinking and deciding. This results in a story morphology that generates stories with an optional 'playing' event at the beginning, a pursuit event, followed by an optional 'emotion' event, followed by the fleeing and evaluation events.

---

[2] Using a geometric function with the above mentioned parameter value to describe the dependence of the prior upon model size was arrived at by a cursory exploration of different possibilities and parameter values. To guard against charges that parameter tuning is responsible for the results, later work will need to explore the sensitivity of the algorithm to variations in choice of prior.

[3] Equation 1 does not define an actual probability distribution, but because we are interested in finding the morphology that maximizes the posterior, not the actual value of the posterior, this is of little concern.

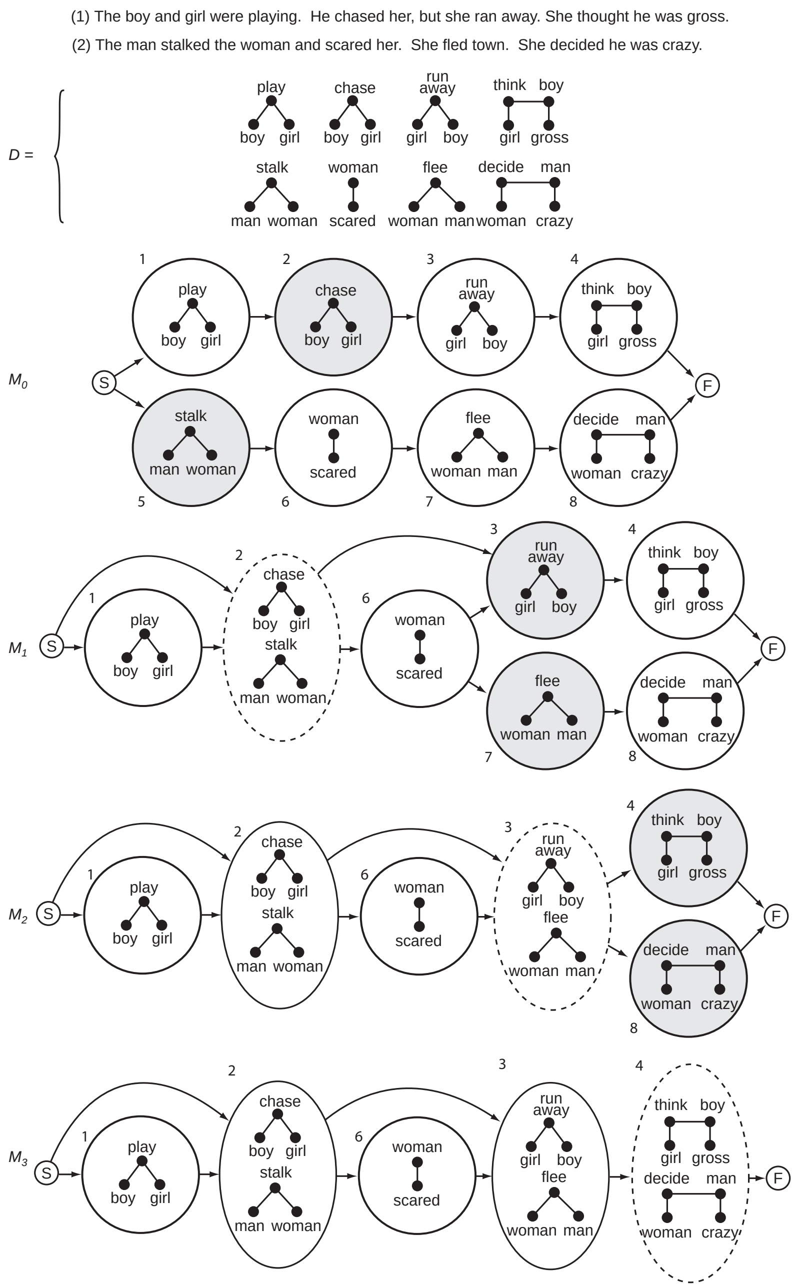

Figure 2: Example of Analogical Story Merging on two simple stories. Shown is a series of merges leading to the model that maximizes the posterior under the described parameters. The final model can generate not only the two input stories, but an additional two stories that alternatively include or exclude both nodes 1 and 6. Thus the model has generalized beyond the two input examples.

Example Morphology

I demonstrate the output of Analogical Story Merging on a small story corpus. The set comprises summaries of five plays by Shakespeare, namely, *Macbeth*, *Hamlet*, *Julius Caesar*, *Othello*, and *Taming of the Shrew*. The summaries were originally written in simple controlled English for another analogy system (Winston, 1980); the original controlled English is given in (Winston, 1981). Each summary contains between 7 and 11 events; each event is annotated in the figure with the first two letters of the story from which it came (e.g., the event represented by state 1, *sisters_predict_murder*, comes from ***MA****cbeth*).

The generated morphology is interesting for a number of reasons. First, the model captures important plot similarities and differences, indicating where one plot branches off from another, and where they merge together again. The paths through the model representing the four tragedies share states, while the path representing the single comedy, *Taming of the Shrew*, is in its own separate branch. *Hamlet* and *Macbeth* share a main branch {2,5,6}, with *Hamlet* detailing *persuasion* leading to the murder, and *Macbeth* adding a detour through the sequence {7,8,9,10} between states 2 and 5. They are more similar to each other than to *Julius Caesar*, which shares only the initial *persuade-murder* sequence. These similarities are consistent with analogical mapping considerations, in that a pairwise mapper comes to the same similarity conclusions (Winston, 1980). Also, events that are similarly key are grouped together. *Macbeth*, *Hamlet*, and *Julius Caesar* all have murders as motivating events; these are grouped into states 4 and 7. Similarly, both *Hamlet* and *Macbeth* conclude with a `revenge achieved' event, involving the killing of the murderer(s); these are grouped together into state 6.

Second, the morphology generalizes from the stories presented. This generalization is especially evident at states 9, 10, 20, 22, and 24, where multiple events occurring in sequence have been collapsed to a single state that allows an arbitrary number of events of those type to occur. State 10 contains the event where the Ghost orders Hamlet to kill Claudius, and might be thought of as a generalized ‘conspiracy’ event. States 12 and 13 contain killing or attacking events close in time, and each can be thought of as a ‘fight.’

Third, the morphology captures when something unusual happens, and sets it off by itself. This is most evident with the whole of the *Taming of the Shrew*, which is sufficiently dissimilar from any other play in the set that is relegated to its own branch. This also covers state 1, the event in *Macbeth* involving the three witches predicting things to come; despite strong starting similarities between *Macbeth*, *Hamlet*, and *Julius Caesar*, there is no analogous event, semantically, structurally, or temporally, in any other story in the set, and so the sisters are set off by themselves.

There are a number of peculiarities to note. First, it is curious that states 4 and 7 have not been merged. Their constituent events are certainly similar, and the events in question occur one after another. A similar sequence of four *attack* events are merged into the single state 21. So why not merge 4 and 7? The only story in which multiple murdering events occur in sequence is *Julius Caesar*[4], and thus merging those states together decreases $P(D|M)$ more than the increase in prior gained from reducing the model size by one state. Second, it is curious that states 5 and 13 were not merged, which would have agglomerated all the suicides that occur at the ends of the stories. However, merging the two would make it more likely that a Julius-Caesar-like story would have an additional killing at the end (rather than a suicide), and this movement of probability mass away from the original data was not offset by the reduction in model size. In general all questions of “to merge or not to merge” reduce to this tradeoff between compacting the model and losing fit with the original data.

[4] Caesar's murder is represented as two events, first a murder by Brutus, then a murder by Cassius.

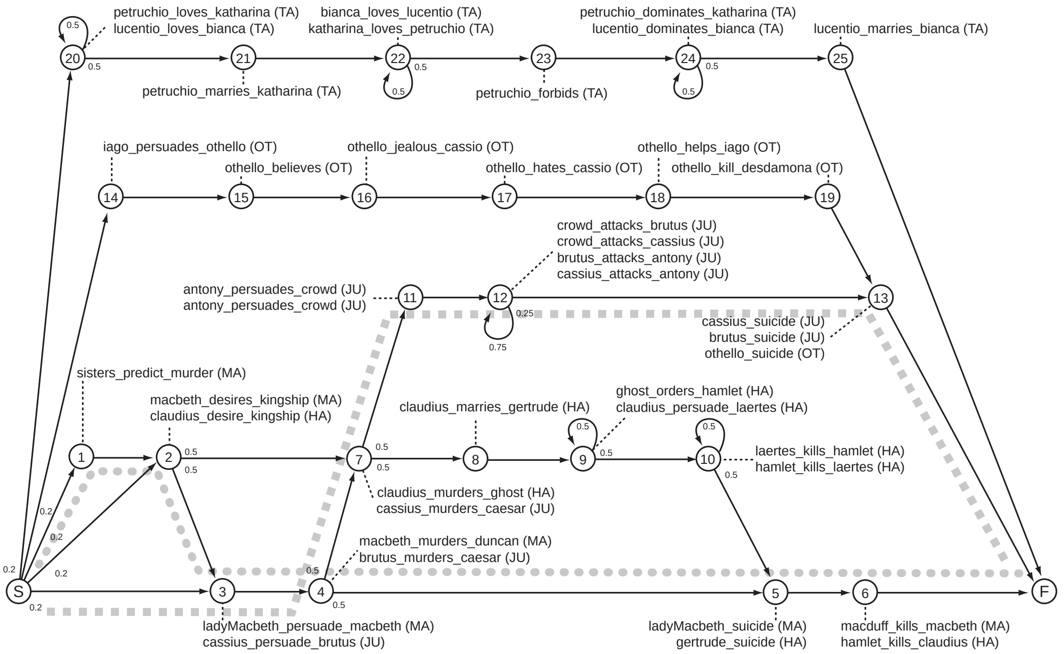

Figure 3: The morphology produced by running Analogical Story Merging over a set of five summaries of Shakespearean plays: *Macbeth*, *Hamlet*, *Julius Caesar*, *Othello*, and *Taming of the Shrew*. Each event is marked with the first two letters of the play from which it came. As an example, the shaded dashed lines indicate the path of *Macbeth* (circle dashes), *Julius Caesar* (square dashes), and their overlap, which occurs in states 3 and 4. *Macbeth* and *Hamlet* have the most plot overlap, as expected, followed by *Julius Caesar* against *Macbeth* and *Hamlet* (Winson, 1980). *Othello* shares the final suicide with *Julius Caesar*. The *Taming of the Shrew* is the only comedy, and is set off in its own completely separate track.

Analogical Story Merging Parameters

The Analogical Story Merging algorithm has been implemented as set of Java™ 6.0 libraries. The design choices and parameters follow closely on the steps outlined above. First, the implementation allows a user to vary how an initial model is constructed from the set of examples. Second, the merge operation can be modified. Third, it allows an arbitrary prior. Finally, there is no restriction on search technique.

Constructing the Initial Model

I chose to construct the initial model from the event sequences in the timeline of the stories, but there are numerous other options, depending on the type of morphology under consideration. One might use the events ordered as they are presented in the narrative (which can be different than the timeline). One might exclude some story events from being states (perhaps they are *prima facie* inconsequential), or include other non-events as states. One might make a single state out of one or more events. Or, one might choose a completely different level of semantics to define the states, such as causal or temporal relations.

State Merging Operation

The merge operation used here does not modify the internals of the states, rather treating states as bags of events. But a merge operation might invoke some sort of generalization, e.g., an analogical generalizer like SEQL (Kuehne, Forbus, Gentner, & Quinn, 2000), to create a single core symbol for the state from the symbols of the merged states. Such a merge operation would require the definition of a more complicated function for measuring the probability of the data given the model, since events observed in the data would not be emitted from the generalized state.

Forms of the Prior

The example presented was a modification of a geometric prior with parameter $p=0.95$. Nevertheless, if there is strong evidence as to the actual size of the morphology in question, this can be incorporated into the prior using, say, a beta distribution. Too, the portions of the prior which take account of the contents of the states can vary widely. In the example presented the modulation is all-or-none, requiring all events merged into a state to be pairwise similar. But the amount of probability assigned to a state could vary according to the magnitude of similarity of the events in the state, however that may be defined. In particular, it could be especially valuable to consider causal or temporal relations connecting events when calculating their similarity – many events that should be grouped into the same morphological state share causal connections with similar key events in the story.

Search Technique

Finally, the whole range of search techniques is available for searching the merge space. For even a small set of example stories the merge space can get quite large, so one will want to apply some intelligence to the design of the search algorithm.

Next Steps

Where to now that Analogical Story Merging is in hand? The first is, clearly, to vet the algorithm on larger sets of stories and to explore more fully the parameter space of the algorithm. Collecting large, semantically-annotated story corpora has been an impediment to research in analogy for quite some time. Fortunately, I have recently completed the beta version of a tool called the *Story Workbench* (Finlayson, 2008) that will help overcome this roadblock, and allow efficient and accurate assembly of large sets of semantically annotated stories. Russian folktales are the obvious choice for the first major story set to be collected, because the morphology so generated can be compared directly with Propp's.

Once corpora of appropriate size have been assembled, and reasonable morphologies extracted, the psychological reality of story morphologies can be investigated. Are people sensitive to the pieces of the morphologies identified by the algorithm? Do the morphologies really describe the constraints on generation of folktales for the culture in question? These issues can be explored through Bartlett-like story recall experiments (Bartlett, 1920, 1932), where story stimuli are modified using the morphology as a guide, with important pieces removed, or foreign pieces inserted.

Contributions

The contributions of the work described in this paper are as follows. First, I have suggested morphologies as a potentially powerful computational window into cultural thought. Second, I have specified a computationally tractable representation for morphologies, namely an HMM-like structure the states of which contain story events. Third, I have implemented an algorithm I call *Analogical Story Merging* that allows the derivation of a morphology from a set of stories. Finally, I have demonstrated the output of the algorithm when run over a story set containing five summaries of Shakespearean plays.

Acknowledgments

This work was supported in part by the Air Force Office of Scientific Research under grant number A9550-05-1-0321. Many thanks to Michael Klein, Whitman Richards, Mac Schwager, Peter Szolovits, and Patrick Winston for comments on earlier drafts, and to Josh Tenenbaum for advice on the work in general.

References

Bartlett, F.C. (1920). Some experiments on the reproduction of folk-stories. *Folklore, **31***, 30-47.

Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

Colby, B.N. (1973). A partial grammar of Eskimo folktales. *American Anthropologist, **75***, 645-662.

Dundes, A. (1964). *The morphology of North American Indian folktales*. Folklore Fellows Communications.

Falkenhainer, B., Forbus, K.D., Gentner, D. (1989). The Structure-Mapping Engine: Algorithm and examples. *Artificial Intelligence, **43***, 1-63.

Finlayson, M.A. (2008). Collecting semantics in the wild: The Story Workbench. In J. Beal, P. Bello, N. Cassimatis, M. Coen, & P.H. Winston (Eds.), *AAAI Fall Symposium on Naturally Inspired Artificial Intelligence* (p. 46-53). Washington DC: AAAI Press.

Kuehne, S.E., Forbus, K.D., Gentner, D., & Quinn, B. (2000). SEQL: Category learning as progressive abstraction using structure mapping. In *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society* (p. 770-775).

Propp, V. (1968). *Morphology of the folktale*. Austin: University of Texas Press.

Stolcke, A., & Omohundro, S. (1994). Inducing probabilistic grammars by Bayesian model merging. In *Grammatical Inference and Applications* (Lecture Notes in Computer Science Vol. 862, pp. 106-118). Berlin: Springer.

Winston, P.H. (1980). Learning and reasoning by analogy. *Communications of the ACM, **23***, 689-703.

Winston, P.H. (1981). *Learning new principles from precedents and exercises: The details.* MIT AI Laboratory Technical Report No. 632. Located on the web at *http://hdl.handle.net/1721.1/6356*