Feuding Families and Former Friends: Unsupervised Learning for Dynamic Fictional Relationships

Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi,
Jordan Boyd-Graber, and Hal Daumé III

University of Maryland, College Park
University of Colorado, Boulder

How can we describe a fictional relationship between two characters?

How can we describe a fictional relationship between two characters?

- isn't this easy? we can assign it a single label (or *relationship descriptor*) from a predetermined set

Friend or foe?

How can we describe a fictional relationship between two characters?

- isn't this easy? we can assign it a single label (or *relationship descriptor*) from a predetermined set

Friend or foe?

Peter Pan and Captain Hook (*Peter Pan*)

How can we describe a fictional relationship between two characters?

- isn't this easy? we can assign it a single label (or *relationship descriptor*) from a predetermined set

Friend or foe?

Peter Pan and Captain Hook (*Peter Pan)*

How can we describe a fictional relationship between two characters?

- isn't this easy? we can assign it a single label (or *relationship descriptor*) from a predetermined set

Friend or foe?

Peter Pan and Captain Hook (*Peter Pan*)

Frodo and Sam (*Lord of the Rings*)

How can we describe a fictional relationship between two characters?

- isn't this easy? we can assign it a single label (or *relationship descriptor*) from a predetermined set

Friend or foe?

Peter Pan and Captain Hook (*Peter Pan*)

Frodo and Sam (*Lord of the Rings*)

How can we describe a fictional relationship between two characters?

- isn't this easy? we can assign it a single label (or *relationship descriptor*) from a predetermined set

Friend or foe?

Peter Pan and Captain Hook (*Peter Pan*)

Frodo and Sam (*Lord of the Rings*)

Winston and Julia (*1984*)

How can we describe a fictional relationship between two characters?

- isn't this easy? we can assign it a single label (or *relationship descriptor*) from a predetermined set

Friend or foe?

Peter Pan and Captain Hook (*Peter Pan*)

Frodo and Sam (*Lord of the Rings*)

Winston and Julia (*1984*) **???**

How can we describe a fictional relationship between two characters?

- isn't this easy? we can assign it a single label (or *relationship descriptor*) from a predetermined set

Friend or foe?

Peter Pan and Captain Hook (*Peter Pan*)

Frodo and Sam (*Lord of the Rings*)

Winston and Julia (*1984*) **???**

Harry Potter and Sirius (*Prisoner of Azkaban*)

How can we describe a fictional relationship between two characters?

- isn't this easy? we can assign it a single label (or *relationship descriptor*) from a predetermined set

Friend or foe?

Peter Pan and Captain Hook (*Peter Pan*)

Frodo and Sam (*Lord of the Rings*)

Winston and Julia (*1984*) **???**

Harry Potter and Sirius (*Prisoner of Azkaban*) **???**

How can we describe a fictional relationship between two characters?

- what if we treat relationships as sequences (or *trajectories*) of descriptors? (Chaturvedi et al., 2016)

Tom Sawyer and Becky Thatcher:

friends -> foes -> friends

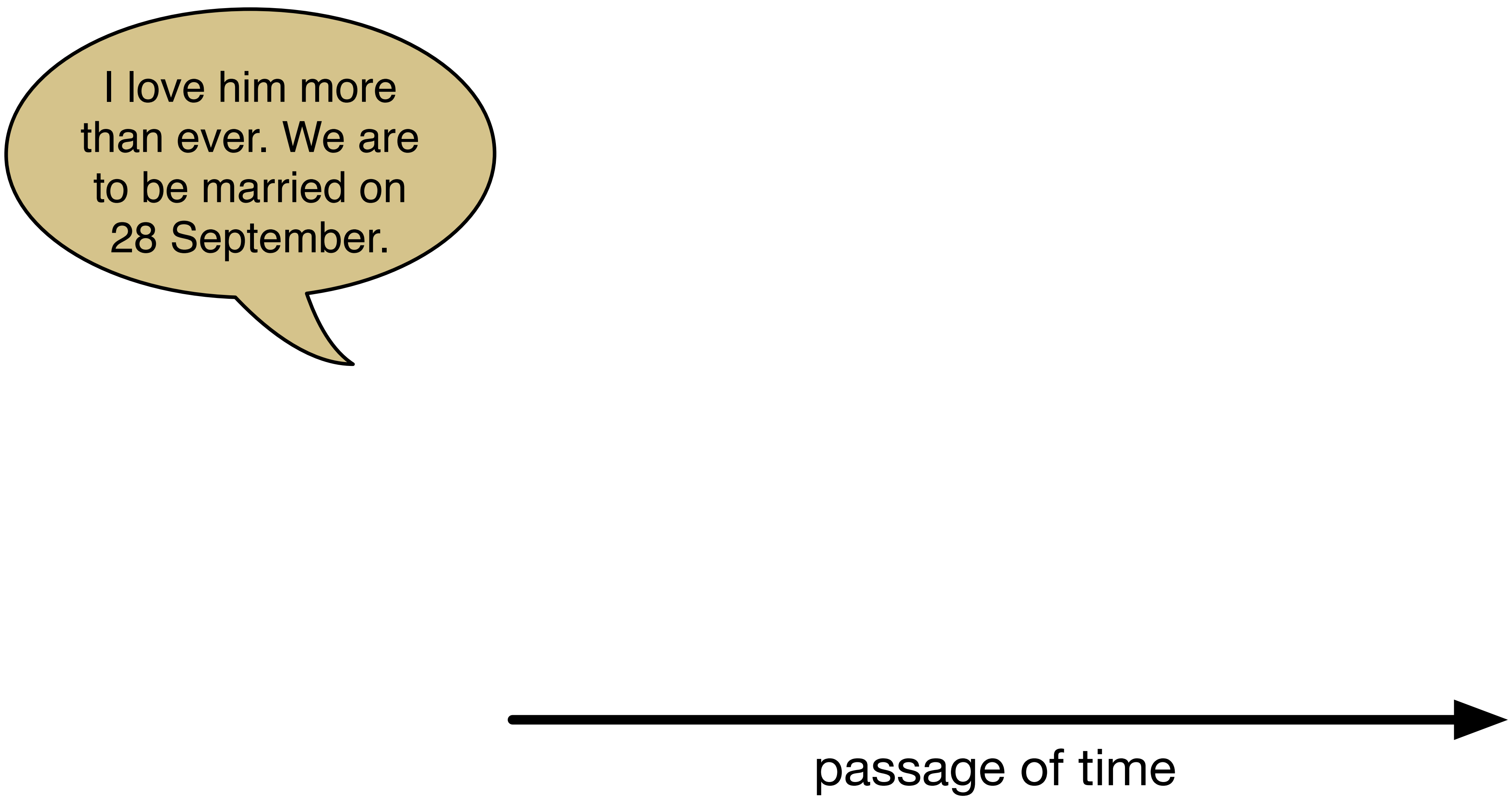
How can we describe a fictional relationship between two characters?

- what if we treat relationships as sequences (or *trajectories*) of descriptors? (Chaturvedi et al., 2016)

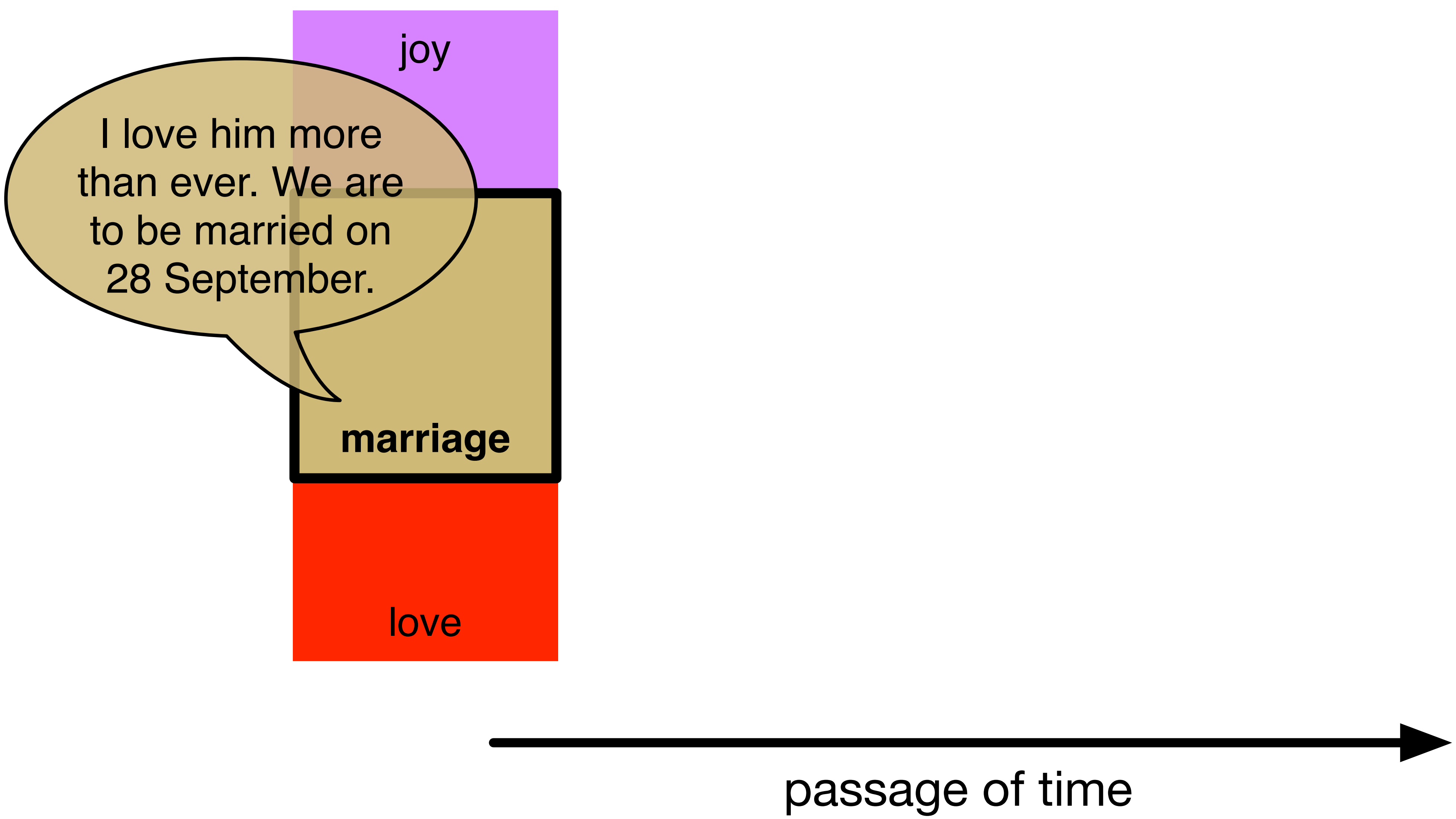
Tom Sawyer and Becky Thatcher:
friends -> foes -> friends

- limited by fixed descriptor set
- required expensive annotations
- limited to plot summaries

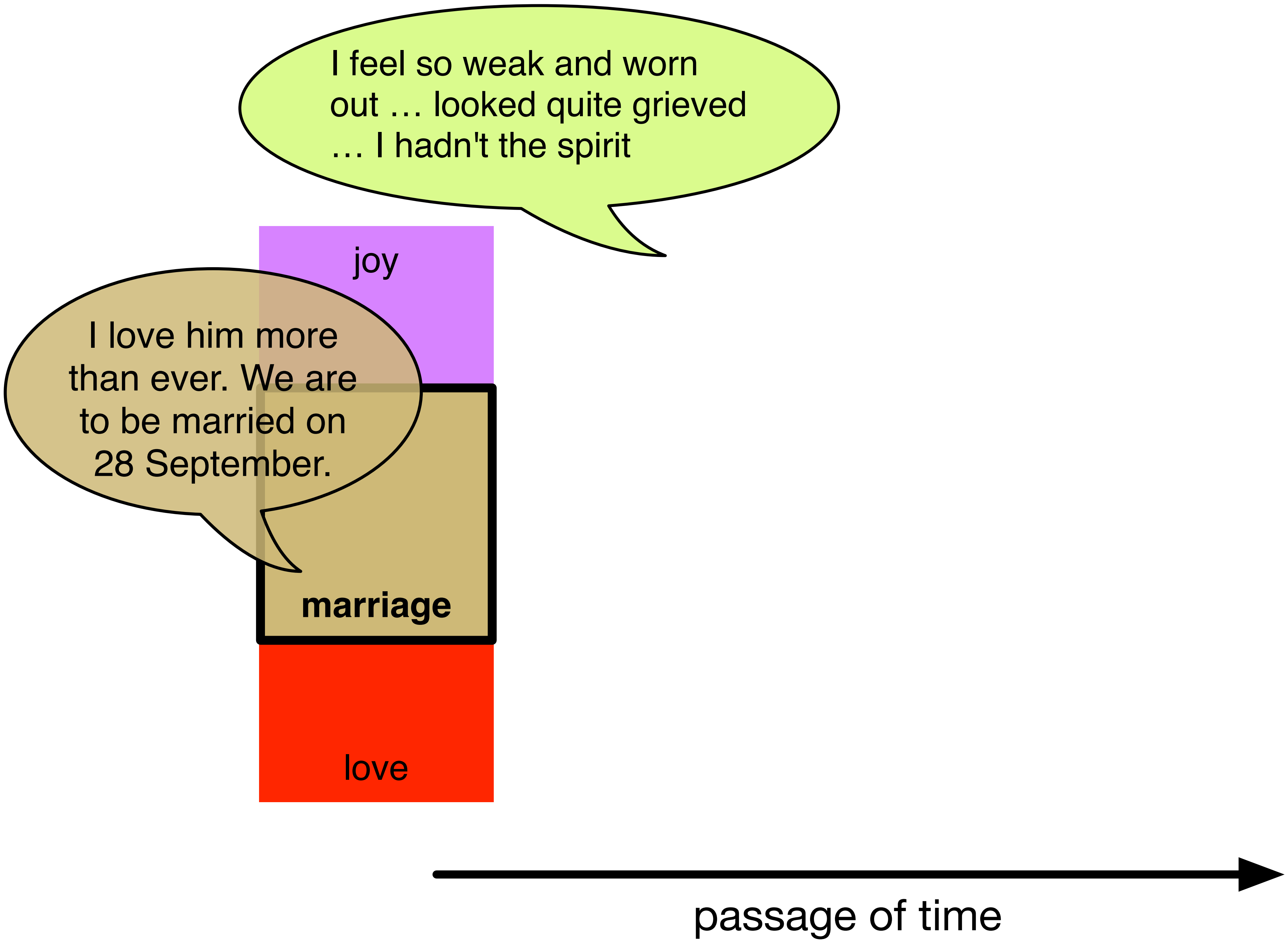
Arthur and Lucy (*Dracula*)



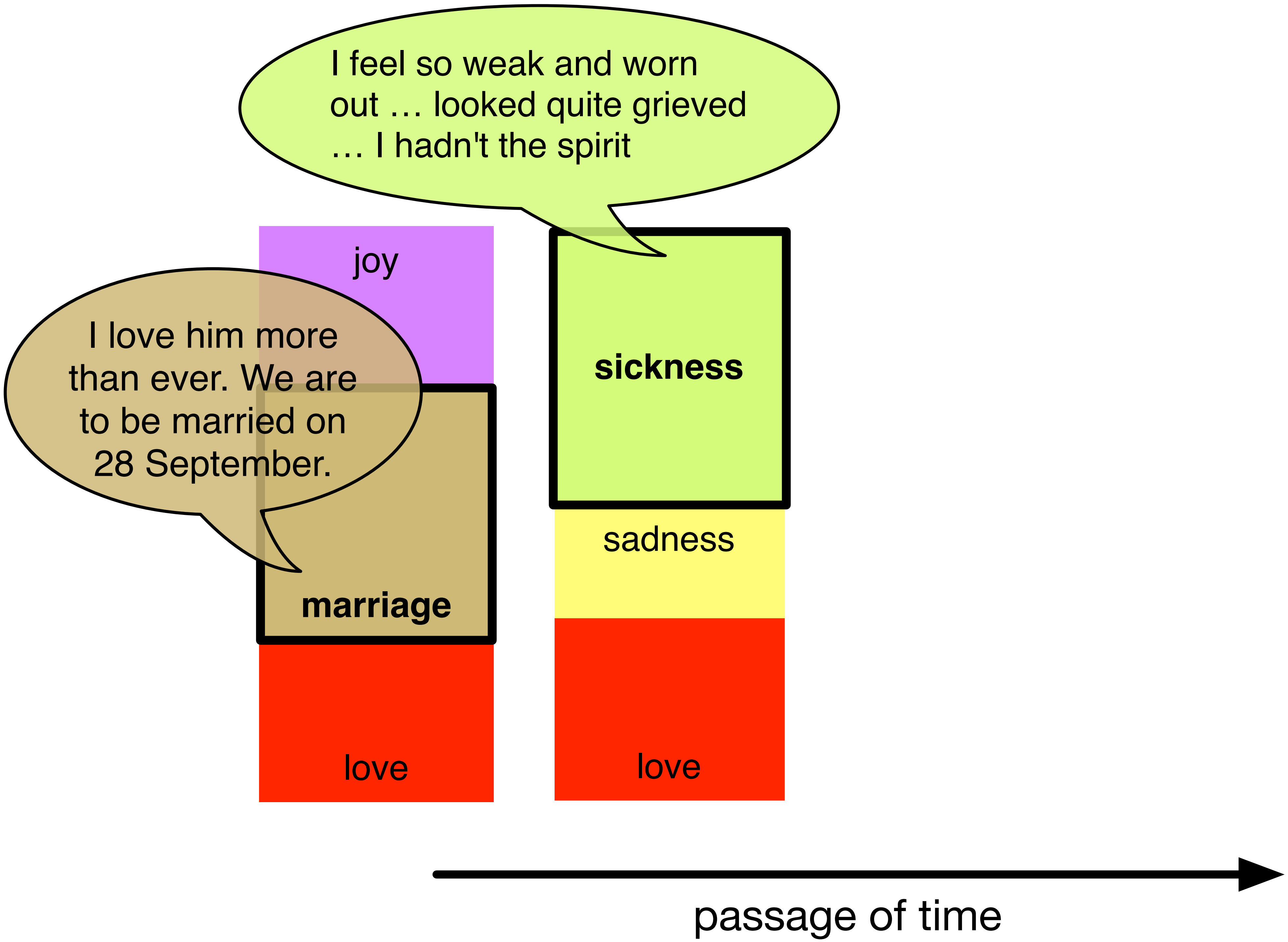
Arthur and Lucy (*Dracula*)



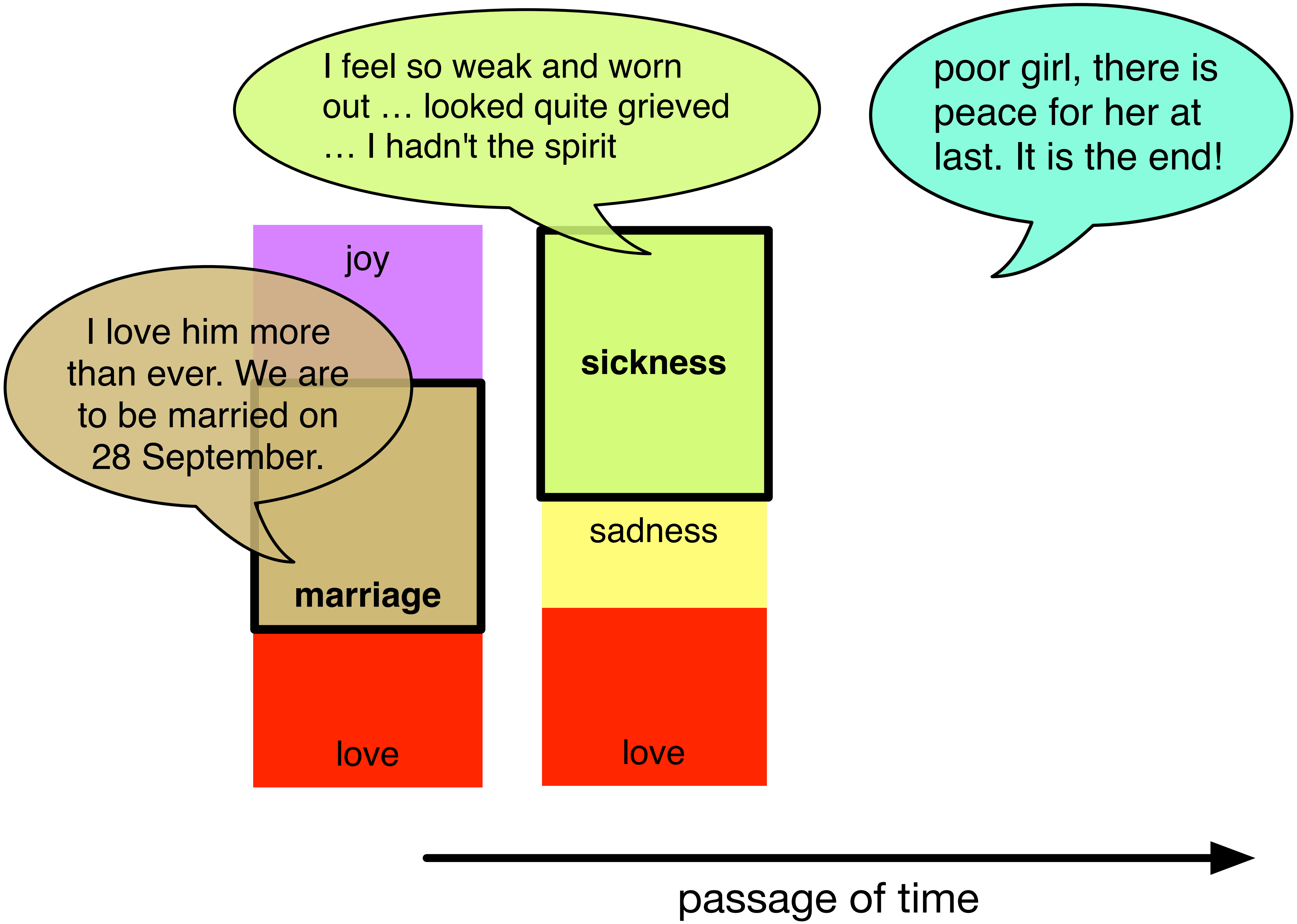
Arthur and Lucy (*Dracula*)



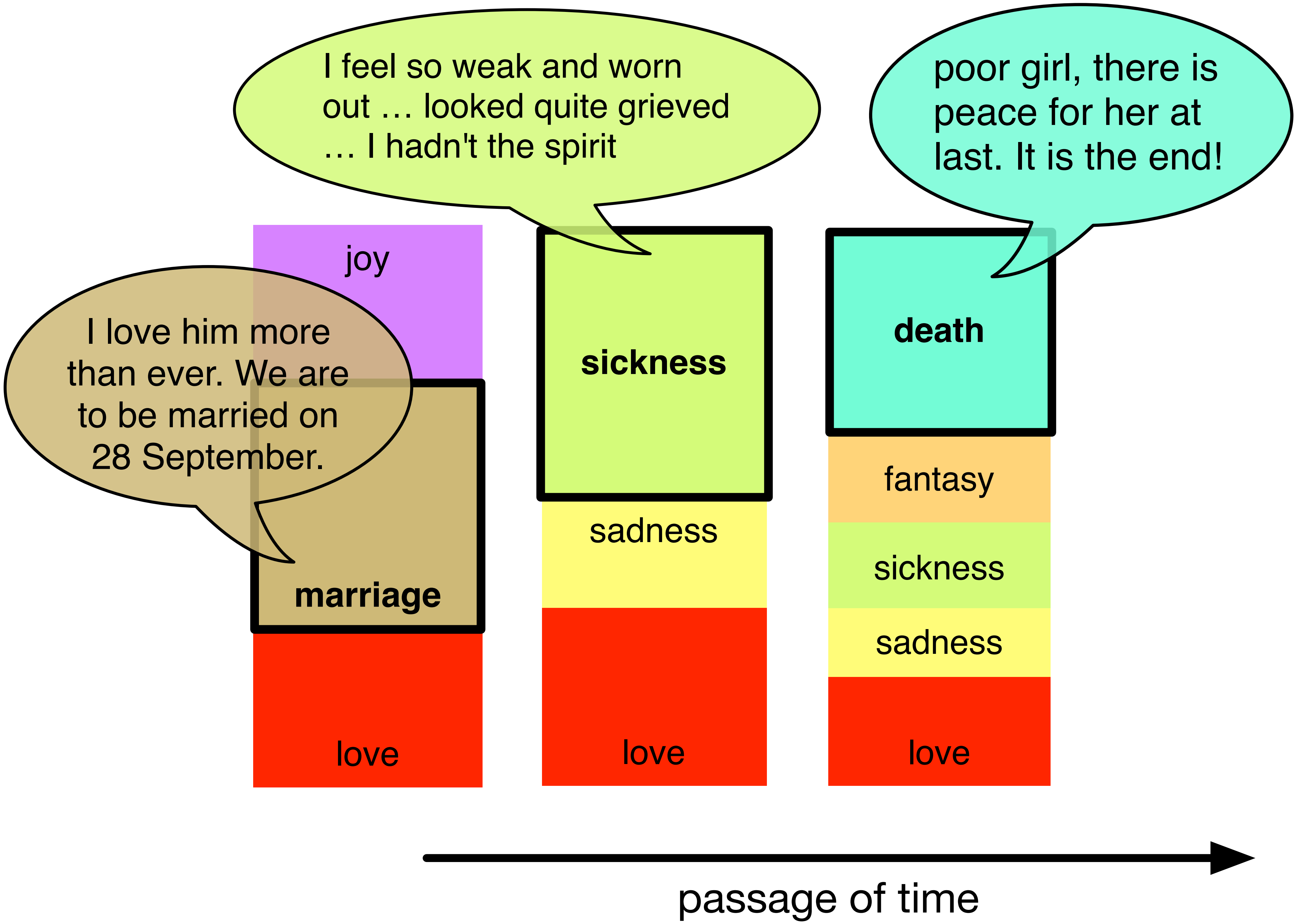
Arthur and Lucy (*Dracula*)



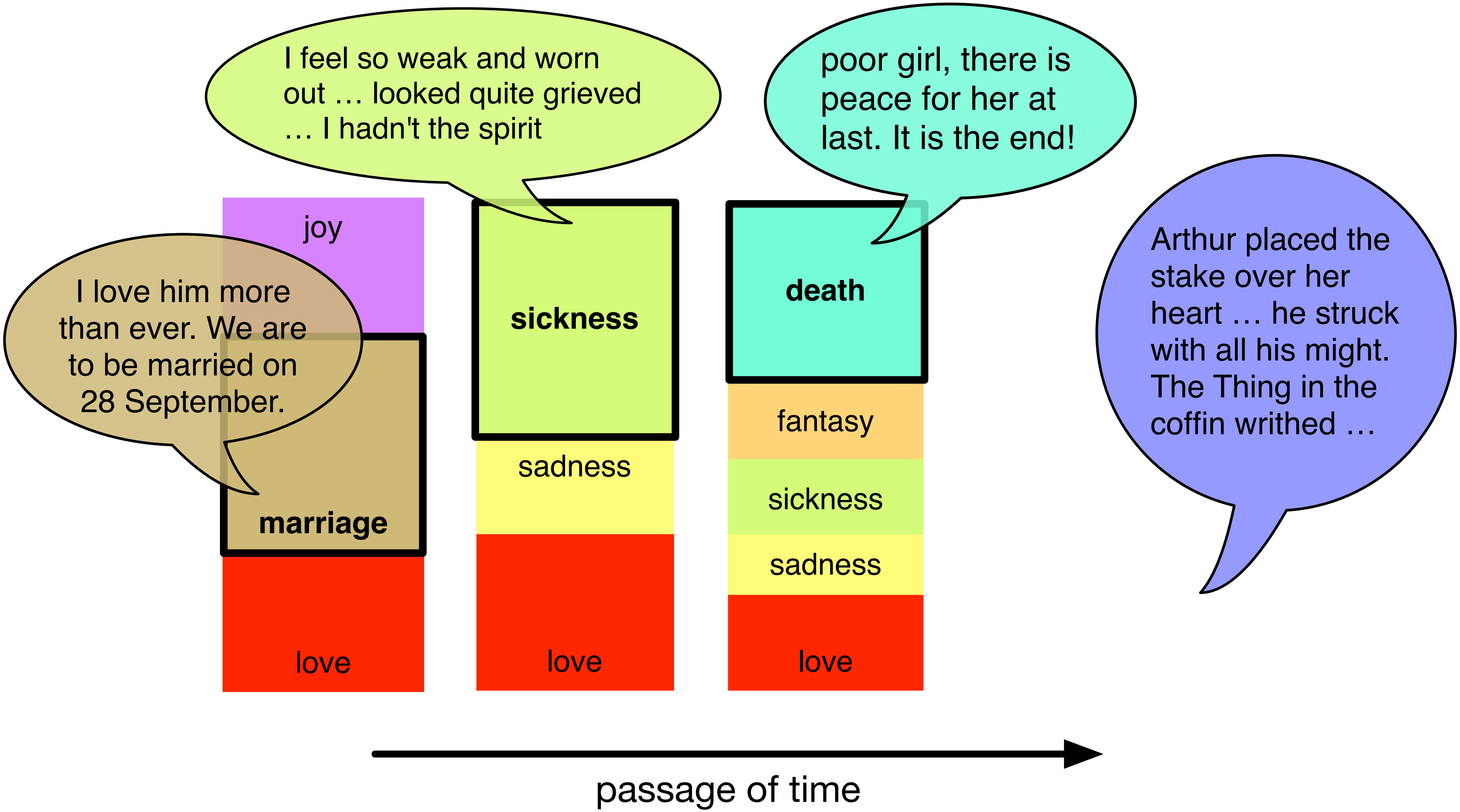
Arthur and Lucy (*Dracula*)



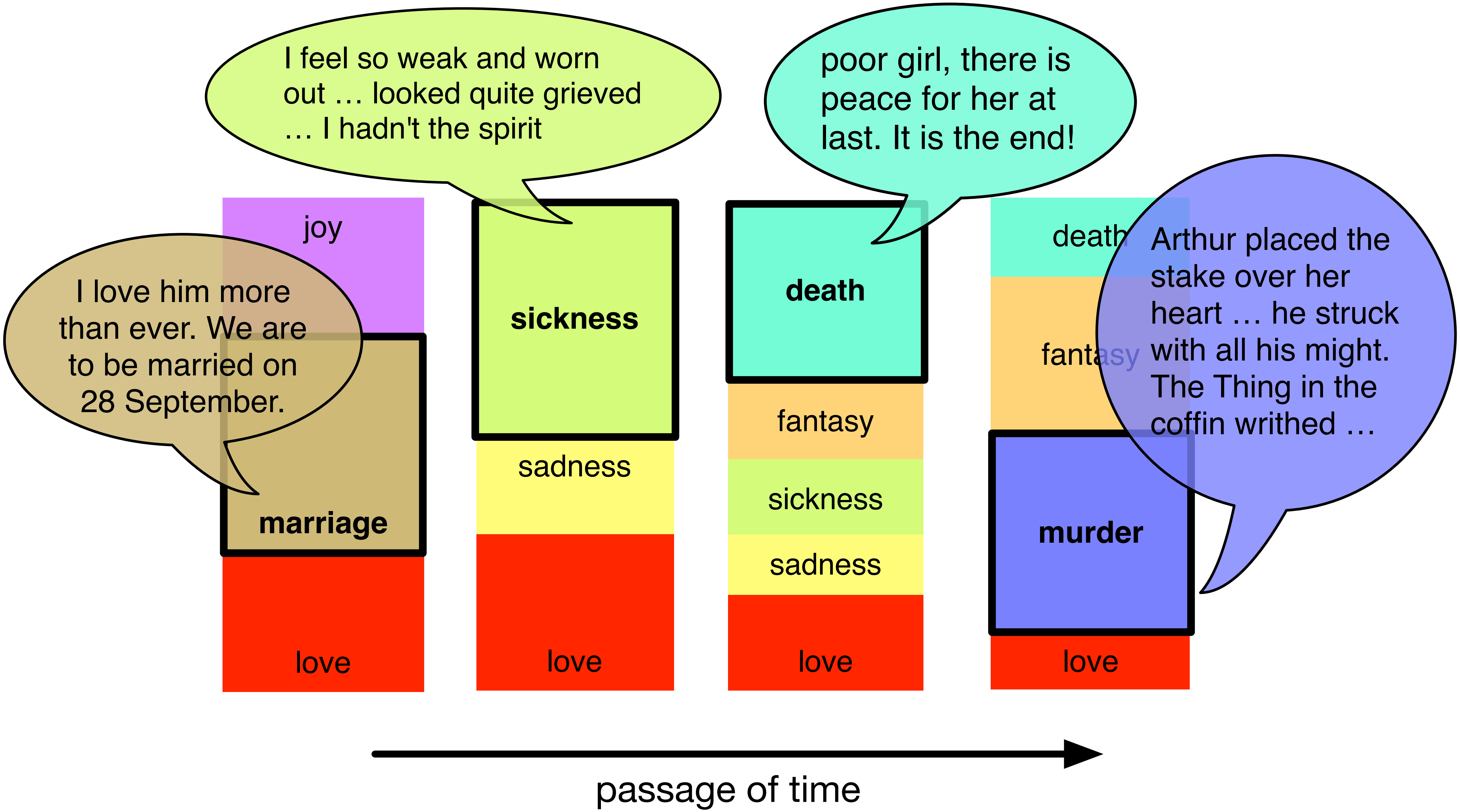
Arthur and Lucy (*Dracula*)



Arthur and Lucy (*Dracula*)



Arthur and Lucy (*Dracula*)



Arthur and Lucy (*Dracula*)

Why is this a worthwhile problem?

- "Distant reading" (Moretti, 2005) can help humanities scholars collect examples of specific relationship types

"Do Jane Austen's female and male protagonists have a pattern in their evolving relationship (e.g., mutual disdain followed by romantic love)?"
(Butler, 1975; Stovel, 1987; Hinant, 2006)

"Do certain authors or novels portray relationships of desire more than others?"
(Polhemus, 1990)

"Can we detect positive or negative subtext underlying meals between two characters?"
(Foster, 2009; Cognard-Black et al., 2014)

Outline

- Dataset: character interactions
- RMN: relationship modeling network
- Experiments: coherent descriptors, interpretable trajectories
- Analysis: RMN's strengths and weaknesses

A Dataset of Character Interactions

- For each pair of characters in a particular book, we extract all **spans** of text that contain mentions to both characters

A Dataset of Character Interactions

- For each pair of characters in a particular book, we extract all **spans** of text that contain mentions to both characters

t=0

“If anyone was ever minding his business, it was I," Ignatius breathed. "Please. We must stop. I think I'm going to have a hemorrhage.” "Okay." Mrs. Reilly looked at her son's reddening face and realized that he would very happily collapse at her feet just to prove his point.”

A Dataset of Character Interactions

- For each pair of characters in a particular book, we extract all **spans** of text that contain mentions to both characters

t=0

> “If anyone was ever minding his business, it was I," Ignatius breathed. "Please. We must stop. I think I'm going to have a hemorrhage.” "Okay." Mrs. Reilly looked at her son's reddening face and realized that he would very happily collapse at her feet just to prove his point.”

t=1

> “Ignatius belched the gas of a dozen brownies trapped by his valve. "Grant me a little peace….” "You know I appreciate you, babe," Mrs. Reilly sniffed. "Come on and gimme a little goodbye kiss like a good boy.”

A Dataset of Character Interactions

- For each pair of characters in a particular book, we extract all **spans** of text that contain mentions to both characters

t=0	“If anyone was ever minding his business, it was I," Ignatius breathed. "Please. We must stop. I think I'm going to have a hemorrhage.” "Okay." Mrs. Reilly looked at her son's reddening face and realized that he would very happily collapse at her feet just to prove his point.”
t=1	“Ignatius belched the gas of a dozen brownies trapped by his valve. "Grant me a little peace….” "You know I appreciate you, babe," Mrs. Reilly sniffed. "Come on and gimme a little goodbye kiss like a good boy.”
t=2	Mrs. Reilly looked at her son slyly and asked, "Ignatius, you sure you not a communiss?" "Oh, my God!" Ignatius bellowed. "Every day I am subjected to a McCarthyite witchhunt in this crumbling building. No!"

A Dataset of Character Interactions

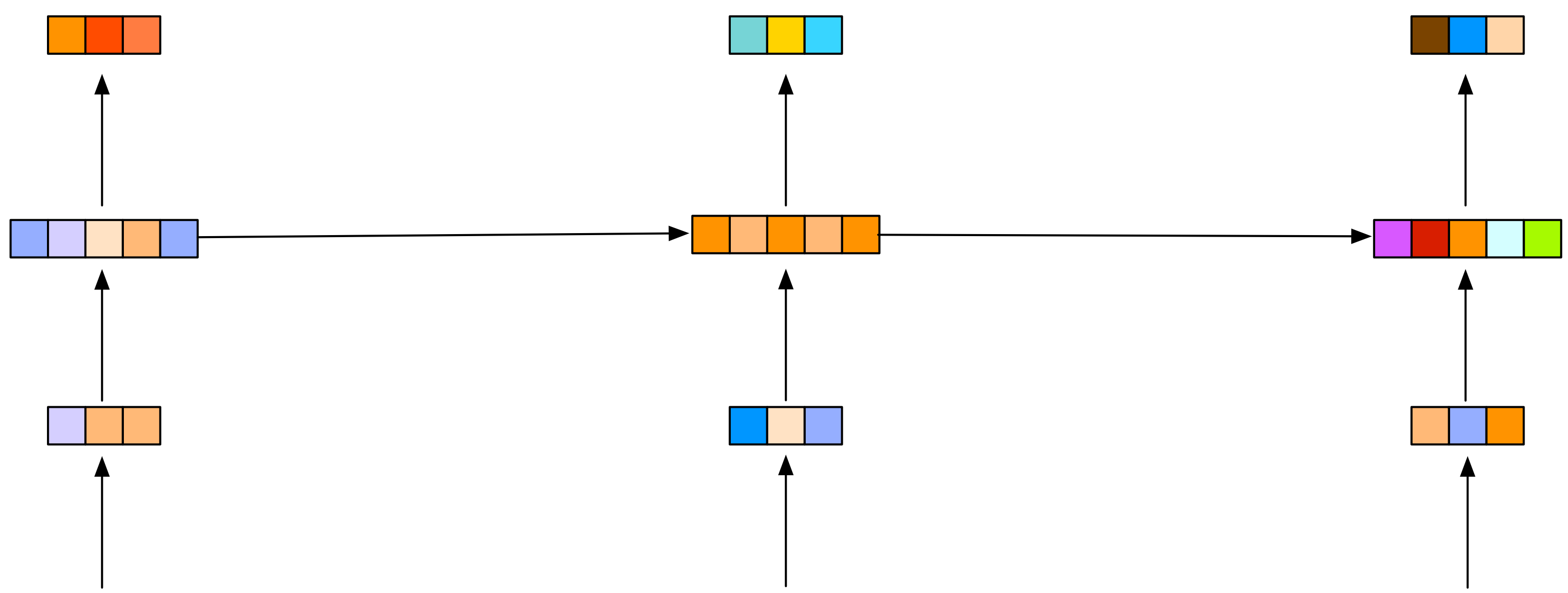
- 1,383 novels from Project Gutenberg and other Internet sources
  - Genres represented include romance, mystery, and fantasy
  - Preprocessed with David Bamman's BookNLP pipeline
  - Each span is a 200-token window centered around a character mention
- 20,013 unique character pairs and 380,408 spans

Relationship Modeling Network (RMN)

- recurrent autoencoder with dictionary learning

Relationship Modeling Network (RMN)

- recurrent autoencoder with dictionary learning



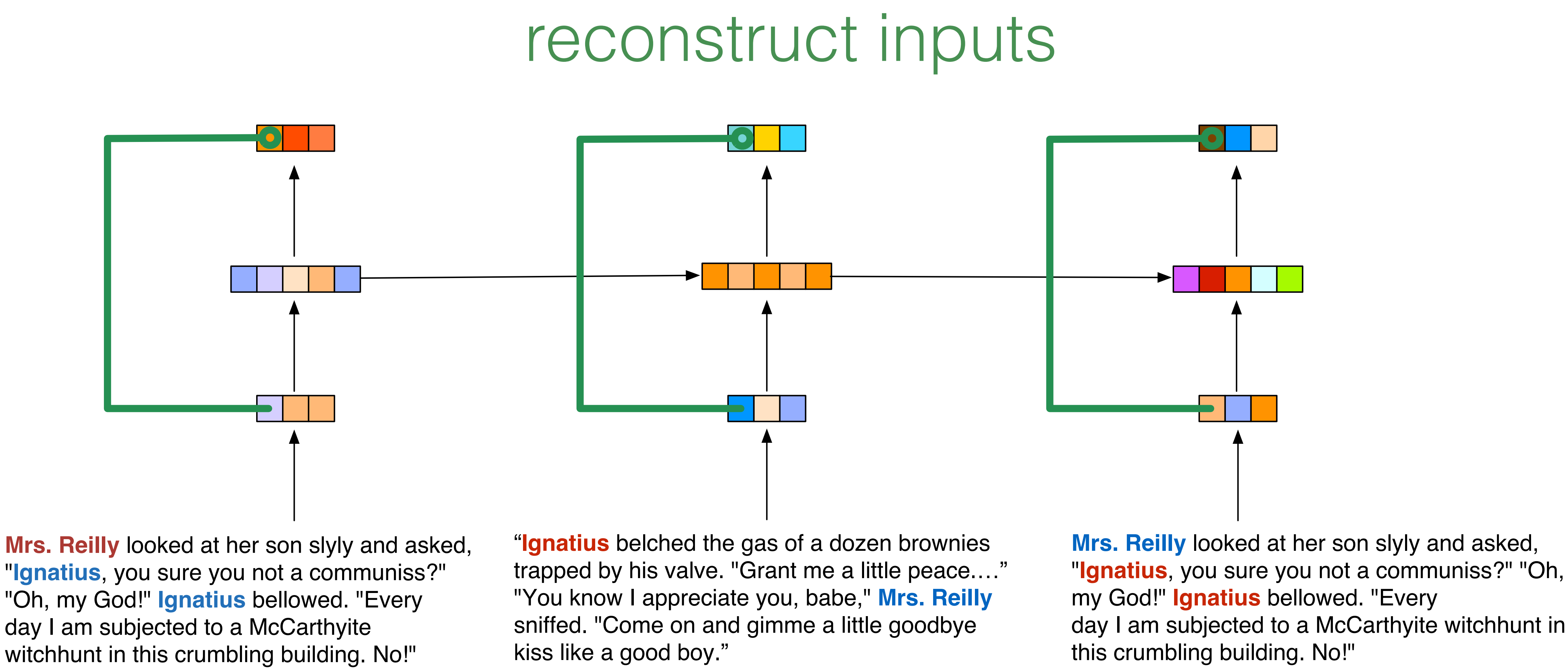
Mrs. Reilly looked at her son slyly and asked, "**Ignatius**, you sure you not a communiss?" "Oh, my God!" **Ignatius** bellowed. "Every day I am subjected to a McCarthyite witchhunt in this crumbling building. No!"

"**Ignatius** belched the gas of a dozen brownies trapped by his valve. "Grant me a little peace.…" "You know I appreciate you, babe," **Mrs. Reilly** sniffed. "Come on and gimme a little goodbye kiss like a good boy."

Mrs. Reilly looked at her son slyly and asked, "**Ignatius**, you sure you not a communiss?" "Oh, my God!" **Ignatius** bellowed. "Every day I am subjected to a McCarthyite witchhunt in this crumbling building. No!"

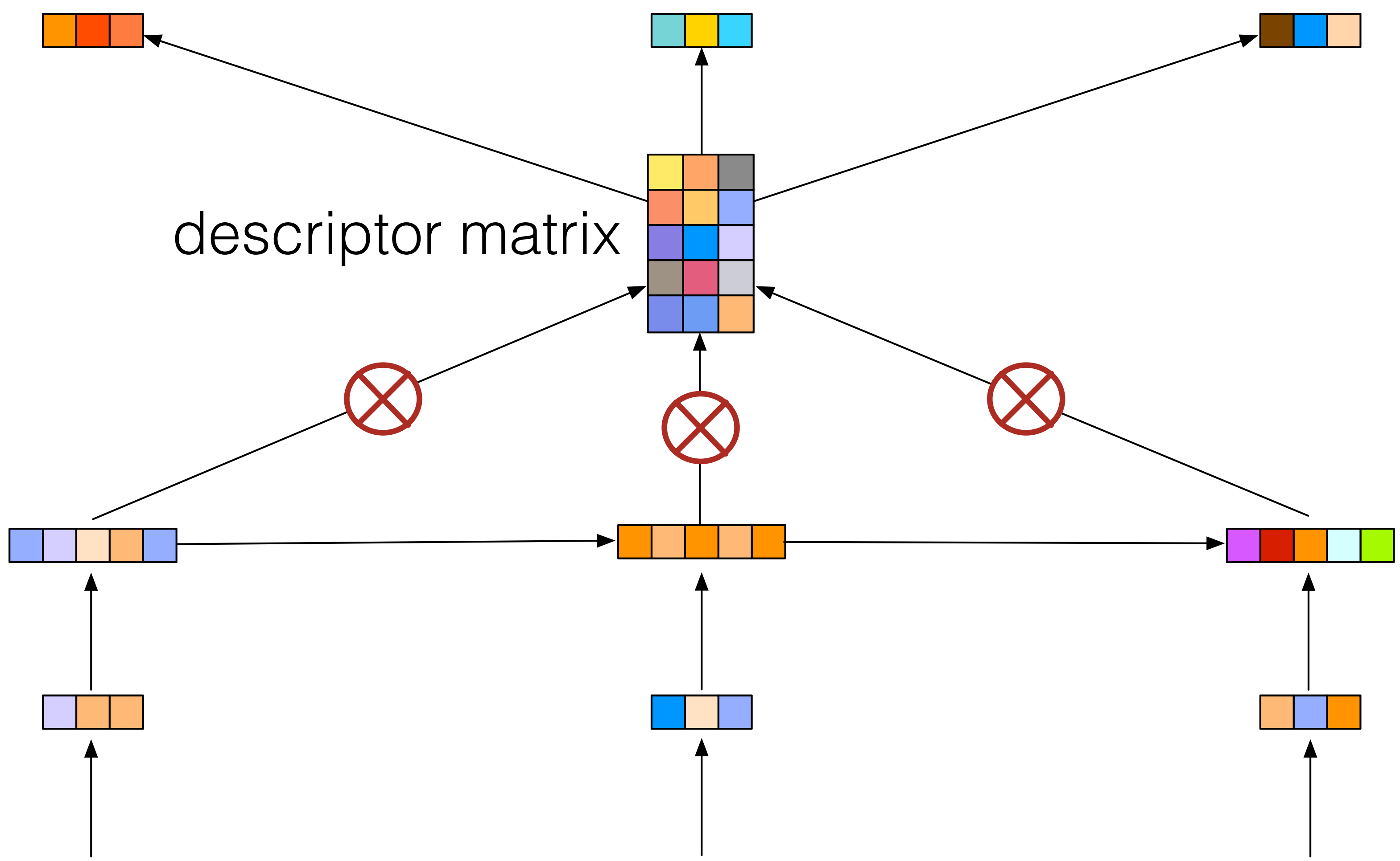
Relationship Modeling Network (RMN)

- recurrent autoencoder with dictionary learning



Relationship Modeling Network (RMN)

- recurrent autoencoder with dictionary learning



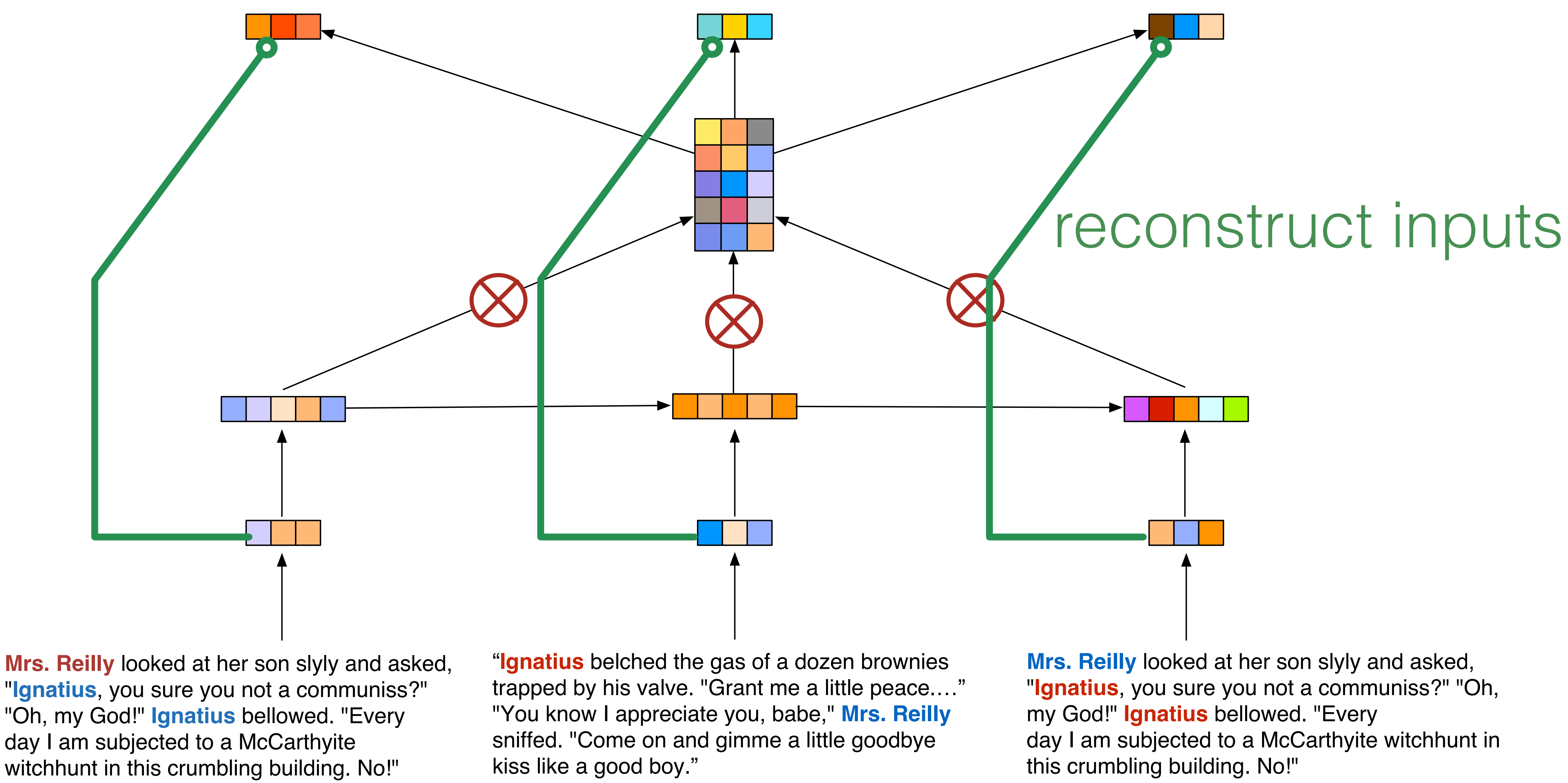
Mrs. Reilly looked at her son slyly and asked, "**Ignatius**, you sure you not a communiss?" "Oh, my God!" **Ignatius** bellowed. "Every day I am subjected to a McCarthyite witchhunt in this crumbling building. No!"

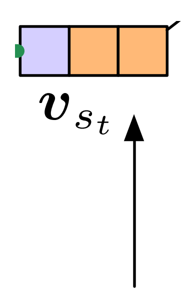
"**Ignatius** belched the gas of a dozen brownies trapped by his valve. "Grant me a little peace…." "You know I appreciate you, babe," **Mrs. Reilly** sniffed. "Come on and gimme a little goodbye kiss like a good boy."

Mrs. Reilly looked at her son slyly and asked, "**Ignatius**, you sure you not a communiss?" "Oh, my God!" **Ignatius** bellowed. "Every day I am subjected to a McCarthyite witchhunt in this crumbling building. No!"

Relationship Modeling Network (RMN)

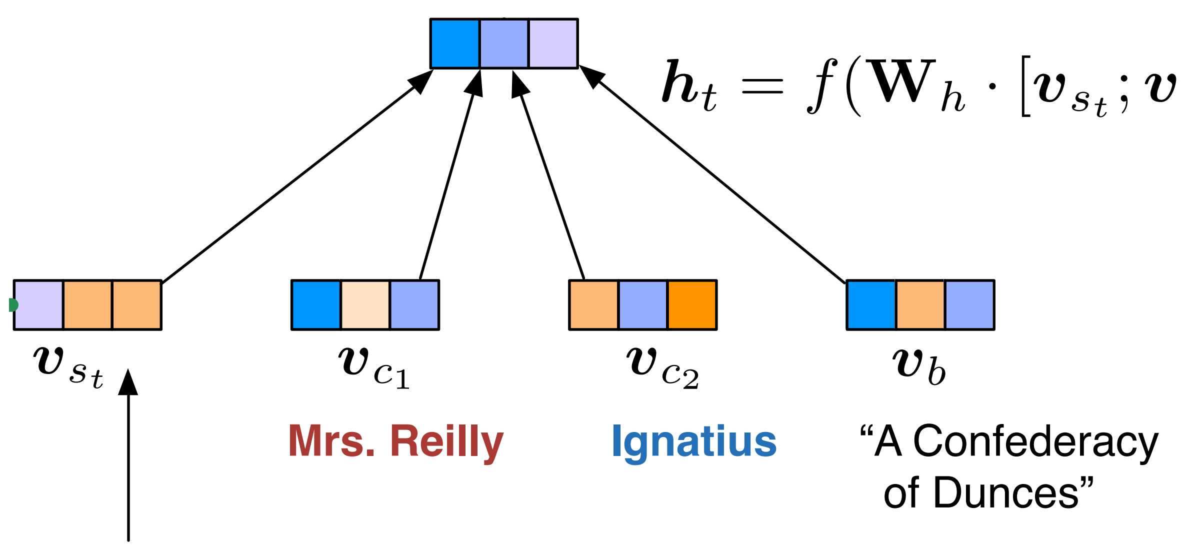
- recurrent autoencoder with dictionary learning





1. word embedding average

Mrs. Reilly looked at her son slyly and asked, "**Ignatius**, you sure you not a communiss?" "Oh, my God!" **Ignatius** bellowed. "Every day I am subjected to a McCarthyite witchhunt in this crumbling building. No!"

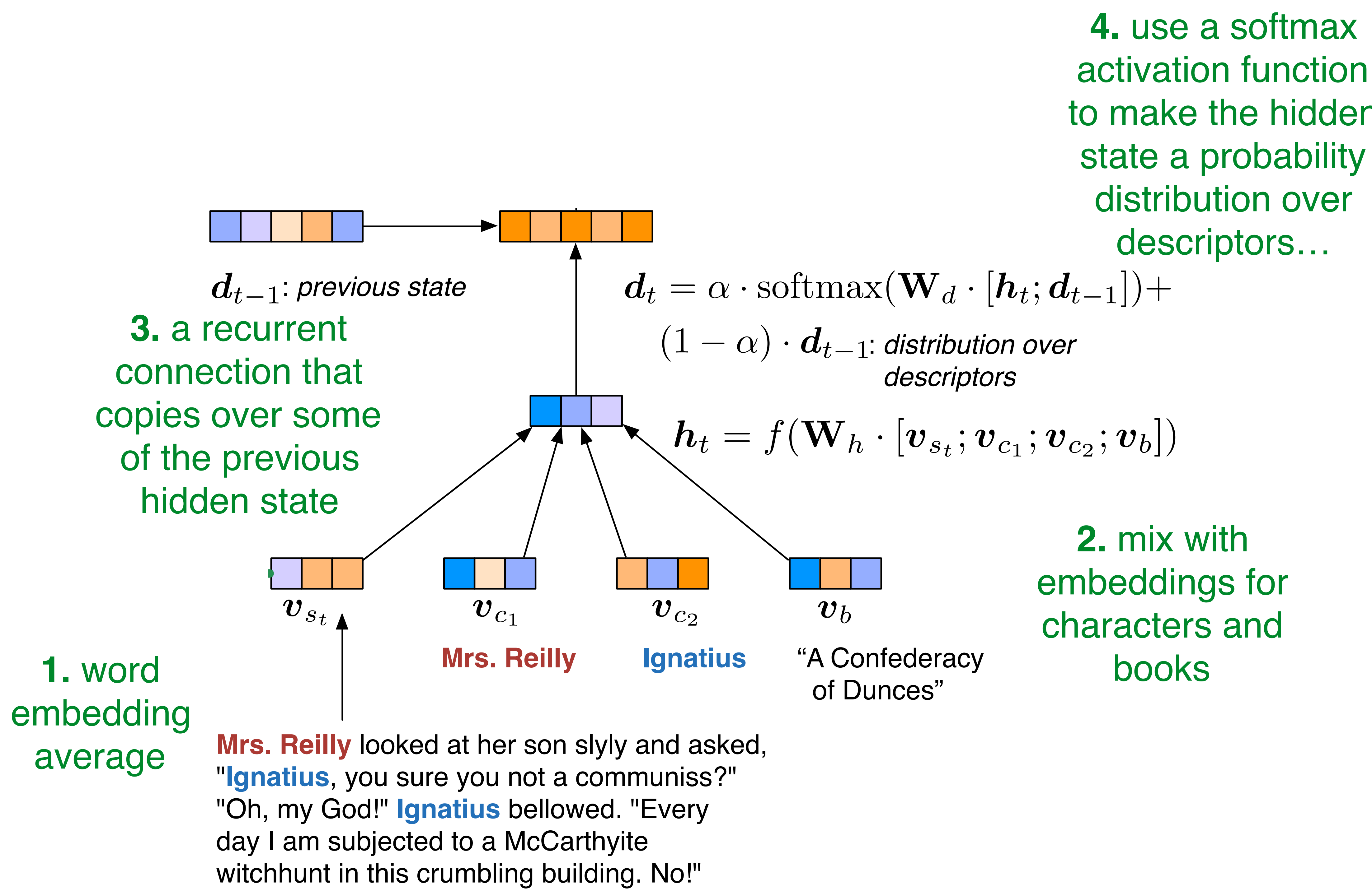


$$\boldsymbol{h}_t = f(\mathbf{W}_h \cdot [\boldsymbol{v}_{s_t}; \boldsymbol{v}_{c_1}; \boldsymbol{v}_{c_2}; \boldsymbol{v}_b])$$

1. word embedding average

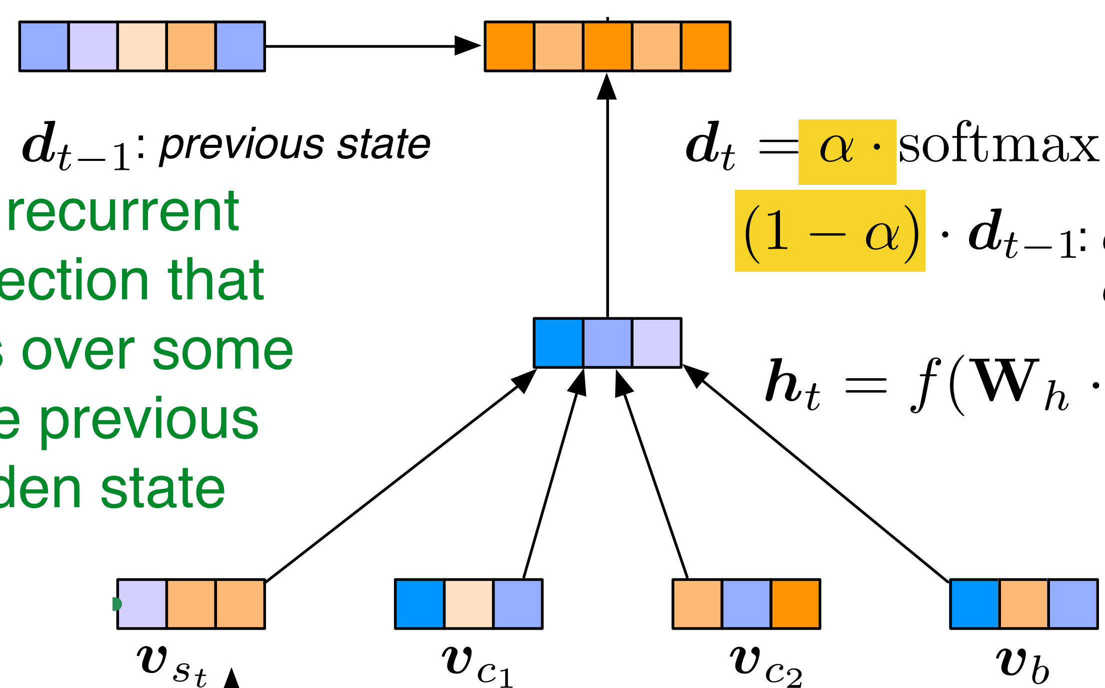
Mrs. Reilly looked at her son slyly and asked, "**Ignatius**, you sure you not a communiss?" "Oh, my God!" **Ignatius** bellowed. "Every day I am subjected to a McCarthyite witchhunt in this crumbling building. No!"

2. mix with embeddings for characters and books


4. use a softmax activation function to make the hidden state a probability distribution over descriptors…
$\boldsymbol{d}_{t-1}$: *previous state*
$\boldsymbol{d}_t = \alpha \cdot \mathrm{softmax}(\mathbf{W}_d \cdot [\boldsymbol{h}_t ; \boldsymbol{d}_{t-1}]) +$
$(1-\alpha) \cdot \boldsymbol{d}_{t-1}$: *distribution over descriptors*
3. a recurrent connection that copies over some of the previous hidden state
$\boldsymbol{h}_t = f(\mathbf{W}_h \cdot [\boldsymbol{v}_{s_t} ; \boldsymbol{v}_{c_1} ; \boldsymbol{v}_{c_2} ; \boldsymbol{v}_b])$
2. mix with embeddings for characters and books
$\boldsymbol{v}_{s_t}$
$\boldsymbol{v}_{c_1}$
$\boldsymbol{v}_{c_2}$
$\boldsymbol{v}_b$
Mrs. Reilly
Ignatius
“A Confederacy of Dunces”
1. word embedding average
Mrs. Reilly looked at her son slyly and asked, "Ignatius, you sure you not a communiss?" "Oh, my God!" Ignatius bellowed. "Every day I am subjected to a McCarthyite witchhunt in this crumbling building. No!"

4. **4.** use a softmax activation function to make the hidden state a probability distribution over descriptors…

linear interpolation



d_{t-1}: *previous state*

$$d_t = \alpha \cdot \text{softmax}(\mathbf{W}_d \cdot [h_t; d_{t-1}]) + (1-\alpha) \cdot d_{t-1}$$

: *distribution over descriptors*

3. a recurrent connection that copies over some of the previous hidden state

$$h_t = f(\mathbf{W}_h \cdot [v_{s_t}; v_{c_1}; v_{c_2}; v_b])$$

2. mix with embeddings for characters and books

1. word embedding average

Mrs. Reilly

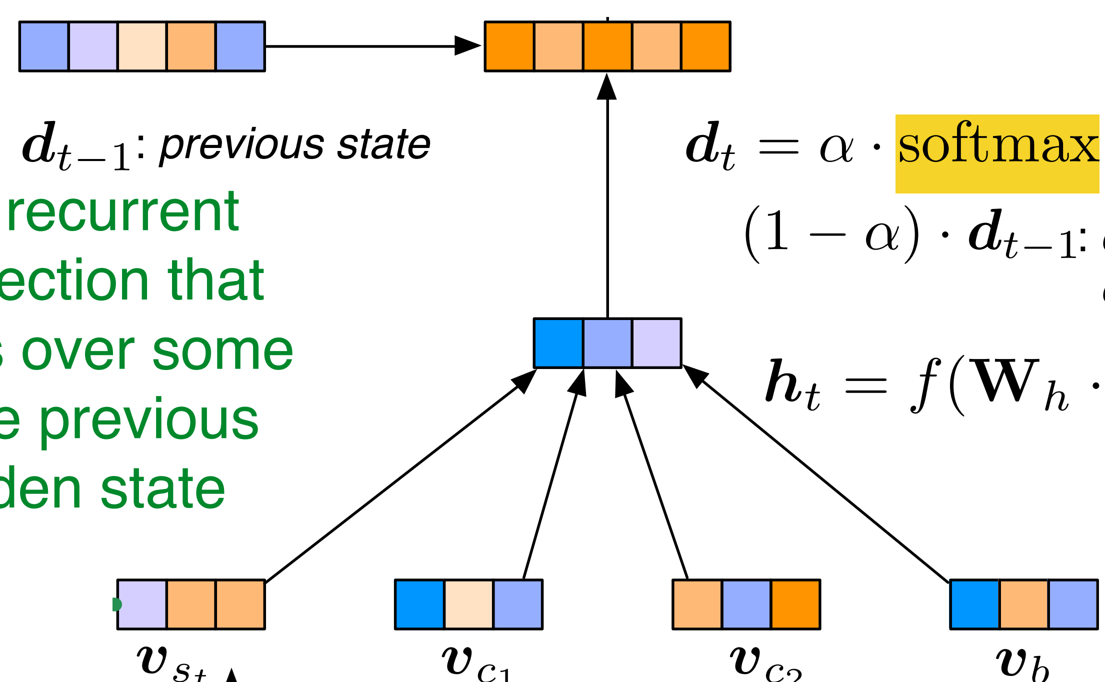
Ignatius

"A Confederacy of Dunces"

Mrs. Reilly looked at her son slyly and asked, "Ignatius, you sure you not a communiss?" "Oh, my God!" Ignatius bellowed. "Every day I am subjected to a McCarthyite witchhunt in this crumbling building. No!"

4. use a softmax activation function to make the hidden state a probability distribution over descriptors…

softmax activation



$$\boldsymbol{d}_t = \alpha \cdot \text{softmax}(\mathbf{W}_d \cdot [\boldsymbol{h}_t; \boldsymbol{d}_{t-1}]) + (1-\alpha) \cdot \boldsymbol{d}_{t-1}$$

distribution over descriptors

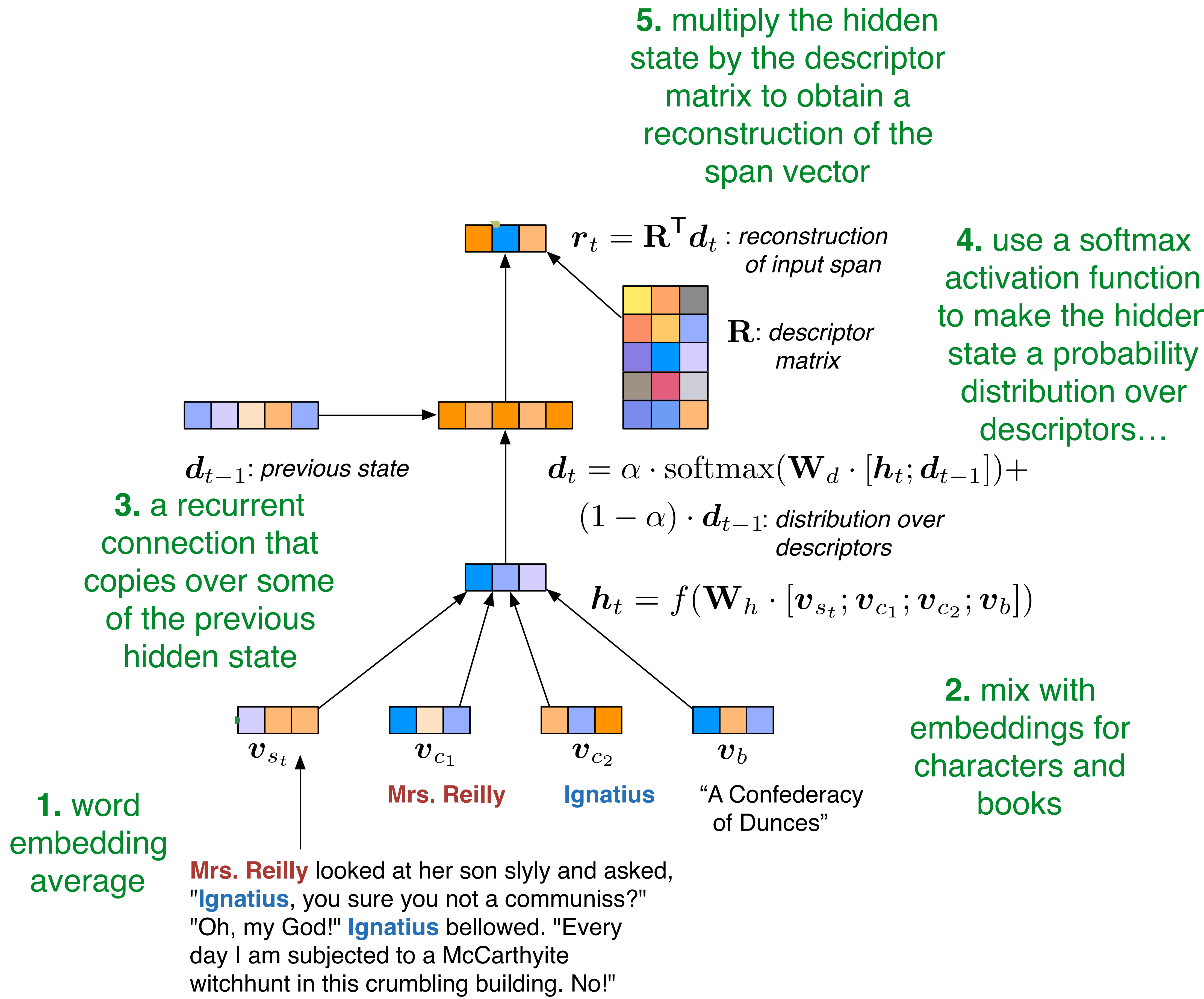
$$\boldsymbol{h}_t = f(\mathbf{W}_h \cdot [\boldsymbol{v}_{s_t}; \boldsymbol{v}_{c_1}; \boldsymbol{v}_{c_2}; \boldsymbol{v}_b])$$

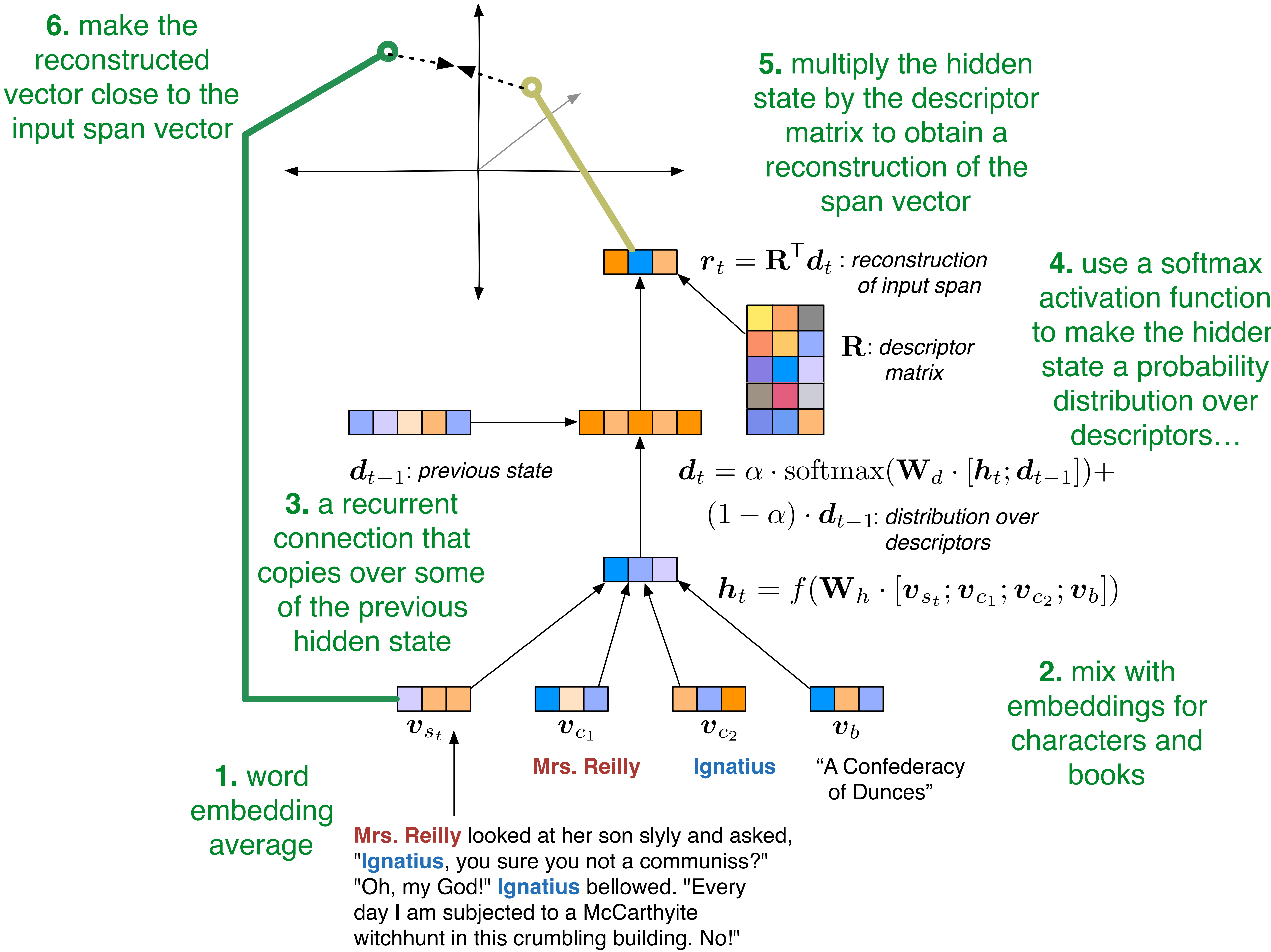
3. a recurrent connection that copies over some of the previous hidden state

2. mix with embeddings for characters and books

1. word embedding average

Mrs. Reilly looked at her son slyly and asked, "**Ignatius**, you sure you not a communiss?" "Oh, my God!" **Ignatius** bellowed. "Every day I am subjected to a McCarthyite witchhunt in this crumbling building. No!"


5. multiply the hidden state by the descriptor matrix to obtain a reconstruction of the span vector
$r_t = \mathbf{R}^{\top} d_t$: *reconstruction of input span*
$\mathbf{R}$: *descriptor matrix*
4. use a softmax activation function to make the hidden state a probability distribution over descriptors…
d_{t-1}: *previous state*
$d_t = \alpha \cdot \mathrm{softmax}(\mathbf{W}_d \cdot [h_t; d_{t-1}]) + (1-\alpha) \cdot d_{t-1}$: *distribution over descriptors*
3. a recurrent connection that copies over some of the previous hidden state
$h_t = f(\mathbf{W}_h \cdot [v_{s_t}; v_{c_1}; v_{c_2}; v_b])$
v_{s_t}
v_{c_1}
v_{c_2}
v_b
Mrs. Reilly
Ignatius
“A Confederacy of Dunces”
2. mix with embeddings for characters and books
1. word embedding average
Mrs. Reilly looked at her son slyly and asked, "Ignatius, you sure you not a communiss?" "Oh, my God!" Ignatius bellowed. "Every day I am subjected to a McCarthyite witchhunt in this crumbling building. No!"


6. make the reconstructed vector close to the input span vector
5. multiply the hidden state by the descriptor matrix to obtain a reconstruction of the span vector
$\boldsymbol{r}_t = \mathbf{R}^{\mathsf{T}} \boldsymbol{d}_t$: *reconstruction of input span*
4. use a softmax activation function to make the hidden state a probability distribution over descriptors…
$\mathbf{R}$: *descriptor matrix*
$\boldsymbol{d}_{t-1}$: *previous state*
$\boldsymbol{d}_t = \alpha \cdot \mathrm{softmax}(\mathbf{W}_d \cdot [\boldsymbol{h}_t; \boldsymbol{d}_{t-1}]) + (1-\alpha) \cdot \boldsymbol{d}_{t-1}$: *distribution over descriptors*
3. a recurrent connection that copies over some of the previous hidden state
$\boldsymbol{h}_t = f(\mathbf{W}_h \cdot [\boldsymbol{v}_{s_t}; \boldsymbol{v}_{c_1}; \boldsymbol{v}_{c_2}; \boldsymbol{v}_b])$
2. mix with embeddings for characters and books
$\boldsymbol{v}_{s_t}$
$\boldsymbol{v}_{c_1}$
$\boldsymbol{v}_{c_2}$
$\boldsymbol{v}_b$
Mrs. Reilly
Ignatius
"A Confederacy of Dunces"
1. word embedding average
Mrs. Reilly looked at her son slyly and asked, "Ignatius, you sure you not a communiss?" "Oh, my God!" Ignatius bellowed. "Every day I am subjected to a McCarthyite witchhunt in this crumbling building. No!"

Labeling the Learned Descriptors

- We compute the nearest word embeddings to each row of the descriptor matrix **R**, which humans use to provide external labels.

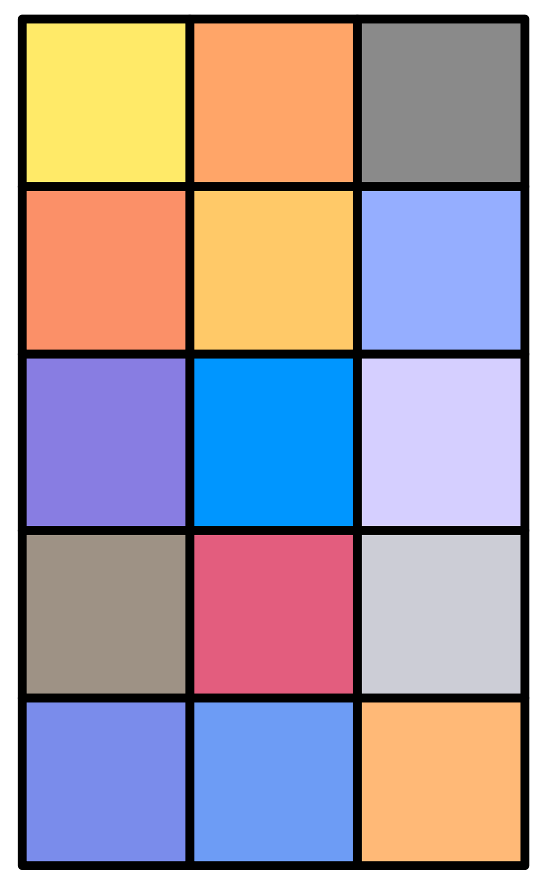
violence: *grenades, guns, bullets*

sadness: *regretful, rueful, pity*

politics: *political, leadership, rule*

fantasy: *cosmic, realms, universe*

suffering: *fear, nightmares, suffer*



Relationship to Topic Models

- RMN outputs ≈ topic model latent variables:
  - descriptor matrix R ≈ topic-word matrices ϕ
  - descriptor weights d_t at each timestep ≈ document-topic assignments z
- Baselines:
  - temporally-oblivious: LDA (Blei et al., 2001), Nubbi (Chang et al., 2008)
  - temporally-aware: HTMM (Gruber et al., 2007)

Experiment 1:

Descriptor Coherence

Do the Descriptors Make Sense?

- Goal: compare the descriptors learned by the RMN to the topics learned by our topic model baselines

- Task: word intrusion (Chang et al., 2009)

  - Workers identify an "intruder" word from a set of words that —other than the intruder—come from the same descriptor

contempt malice condescend praise distaste mock

worship pray devote yourselves gods gather

Do the Descriptors Make Sense?

- Goal: compare the descriptors learned by the RMN to the topics learned by our topic model baselines

- Task: word intrusion (Chang et al., 2009)

    - Workers identify an “intruder” word from a set of words that —other than the intruder—come from the same descriptor

contempt malice condescend ***praise*** *distaste mock*

worship pray devote yourselves gods gather

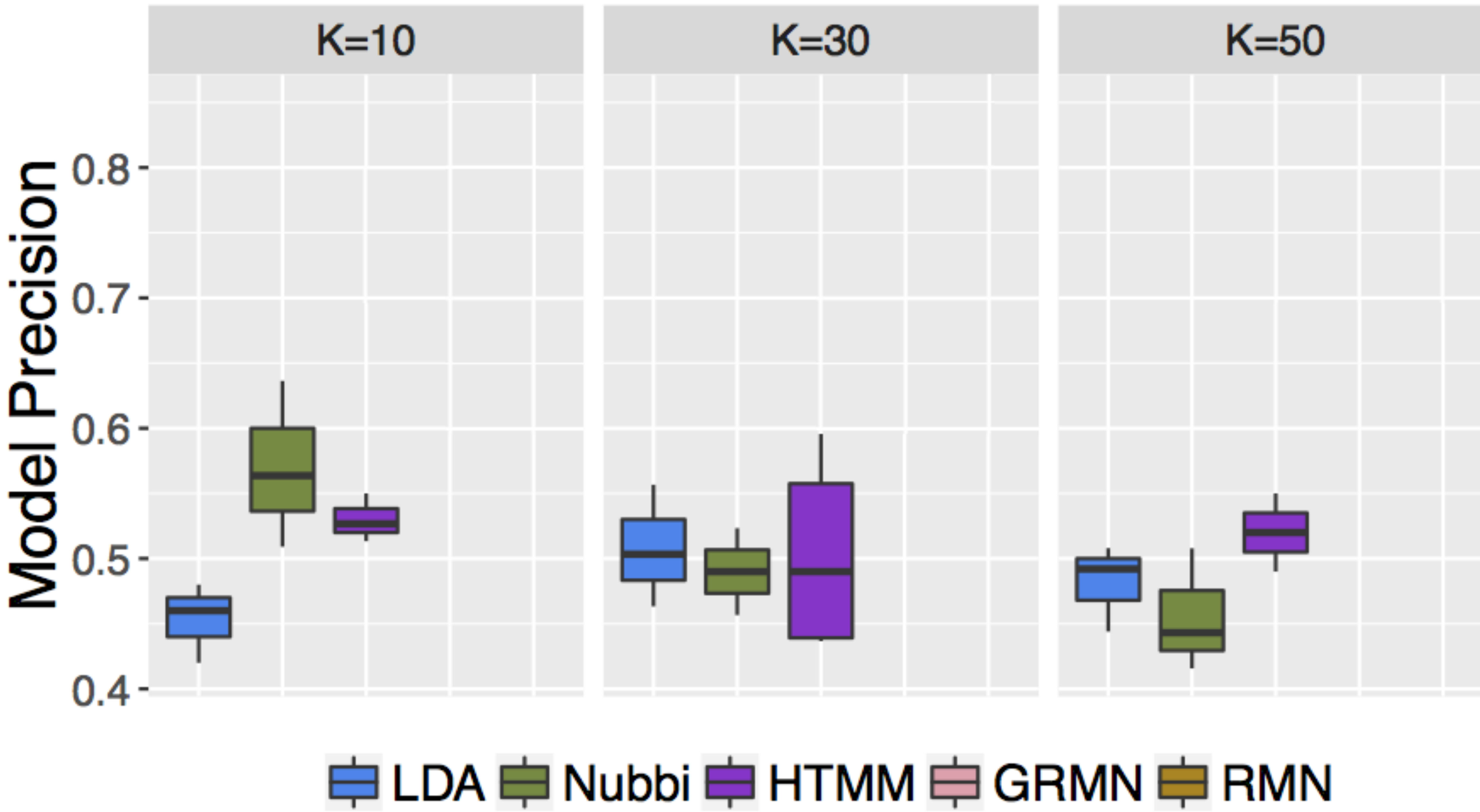
Do the Descriptors Make Sense?

- Goal: compare the descriptors learned by the RMN to the topics learned by our topic model baselines
- Task: word intrusion (Chang et al., 2009)
    - Workers identify an "intruder" word from a set of words that —other than the intruder—come from the same descriptor

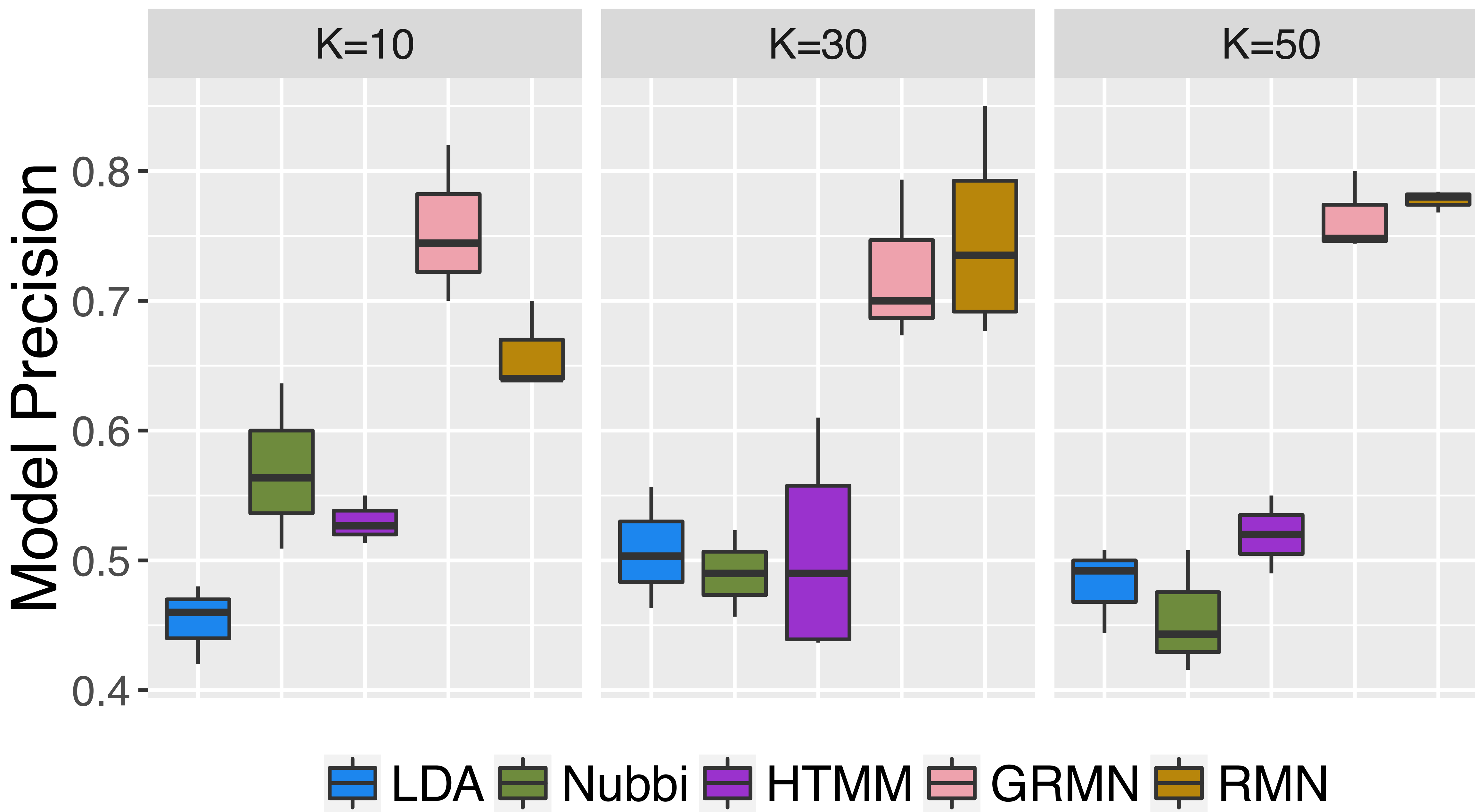
contempt malice condescend ***praise*** *distaste mock*

worship pray devote ***yourselves*** *gods* ***gather***

Do the Descriptors Make Sense?



Do the Descriptors Make Sense?



Coherent Descriptors

RMN

outdoors: *outdoors trail trails hillside grassy slopes*
sadness: *regretful rueful pity pained despondent*
education: *teaching graduate year teacher attended*
love: *love delightful happiness enjoyed enjoyable*
murder: *autopsy arrested homicide murdered*

HTMM

crime: *blood knife pain legs steal*
food: *kitchen mouth glass food bread*
violence: *sword shot blood shouted swung*
boats: *ship boat captain deck crew*
outdoors: *stone rock path darkness desert*

Coherent Descriptors

RMN

outdoors: *outdoors trail trails hillside grassy slopes*
sadness: *regretful rueful pity pained despondent*
education: *teaching graduate year teacher attended*
love: *love delightful happiness enjoyed enjoyable*
murder: *autopsy arrested homicide murdered*

HTMM

crime: *blood knife pain legs steal*
food: *kitchen mouth glass food bread*
violence: *sword shot blood shouted swung*
boats: *ship boat captain deck crew*
outdoors: *stone rock path darkness desert*

Experiment 2:
Trajectory Quality

Visualizing Trajectories

- for all time steps t, compute *argmax* of d_t and stack vertically

Visualizing Trajectories

- for all time steps *t*, compute *argmax* of d_t and stack vertically

time	love	death	money	crime
0	0.95	0.01	0.03	0.01
1	0.8	0.01	0.18	0.01
2	0.4	0.01	0.5	0.09
3	0.3	0.01	0.2	0.5
4	0.2	0.7	0.05	0.05

Visualizing Trajectories

- for all time steps ***t***, compute *argmax* of d_t and stack vertically

time	love	death	money	crime
0	**0.95**	0.01	0.03	0.01
1	0.8	0.01	0.18	0.01
2	0.4	0.01	0.5	0.09
3	0.3	0.01	0.2	0.5
4	0.2	0.7	0.05	0.05

Visualizing Trajectories

- for all time steps *t*, compute *argmax* of d_t and stack vertically

time	love	death	money	crime
0	**0.95**	0.01	0.03	0.01
1	**0.8**	0.01	0.18	0.01
2	0.4	0.01	0.5	0.09
3	0.3	0.01	0.2	0.5
4	0.2	0.7	0.05	0.05

Visualizing Trajectories

- for all time steps *t*, compute *argmax* of d_t and stack vertically

time	love	death	money	crime
0	**0.95**	0.01	0.03	0.01
1	**0.8**	0.01	0.18	0.01
2	0.4	0.01	**0.5**	0.09
3	0.3	0.01	0.2	0.5
4	0.2	0.7	0.05	0.05

Visualizing Trajectories

- for all time steps *t*, compute *argmax* of d_t and stack vertically

time	love	death	money	crime
0	**0.95**	0.01	0.03	0.01
1	**0.8**	0.01	0.18	0.01
2	0.4	0.01	**0.5**	0.09
3	0.3	0.01	0.2	**0.5**
4	0.2	0.7	0.05	0.05

Visualizing Trajectories

- for all time steps t, compute *argmax* of d_t and stack vertically

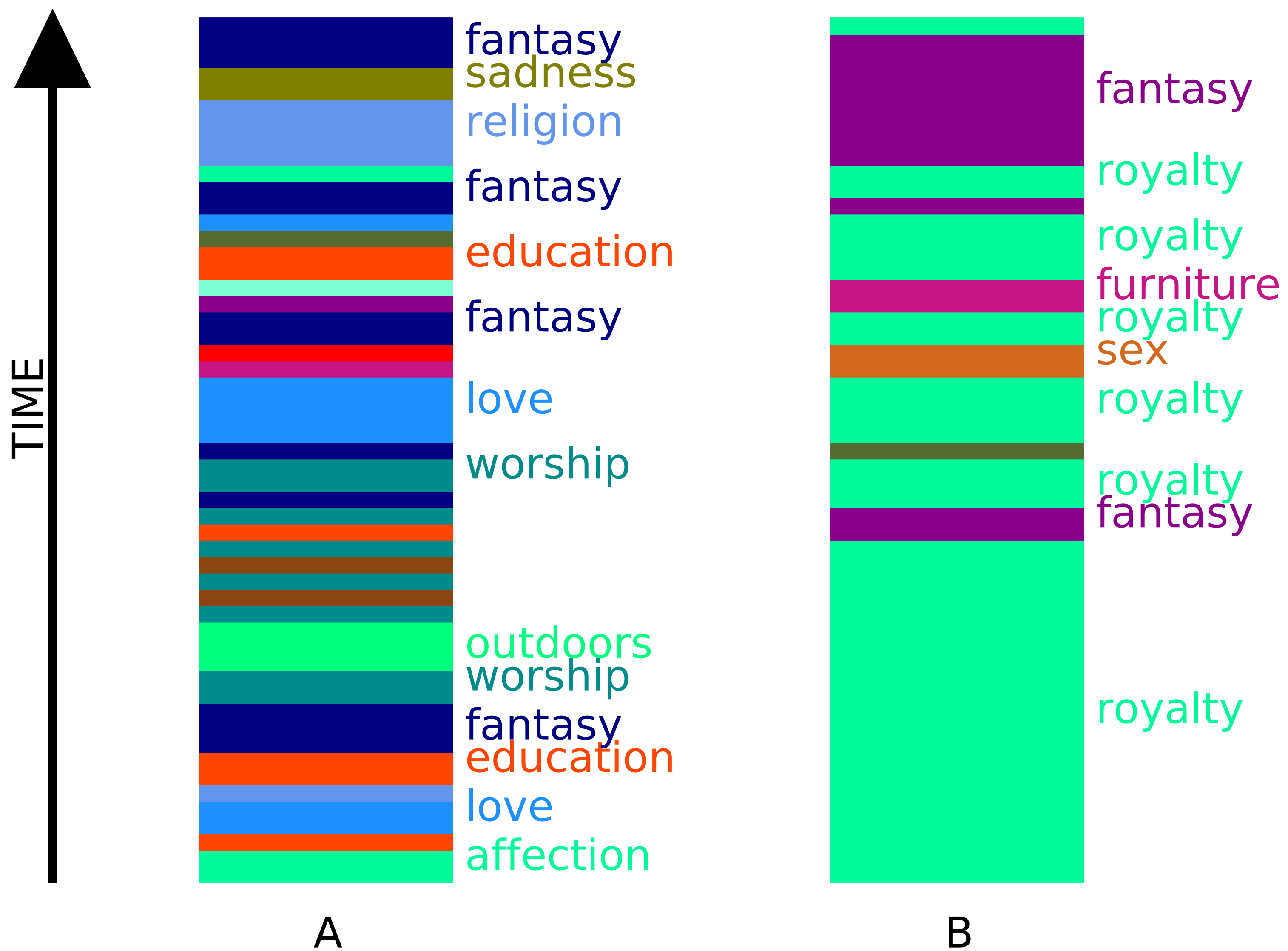
time	love	death	money	crime
0	**0.95**	0.01	0.03	0.01
1	**0.8**	0.01	0.18	0.01
2	0.4	0.01	**0.5**	0.09
3	0.3	0.01	0.2	**0.5**
4	0.2	**0.7**	0.05	0.05

Do the Trajectories Make Sense?

In this task, you will be comparing two timelines of how a relationship between a pair of literary characters changes over time. We will provide you with a summary of the relationship, and your job is to select which of the two timelines (A or B) better captures the content of the summary.

- We crawl Wikipedia and SparkNotes for summaries
- Removing uninformative summaries results in 125 character pairs to evaluate
- Workers prefer the **RMN** to the **HTMM** for 87 out of the 125 relationships (69.6%, Fleiss κ=0.32)

Siddhartha: Siddhartha AND Govinda

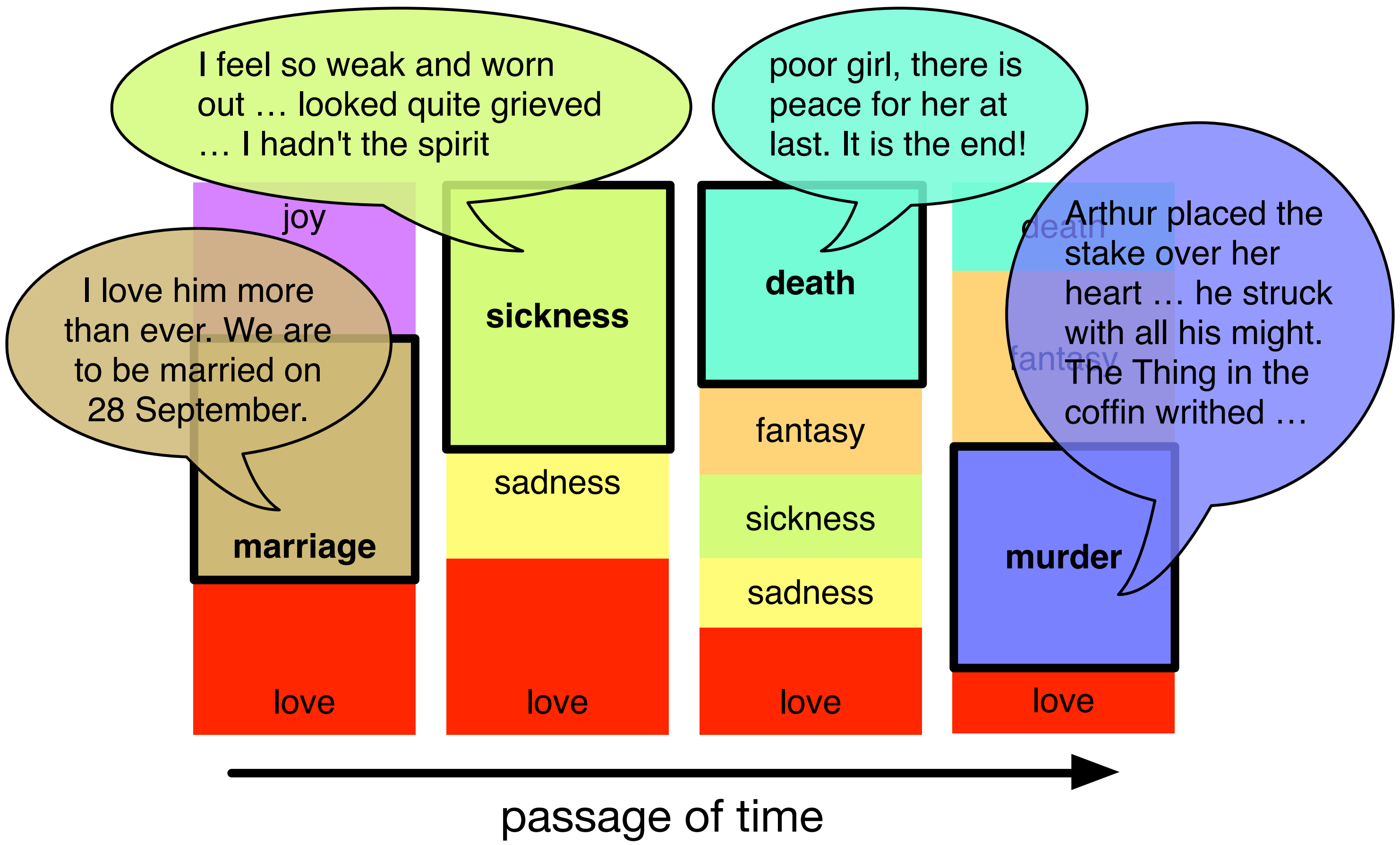


Summary: Govinda is Siddhartha's best friend and sometimes his follower. Like Siddhartha, Govinda devotes his life to the quest for understanding and enlightenment. He leaves his village with Siddhartha to join the Samanas, then leaves the Samanas to follow Gotama. He searches for enlightenment independently of Siddhartha but persists in looking for teachers who can show him the way. In the end, he is able to achieve enlightenment only because of Siddhartha's love for him.

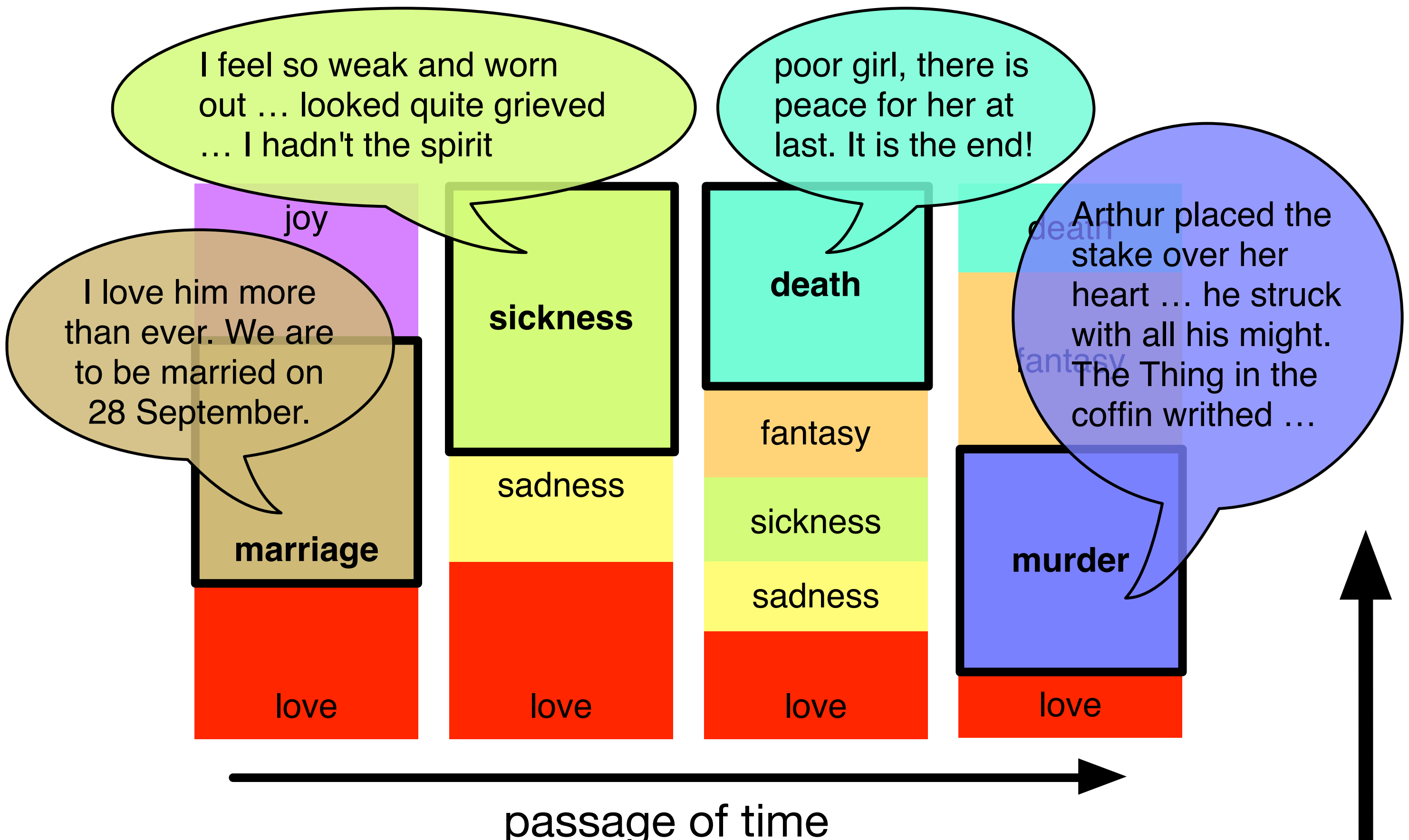
Qualitative Analysis:

Good and Bad Trajectories

Arthur and Lucy "ground-truth":
marriage -> sickness -> death -> murder

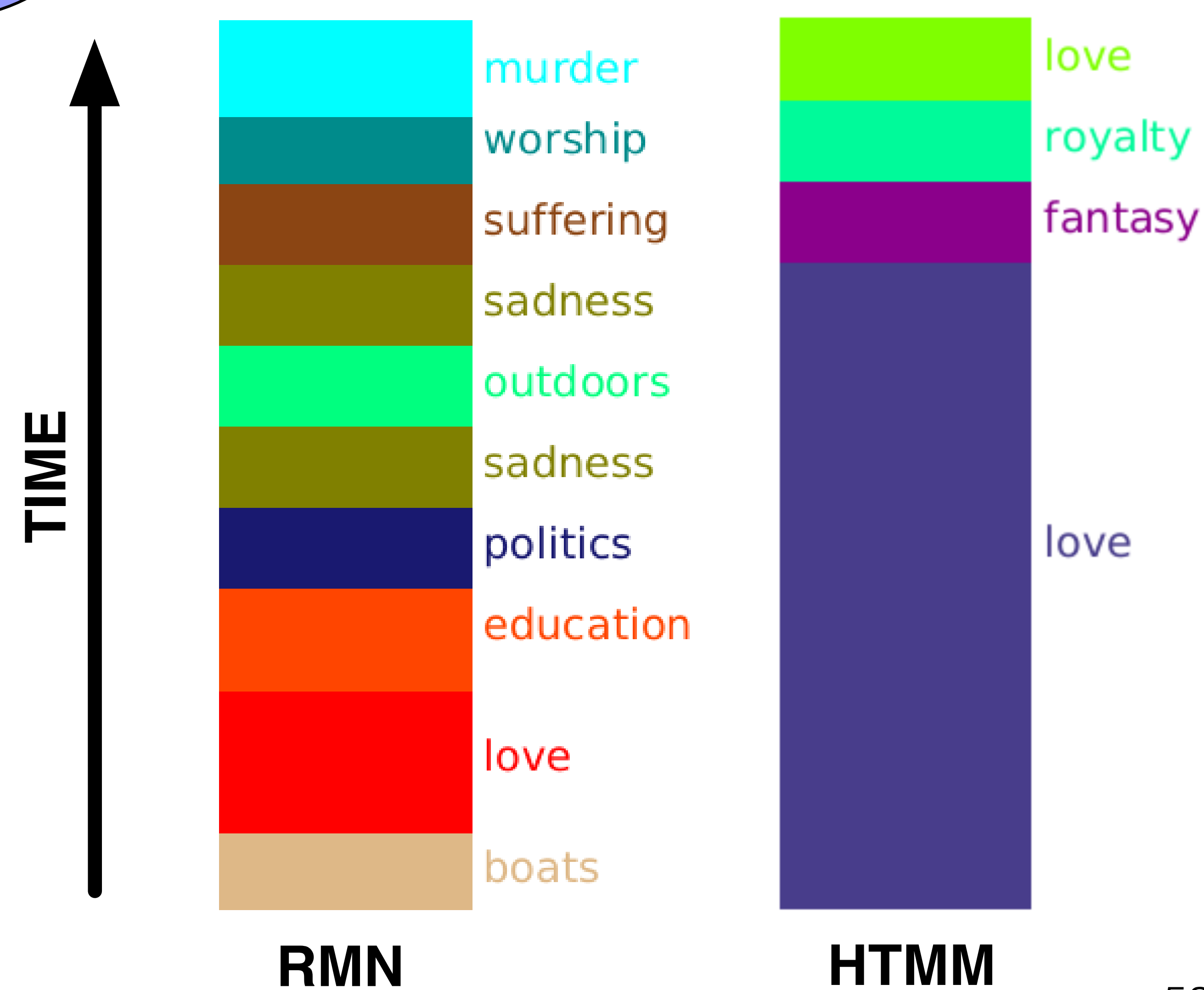


Arthur and Lucy "ground-truth":
marriage -> sickness -> death -> murder

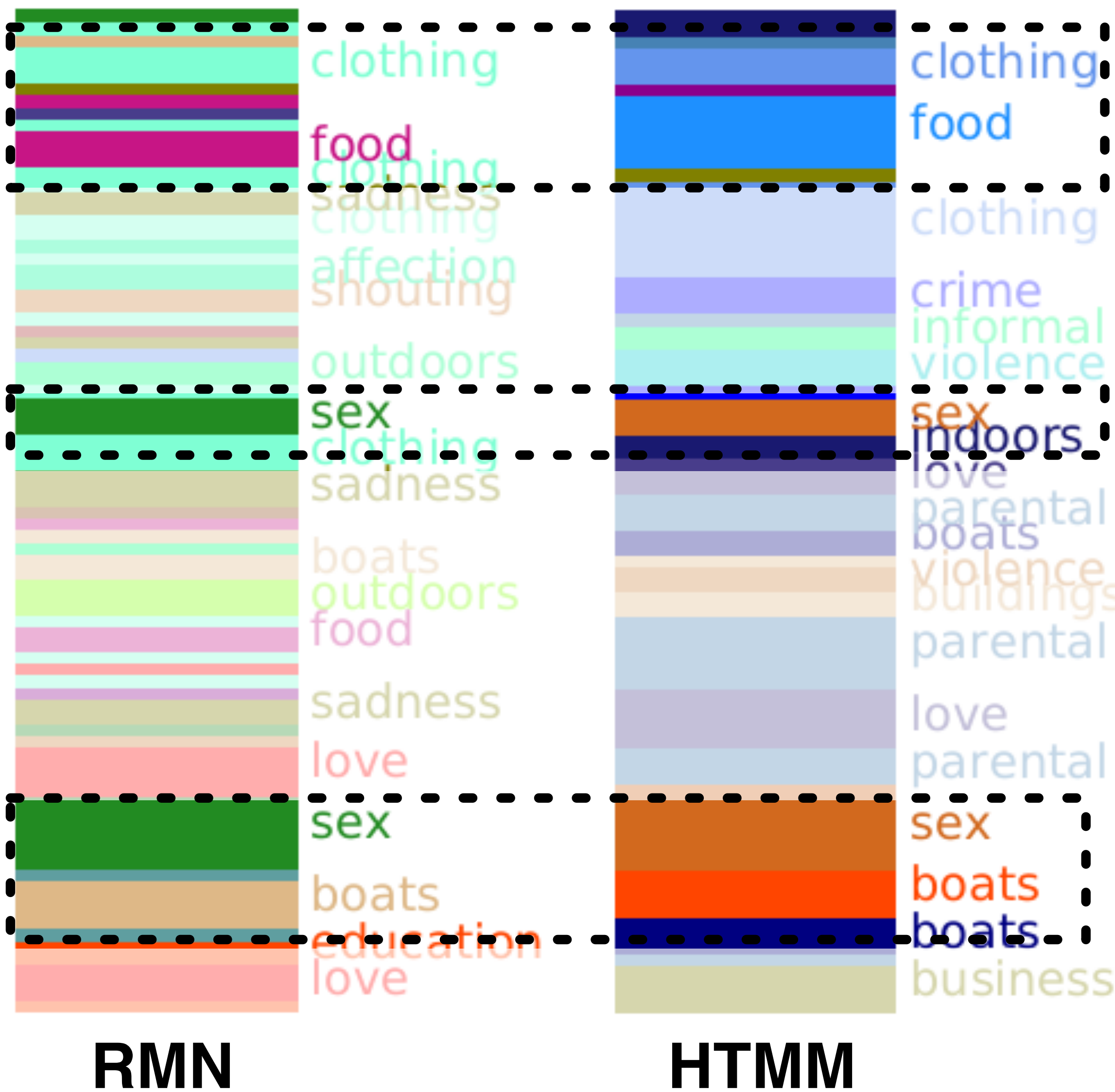


learned trajectories:

***Dracula*: Arthur and Lucy**

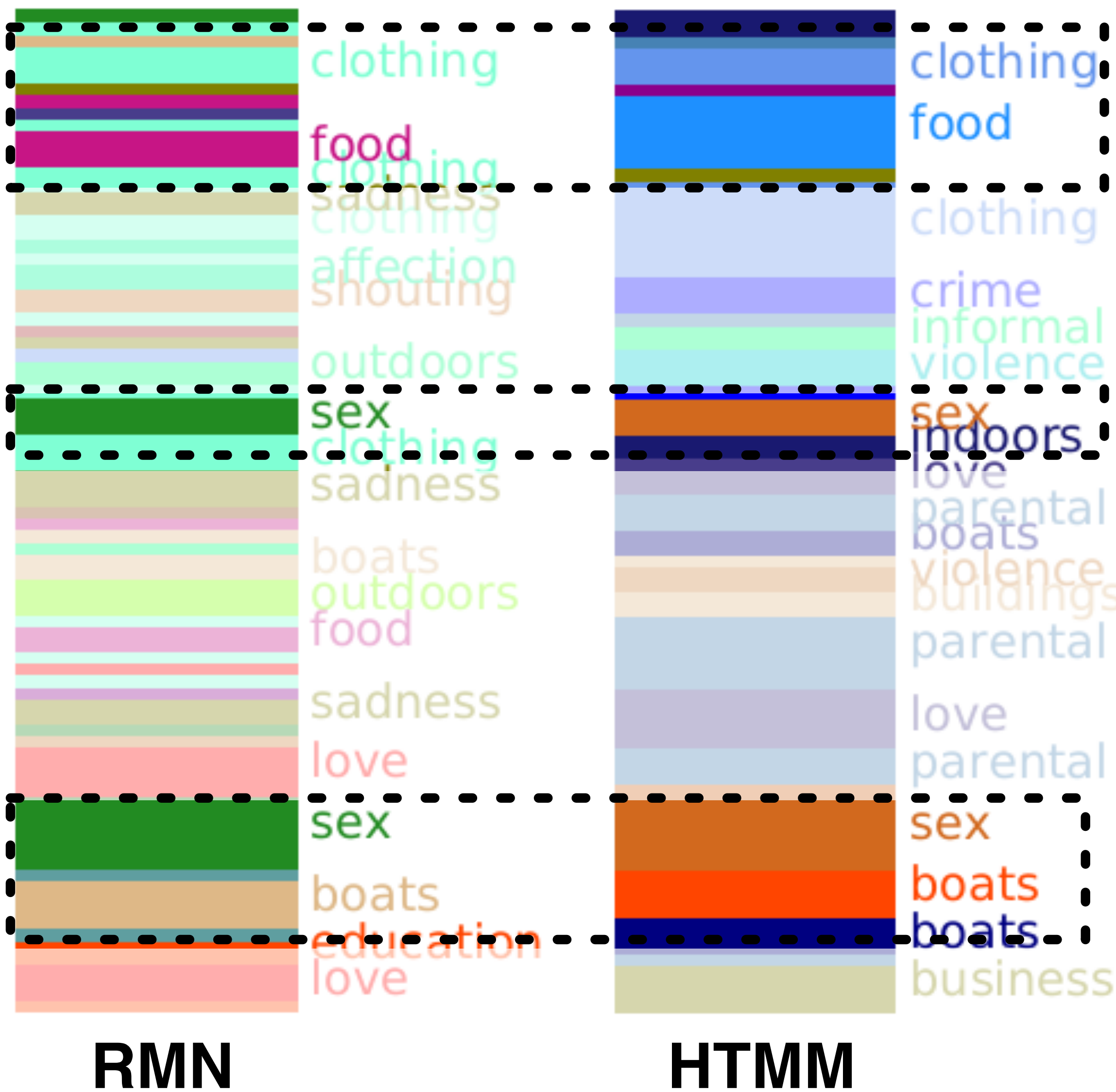


***Storm Island*: David and Lucy**



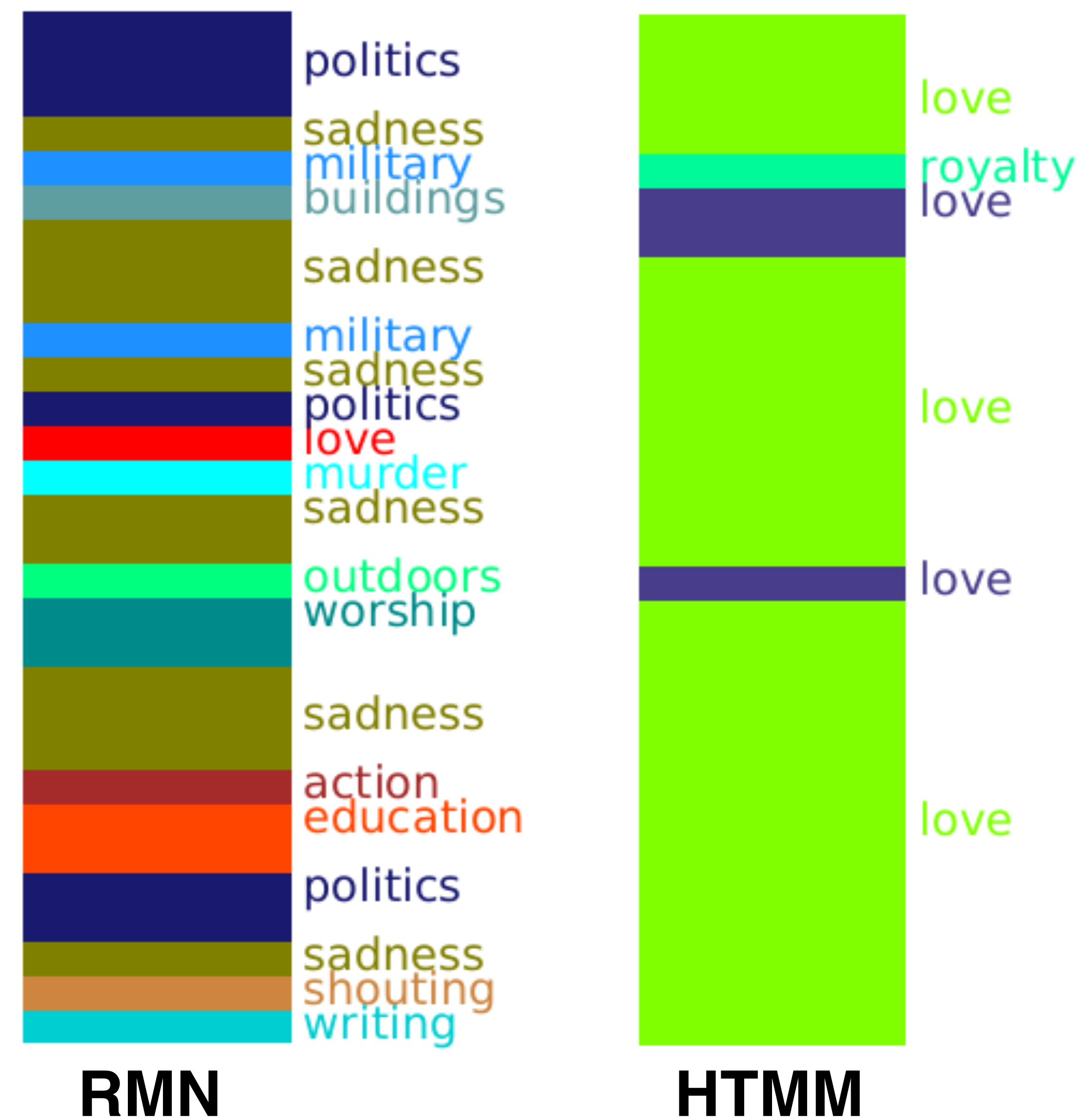
Event-based similarities between the two models

***Storm Island*: David and Lucy**



Event-based similarities between the two models

***A Tale of Two Cities*: Darnay and Lucie**



The RMN is led astray by the novel's sad tone

Qualitative Analysis:
Using Existing Datasets

What Makes a Good Relationship?

- Dataset of Massey et al. (2015) has affinity annotations for relationships in Project Gutenberg
  - 120 non-neutral relationships are also present in our dataset

What Makes a Good Relationship?

- Dataset of Massey et al. (2015) has affinity annotations for relationships in Project Gutenberg
  - 120 non-neutral relationships are also present in our dataset

positive

love	death	sadness
0.9	0.05	0.05
0.8	0.1	0.1
0.6	0.3	0.1
0.7	0.1	0.2
0.8	0.1	0.1

Don Quixote & Sancho Panza
Candide & Cunégonde
Anna Karenina & Vronsky
…

negative

love	death	sadness
0.1	0.7	0.2
0.2	0.3	0.5
0.15	0.25	0.6
0.05	0.65	0.3
0.1	0.2	0.7

Dracula & Jonathan Harker
Dr. Jekyll & Mr. Hyde
Hester Prynne & Chillingworth
…

What Makes a Good Relationship?

- Dataset of Massey et al. (2015) has affinity annotations for relationships in Project Gutenberg
  - 120 non-neutral relationships are also present in our dataset

positive

love	death	sadness
0.9	0.05	0.05
0.8	0.1	0.1
0.6	0.3	0.1
0.7	0.1	0.2
0.8	0.1	0.1
0.76	**0.13**	**0.11**

negative

love	death	sadness
0.1	0.7	0.2
0.2	0.3	0.5
0.15	0.25	0.6
0.05	0.65	0.3
0.1	0.2	0.7
0.12	**0.42**	**0.46**

average the positive and negative trajectories

What Makes a Good Relationship?

- Dataset of Massey et al. (2015) has affinity annotations for relationships in Project Gutenberg
    - 120 non-neutral relationships are also present in our dataset

positive

love	death	sadness
0.9	0.05	0.05
0.8	0.1	0.1
0.6	0.3	0.1
0.7	0.1	0.2
0.8	0.1	0.1
0.76	**0.13**	**0.11**

1. love
2. death
3. sadness

negative

love	death	sadness
0.1	0.7	0.2
0.2	0.3	0.5
0.15	0.25	0.6
0.05	0.65	0.3
0.1	0.2	0.7
0.12	**0.42**	**0.46**

1. sadness
2. death
3. love

Most Positive Descriptors

RMN

education
love
religion
sex

HTMM

love
parental
business
outdoors

Most Positive Descriptors

RMN

education

love

religion

sex

HTMM

love

parental

business

outdoors

Most Negative Descriptors

RMN

politics

murder

sadness

royalty

HTMM

love

politics

violence

crime

Most Positive Descriptors	Most Negative Descriptors
RMN	**RMN**
education	**politics**
love	murder
religion	sadness
sex	royalty
HTMM	**HTMM**
love	love
parental	**politics**
business	violence
outdoors	crime

Why is Politics Negative?

- Both models rank **politics** as highly negative
- The affinity data we look at comes primarily from Victorian-era authors (e.g., Charles Dickens and George Eliot)

Victorian-era authors are "obsessed with otherness… of antiquated social and legal institutions, and of autocratic and/or dictatorial abusive government"
(Zarifopol-Johnston,1995)

Areas for Improvement

- Difficult to evaluate unsupervised relationship modeling, requires considerable human effort
- Our data processing leaves out a lot of information
  - e.g., spans of text in which one but not both characters in a relationship are mentioned
  - only considers *undirected* relationships between *pairs*
- Model performance is directly tied to the quality of character disambiguation and coreference resolution
  - e.g., first person pronouns

Recap

- Introduced the task of unsupervised relationship modeling as well as an *interpretable* neural network architecture, the RMN, for this task

- Found that the RMN generates higher quality descriptors and more interpretable trajectories than topic model baselines

- Future work: collaborate with humanities researchers to help answer literary questions with the RMN

Thanks! Questions?

code/data @ github.com/miyyer/rmn