Sentence-Level Content Planning and Style Specification for Neural Text Generation

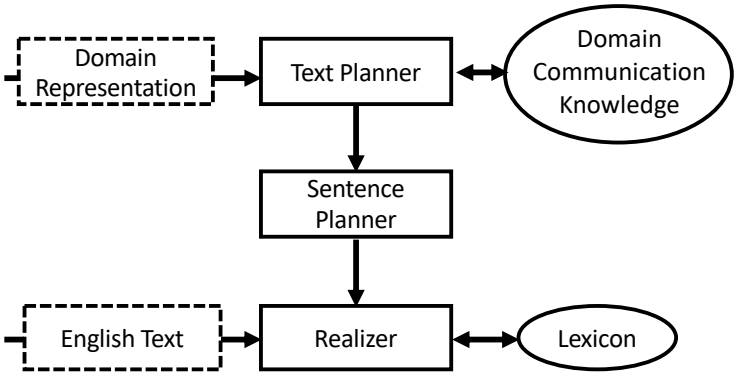
Xinyu Hua and Lu Wang

EMNLP 2019

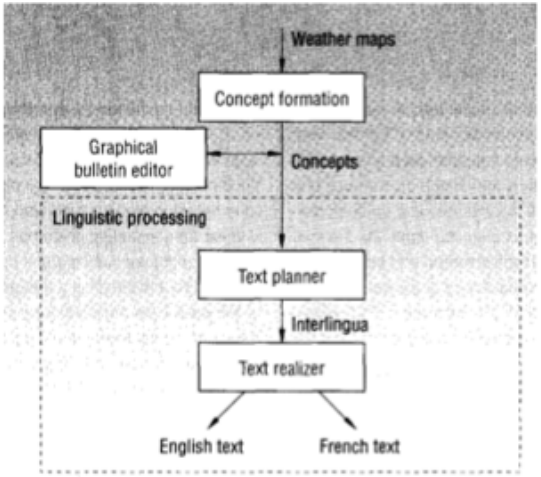
Northeastern University

Motivation

- Text generation in early days



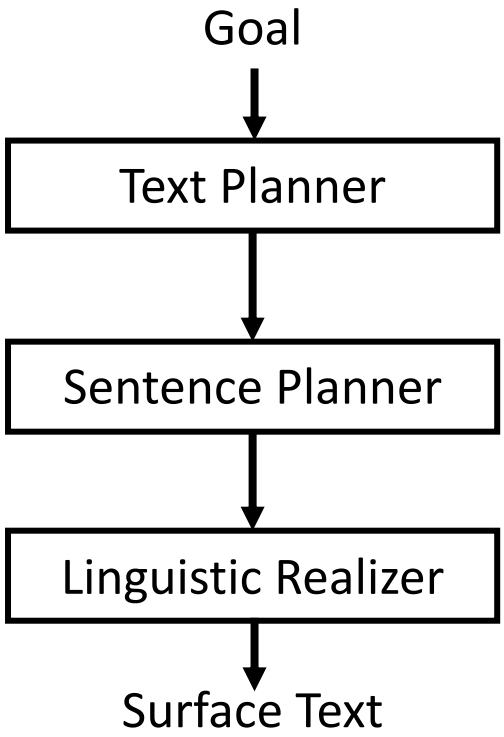
(Joyce [Rambow and Korelsky, 1992])



(FOG [Goldber et al, 1994])

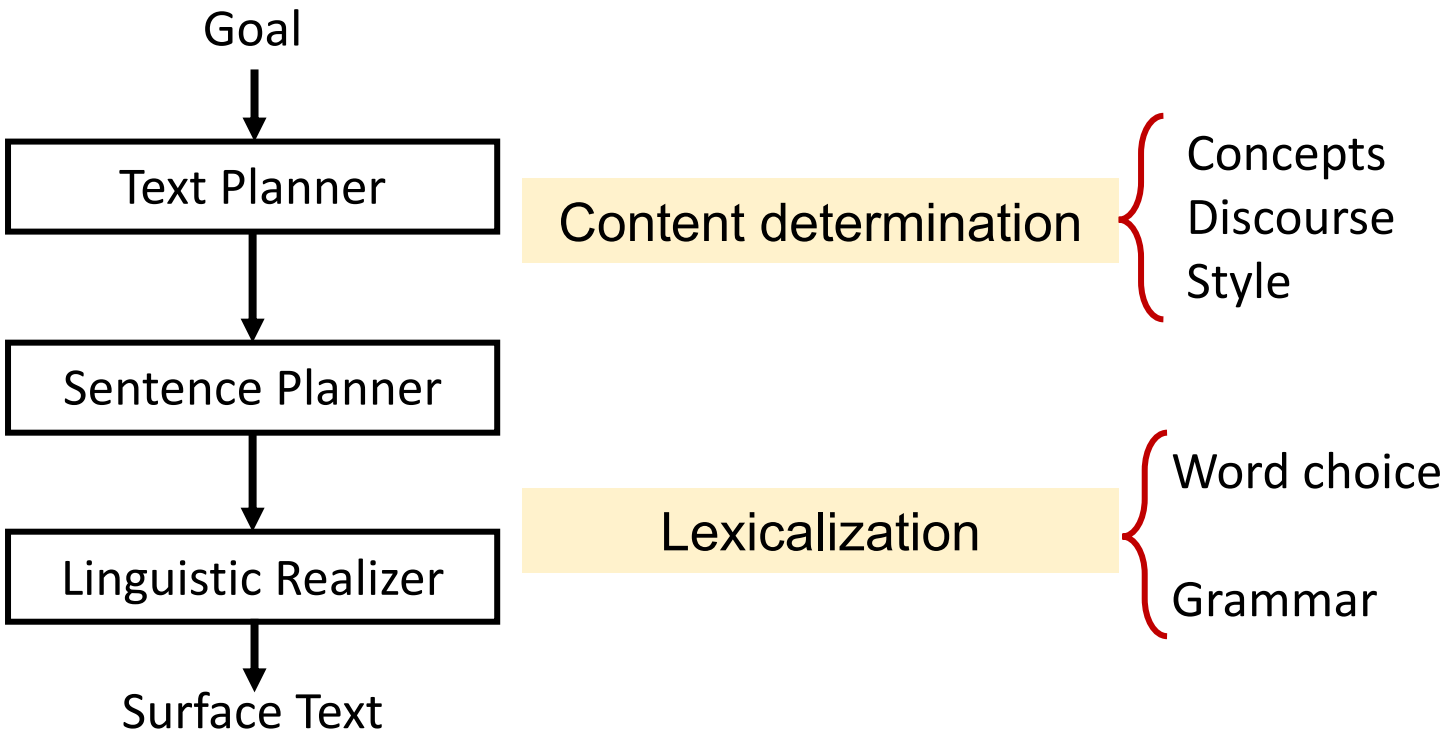
Motivation

- A commonly used architecture [Reiter and Dale, 2000]



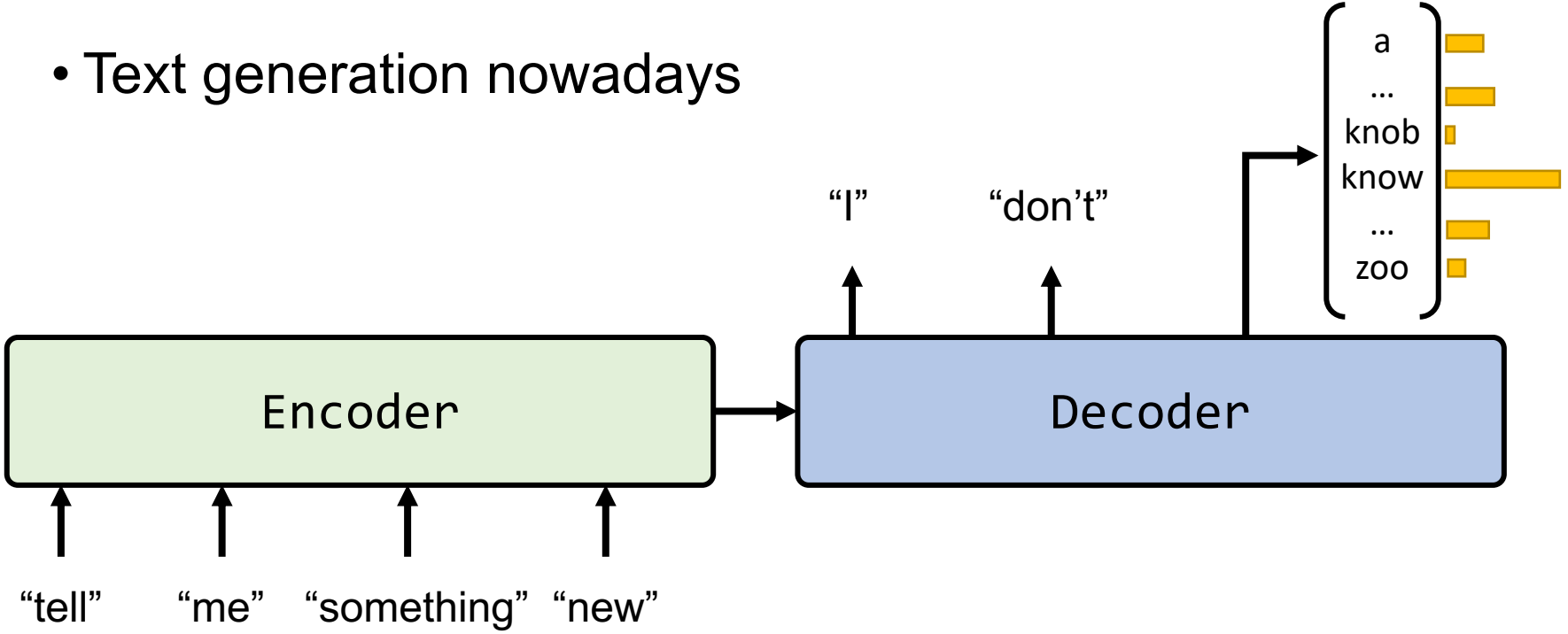
Motivation

- A commonly used architecture [Reiter and Dale, 2000]



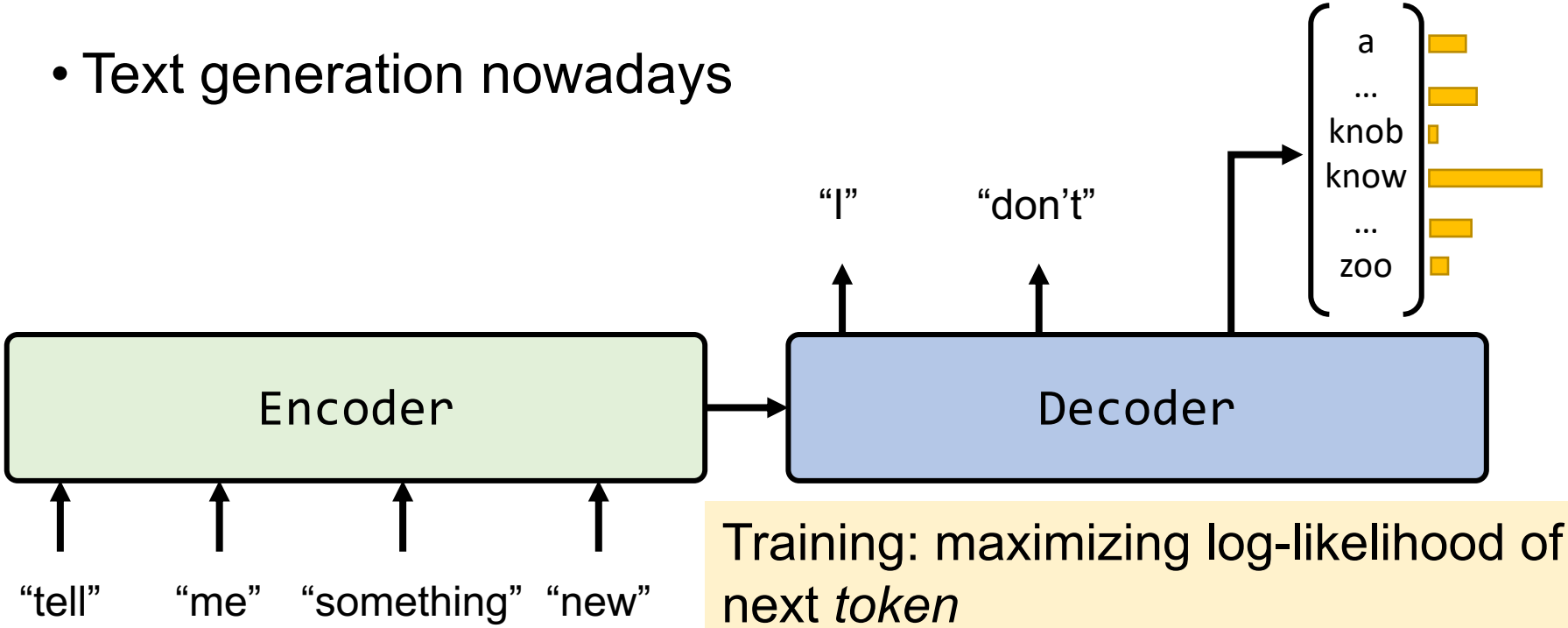
Motivation

- Text generation nowadays



Motivation

- Text generation nowadays



Training: maximizing log-likelihood of next *token*
Inference: autoregressive decoding

Motivation

Is this how humans write?

Motivation

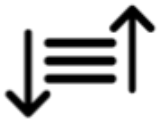
- A cognitive model of writing [Hayes and Flower, 1983]

GENERATING: "*retrieve information relevant to the writing task from long-term memory*".

Motivation

- A cognitive model of writing [Hayes and Flower, 1983]

GENERATING: "*retrieve information relevant to the writing task from long-term memory*".

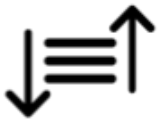


ORGANIZING: "*select the most useful of the materials retrieved by the GENERATING process and to organize them into a writing plan.*"**

Motivation

- A cognitive model of writing [Hayes and Flower, 1983]

GENERATING: "*retrieve information relevant to the writing task from long-term memory*".



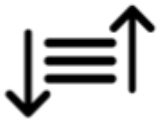
ORGANIZING: "*select the most useful of the materials retrieved by the GENERATING process and to organize them into a writing plan.***"**

EDITING: *"evaluate material with respect to the writing goals."*

Motivation

- A cognitive model of writing [Hayes and Flower, 1983]

GENERATING: "*retrieve information relevant to the writing task from long-term memory*".



ORGANIZING: "*select the most useful of the materials*

em

ls."

- Neural NLG models are great, but they are handling multiple distinct tasks at once.

- Our goal is to separate planning and realization, for better control over the output.

Roadmap

- Motivation
- Tasks
- Model
- Data
- Evaluation
- Conclusion

Roadmap

- Motivation
- Tasks
- Model
- Data
- Evaluation
- Conclusion

Tasks

Subjective
Open-ended
High-entropy

Objective
Factoid
Low-entropy

Tasks

Argument generation

Paper abstract generation

Wikipedia paragraph generation

Subjective
Open-ended
High-entropy

Objective
Factoid
Low-entropy

Argument generation

r/changemyview

Posted by u/CherrySlurpee 16Δ 6 years ago

I think the United States should cut off foreign aid completely. CMV

2011 saw 49.5B in spending on foreign aid. Why aren't US tax dollars being spent to help US

+

cut financial aid | allies | economic benefits | ...

↓

Foreign aid allows for allies in places that are economically advantageous.

Paper abstract generation

AGENDA

Title: Semantic Embeddings from Hashtags

+

Short textual posts | hashtags | Hashtag prediction | ...

↓

We describe a convolutional neural network that learns feature representations for …

Wikipedia paragraph generation

WIKIPEDIA

Title: Artificial intelligence

+

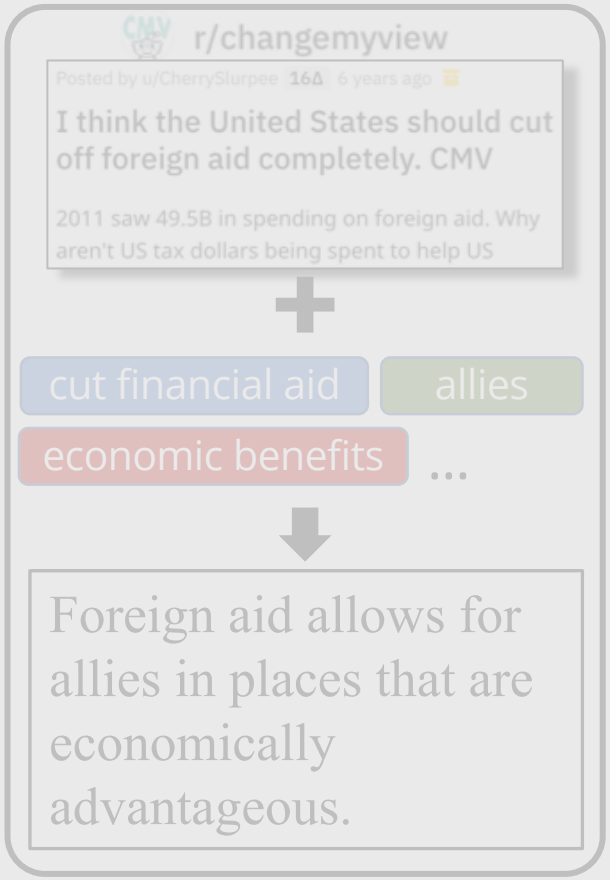
Style: ☐ simple ■ normal

computer science | perceive | artificial | machines

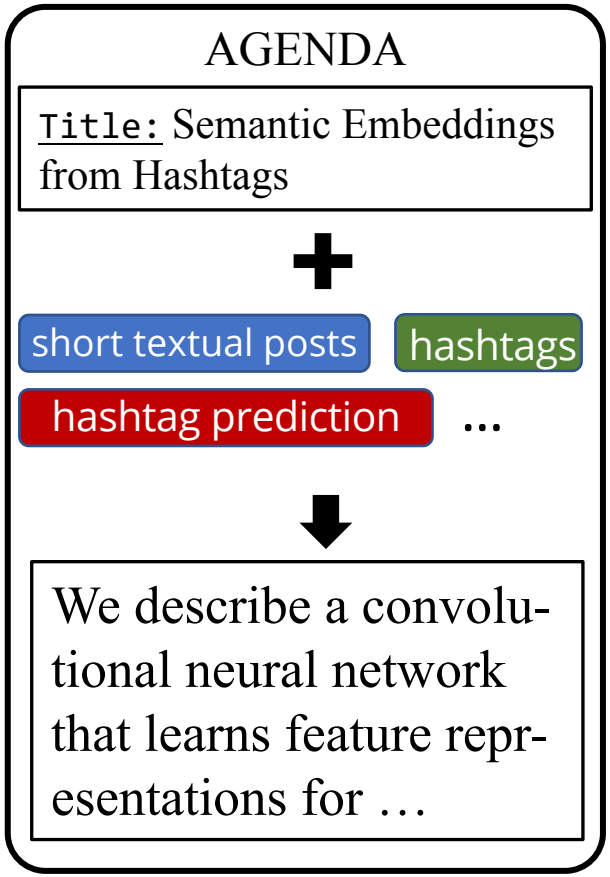
↓

In computer science, artificial intelligence (AI), sometimes called machine intelligence…

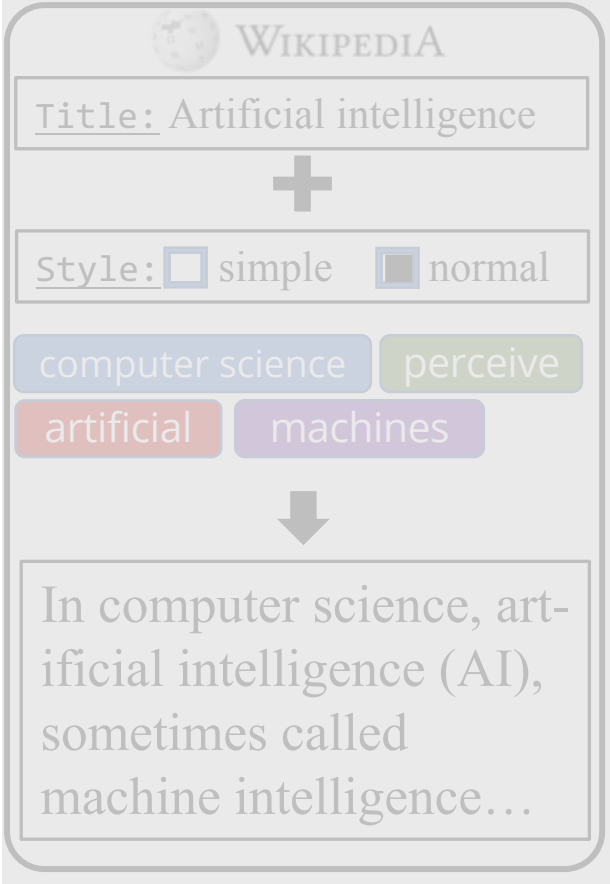
Argument generation



Paper abstract generation



Wikipedia paragraph generation



Argument generation

r/changemyview

Posted by u/CherrySlurpee 16Δ 6 years ago

I think the United States should cut off foreign aid completely. CMV

2011 saw 49.5B in spending on foreign aid. Why aren't US tax dollars being spent to help US

cut financial aid | allies | economic benefits | ...

Foreign aid allows for allies in places that are economically advantageous.

Paper abstract generation

AGENDA

Title: Semantic Embeddings from Hashtags

Short textual posts | hashtags | Hashtag prediction | ...

We describe a convolutional neural network that learns feature representations for …

Wikipedia paragraph generation

WIKIPEDIA

Title: Artificial intelligence

Style: simple normal

computer science | perceive | artificial | machines

In computer science, artificial intelligence (AI), sometimes called machine intelligence…

Roadmap

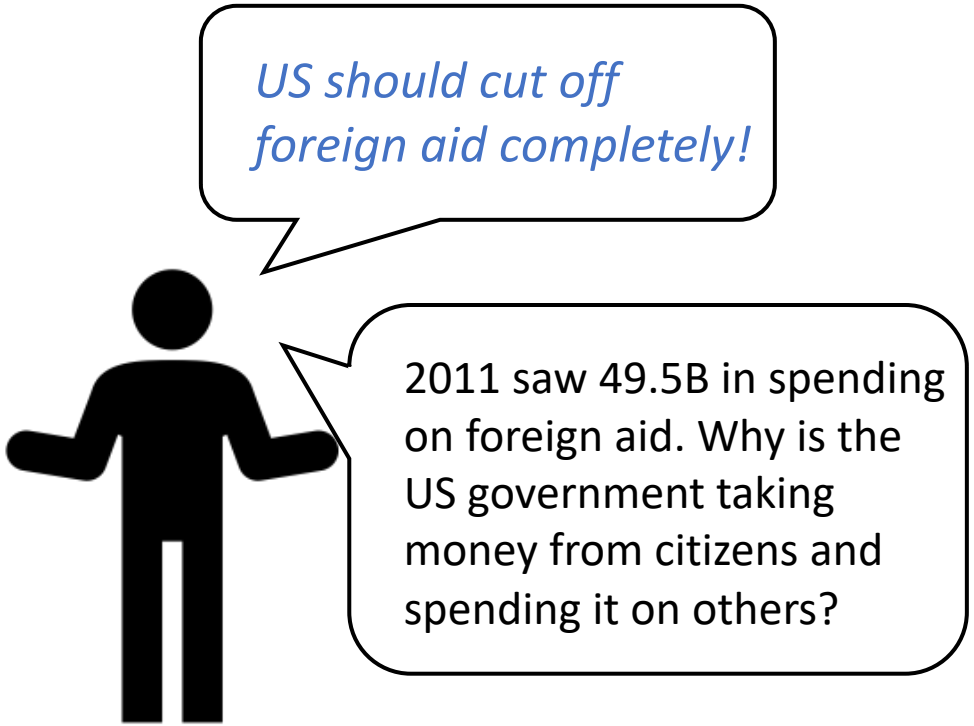
- Motivation
- Tasks
- Model
- Data
- Evaluation
- Conclusion

Model

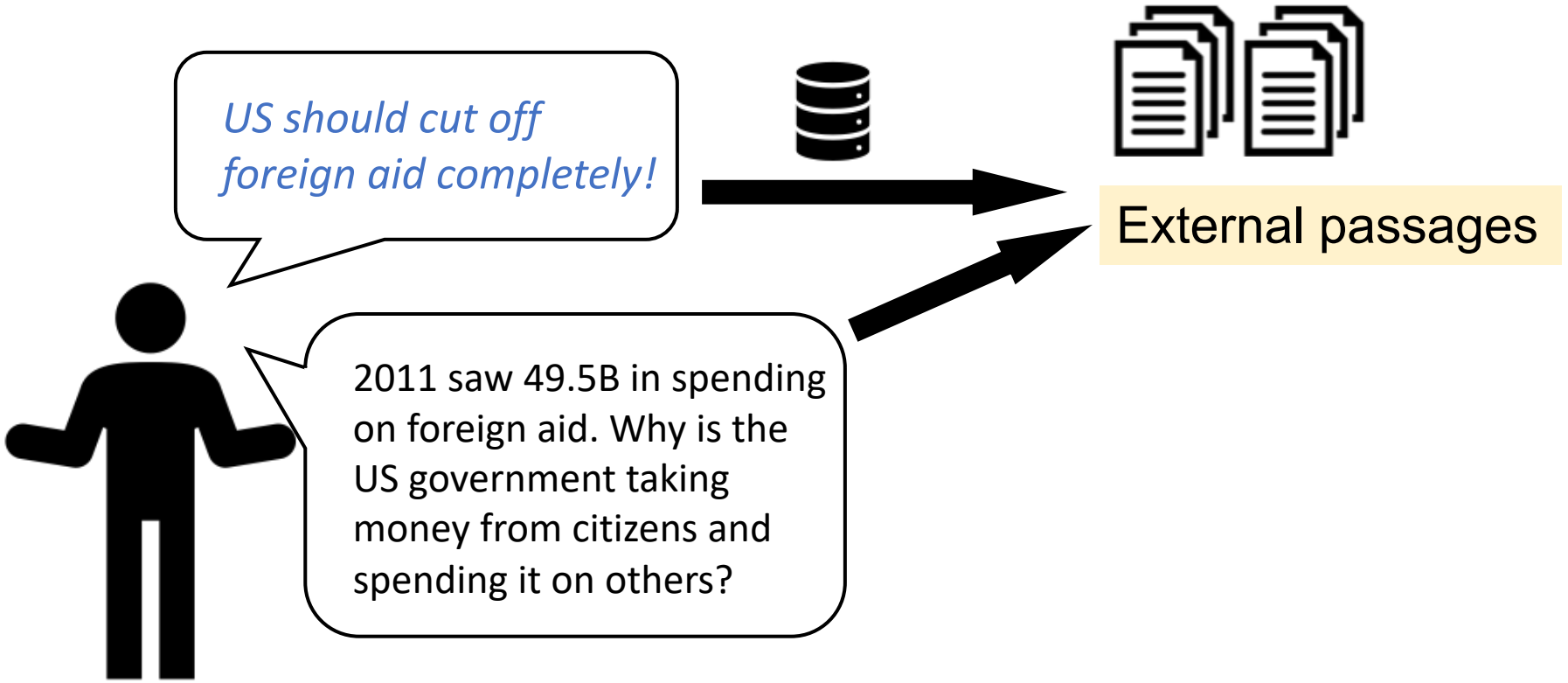
- Example: argument generation

Model

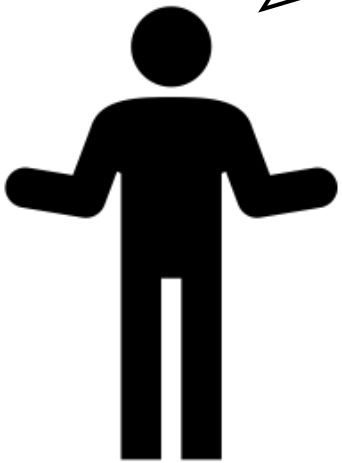
Model


US should cut off foreign aid completely!
2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?

Model


US should cut off foreign aid completely!
External passages
2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?

Model



US should cut off foreign aid completely!

- cut financial aid
- make homosexuality a crime
- uganda
- …
- political bargaining chip

Model

Sentence 1: [political bargaining chip]

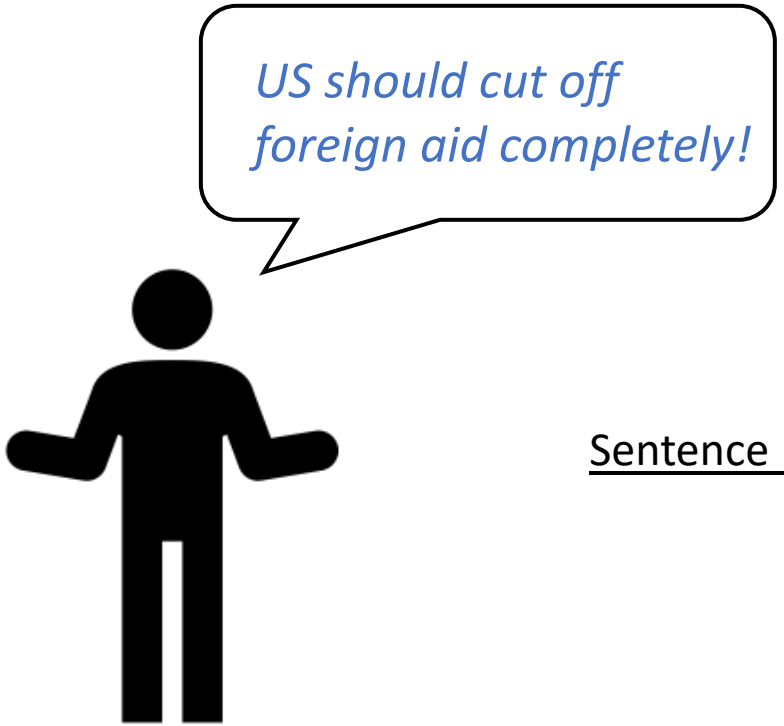
Sentence 2: [cut financial aid; uganda]

Sentence 3: [make homosexuality a crime]

Sentence 4: [NULL]

- cut financial aid
- make homosexuality a crime
- uganda
- …
- political bargaining chip

Model



Sentence 1: [**political bargaining chip**]

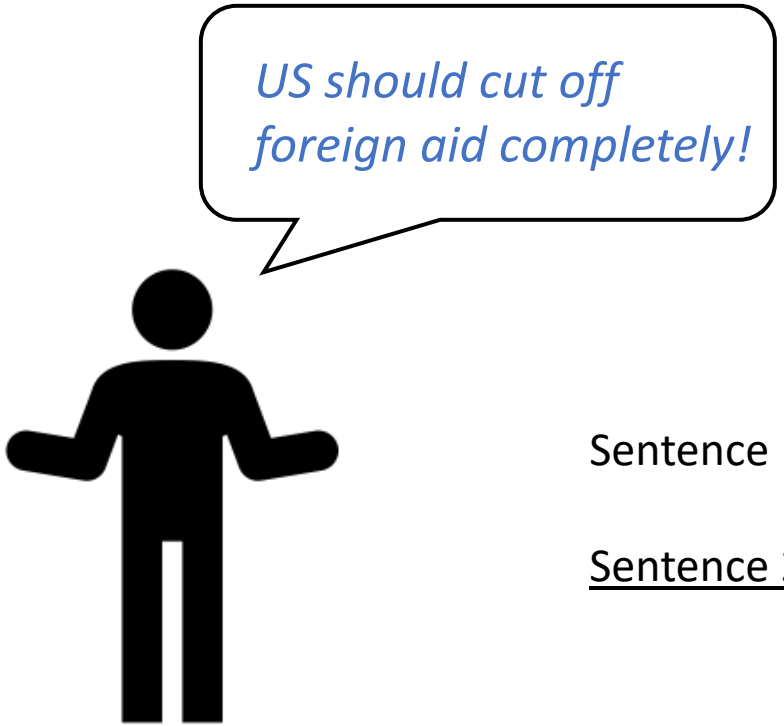
Sentence 2: [cut financial aid; uganda]

Sentence 3: [make homosexuality a crime]

Sentence 4: [NULL]

Sentence 1: It can be a useful political bargaining chip.

Model



Sentence 1: [political bargaining chip]

Sentence 2: [**cut financial aid**; **uganda**]

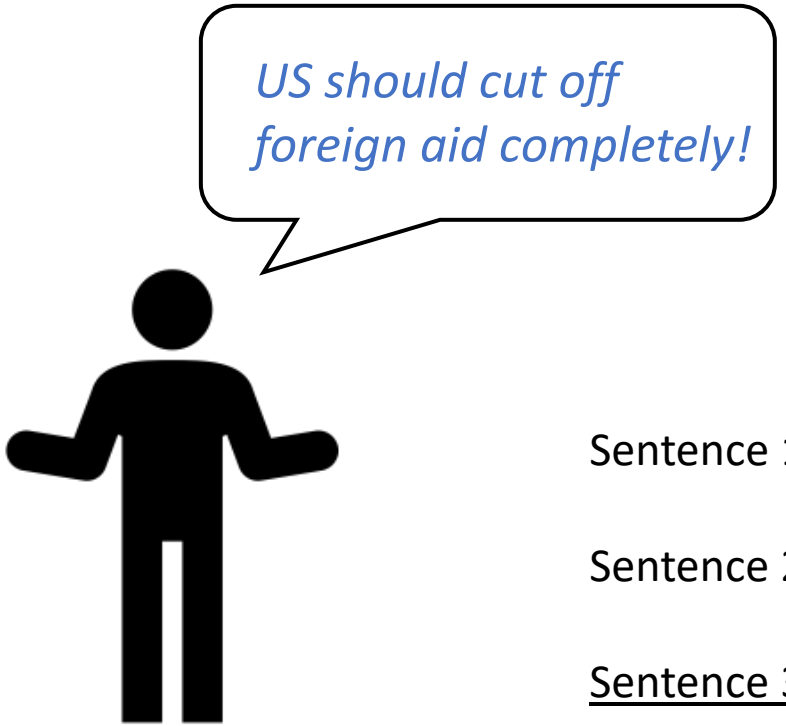
Sentence 3: [make homosexuality a crime]

Sentence 4: [NULL]

Sentence 1: [illegible]an be a useful political bargaining chip.

Sentence 2: US threatened to cut off financial aid to Uganda.

Model



Sentence 1: [political bargaining chip]

Sentence 2: [cut financial aid; uganda]

Sentence 3: [**make homosexuality a crime**]

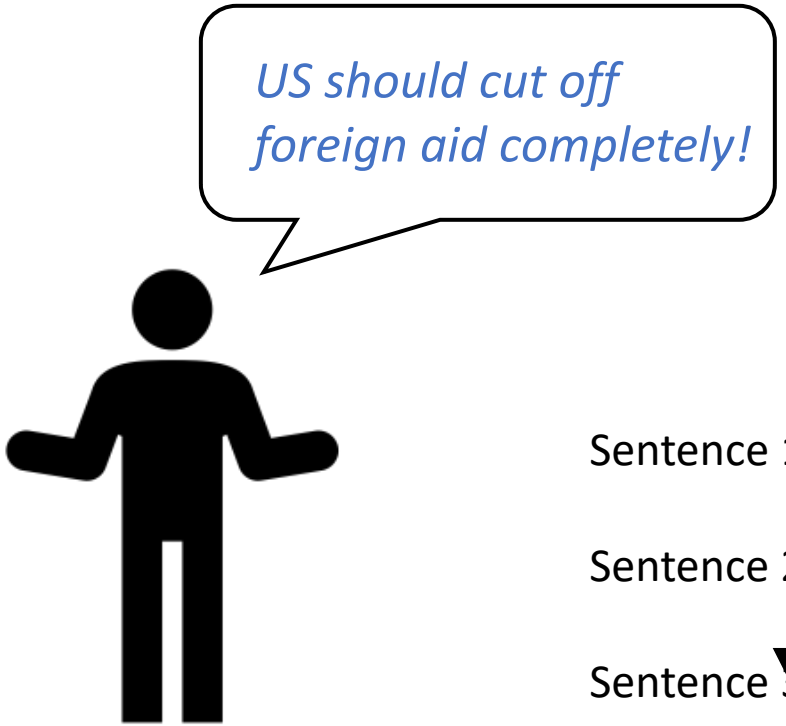
Sentence 4: [NULL]

Sentence 1: It can be a useful political bargaining chip.

Sentence 2: U threatened to cut off financial aid to Uganda.

Sentence 3: Because it planed to criminalize homosexuality.

Model



Sentence 1: [political bargaining chip]

Sentence 2: [cut financial aid; uganda]

Sentence 3: [make homosexuality a crime]

Sentence 4: [**NULL**]

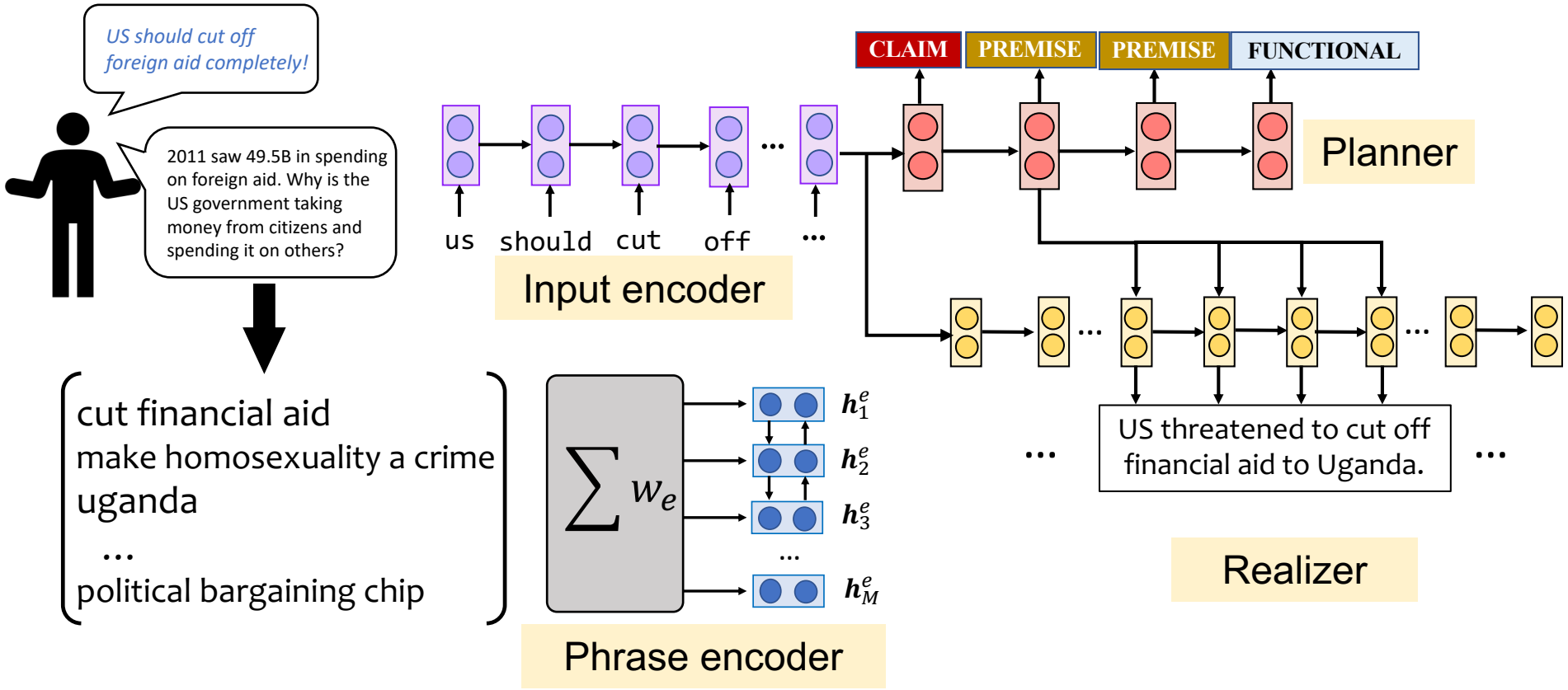
Sentence 1: It can be a useful political bargaining chip.

Sentence 2: US threatened to cut off financial aid to Uganda.

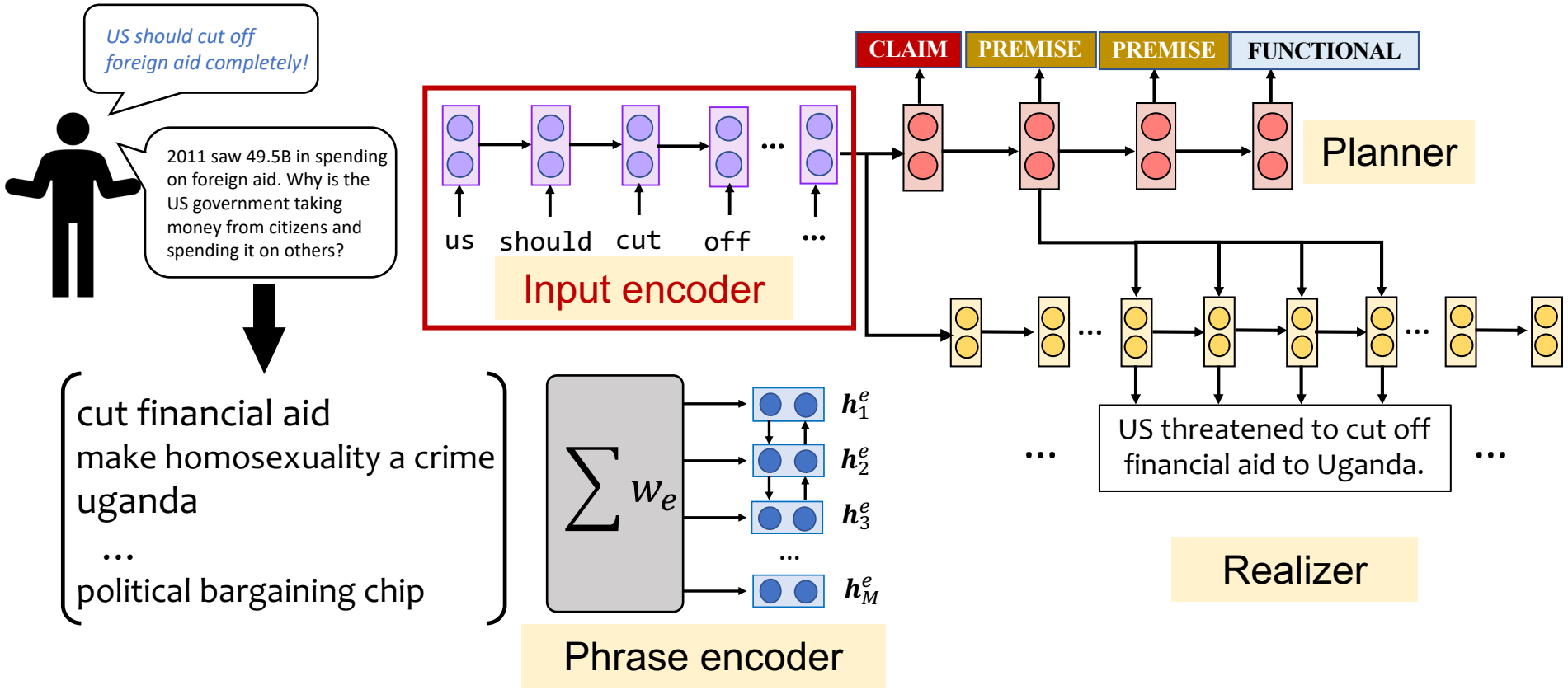
Sentence 3: Because it planed to criminalize homosexuality.

Sentence 4: Please consider change your mind!

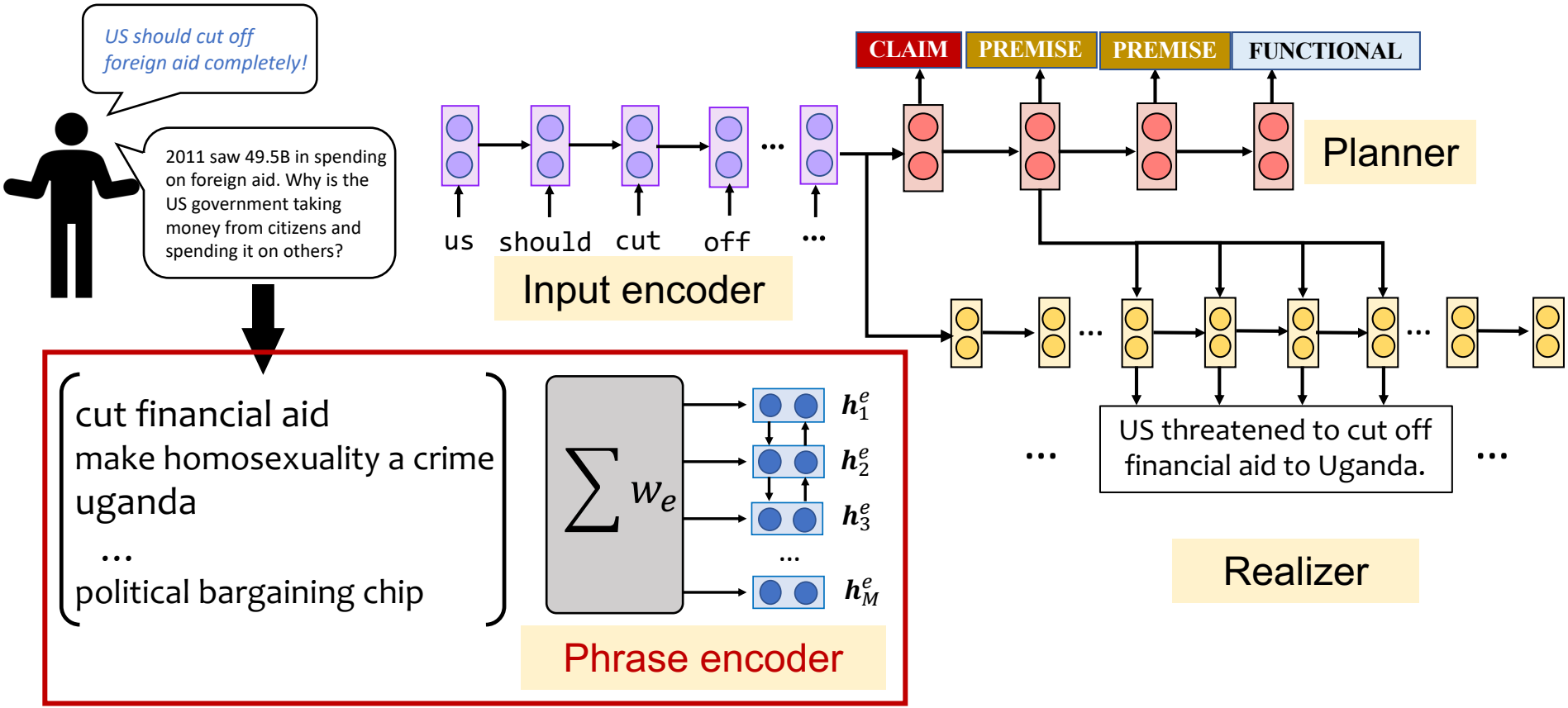
Model


US should cut off foreign aid completely!
2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?
cut financial aid
make homosexuality a crime
uganda
...
political bargaining chip
us
should
cut
off
...
Input encoder
CLAIM
PREMISE
PREMISE
FUNCTIONAL
Planner
US threatened to cut off financial aid to Uganda.
Realizer
$\sum w_e$
h_1^e
h_2^e
h_3^e
h_M^e
Phrase encoder

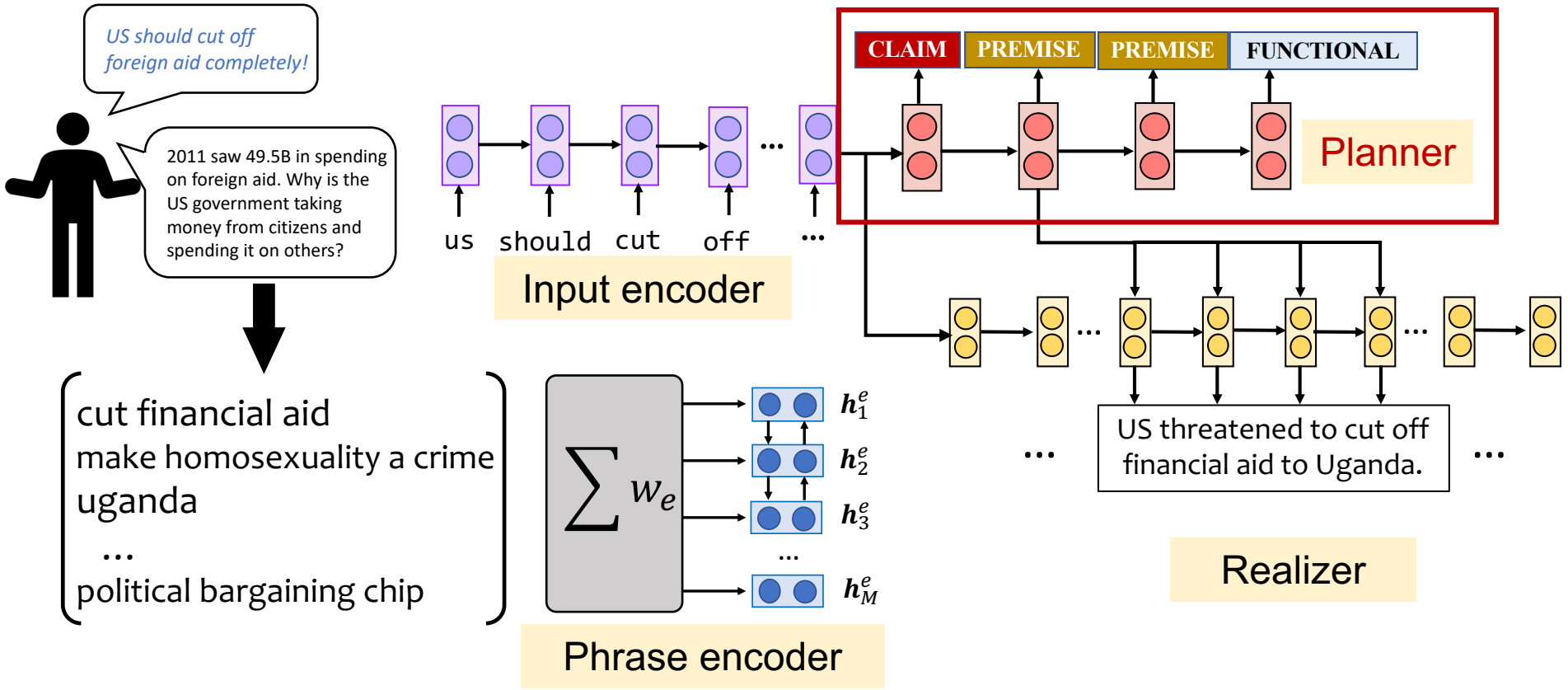
Model


US should cut off foreign aid completely!
2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?
cut financial aid
make homosexuality a crime
uganda
...
political bargaining chip
us should cut off ...
Input encoder
CLAIM
PREMISE
PREMISE
FUNCTIONAL
Planner
US threatened to cut off financial aid to Uganda.
Realizer
$\sum w_e$
h_1^e
h_2^e
h_3^e
...
h_M^e
Phrase encoder

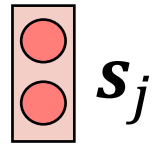
Model


US should cut off foreign aid completely!
2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?
cut financial aid
make homosexuality a crime
uganda
...
political bargaining chip
us
should
cut
off
Input encoder
CLAIM
PREMISE
PREMISE
FUNCTIONAL
Planner
US threatened to cut off financial aid to Uganda.
Realizer
Phrase encoder

Model


US should cut off foreign aid completely!
2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?
cut financial aid
make homosexuality a crime
uganda
...
political bargaining chip
us should cut off ...
Input encoder
CLAIM PREMISE PREMISE FUNCTIONAL
Planner
US threatened to cut off financial aid to Uganda.
Realizer
$\sum w_e$
h_1^e h_2^e h_3^e ... h_M^e
Phrase encoder

Content Planning

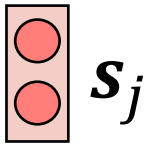


s_j

Planner's hidden states

Sentence 1: [political bargaining chip]

Content Planning

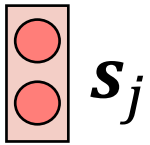


$\boldsymbol{s}_j$

Sentence 1: [**political bargaining chip**]

$$\boldsymbol{s}_j = LSTM(\boldsymbol{s}_{j-1}, \boldsymbol{m}_j)$$

Content Planning



$\boldsymbol{s}_j$

Sentence 1: [political bargaining chip]

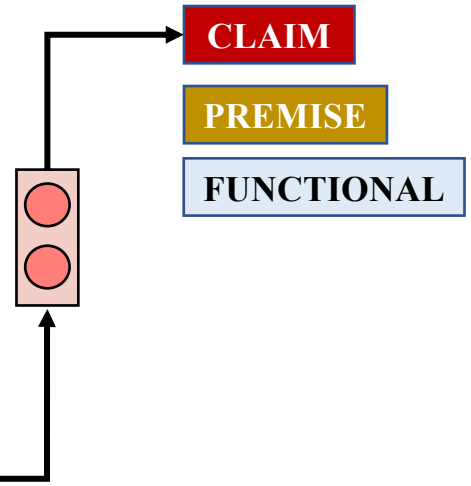
$$\boldsymbol{s}_j = LSTM\left(\boldsymbol{s}_{j-1}, \boldsymbol{m}_j\right)$$

$$\boldsymbol{m}_j = \sum_{sel(k)} \boldsymbol{h}_k^e$$

Content Planning

$$softmax(\mathbf{w}_s^T \tanh(\mathbf{W}^s[\boldsymbol{m}_j; \boldsymbol{s}_j]))$$

Style specification

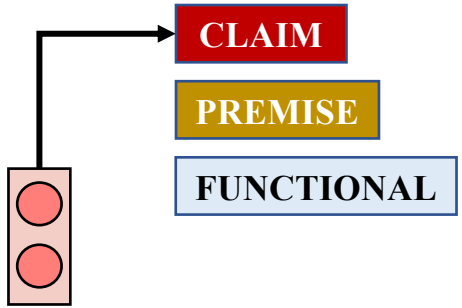


Sentence 1: [political bargaining chip]

Content Planning

$$softmax(\mathbf{w}_s^T \tanh(\mathbf{W}^s[\boldsymbol{m}_j; \boldsymbol{s}_j]))$$

Style specification



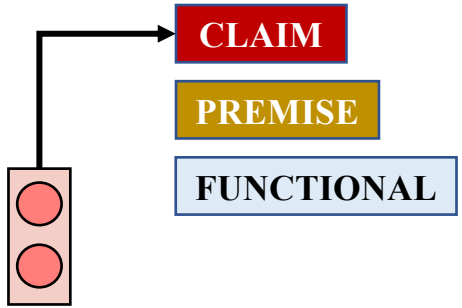
CLAIM: "I believe foreign aid is a useful bargaining chip."

Sentence 1: [political bargaining chip]

Content Planning

$$softmax(\mathbf{w}_s^T \tanh(\mathbf{W}^s[\boldsymbol{m}_j; \boldsymbol{s}_j]))$$

Style specification



Sentence 1: [political bargaining chip]

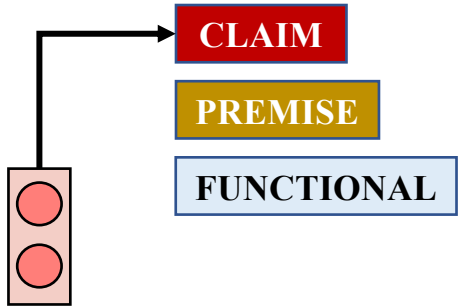
CLAIM: "I believe foreign aid is a useful bargaining chip."

PREMISE: "In 2014, the US cuts aid to Uganda over anti-gay law."

Content Planning

$$softmax(\mathbf{w}_s^T \tanh(\mathbf{W}^s[\boldsymbol{m}_j; \boldsymbol{s}_j]))$$

Style specification



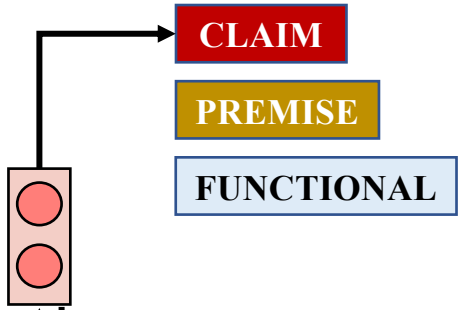
Sentence 1: [political bargaining chip]

CLAIM: "I believe foreign aid is a useful bargaining chip."

PREMISE: "In 2014, the US cuts aid to Uganda over anti-gay law."

FUNCTIONAL: "Please change your mind!"

Content Planning



Sentence 1: [political bargaining chip]

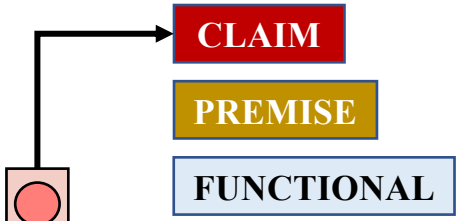
Sentence 2: [cut financial aid; uganda]

Keyphrase selection

$$P\left(\boldsymbol{v}_{j+1,k} = 1 | \boldsymbol{v}_{1:j}\right) = \sigma(\mathbf{w}_v^T \boldsymbol{s}_j + \boldsymbol{q}_j \mathbf{W}^c \boldsymbol{h}_k^e)$$

select *k*-th phrase in (*j*+1)-th sentence

Content Planning



$$\boldsymbol{q}_j = \sum_{k=1}^{|M|} \sum_{r=0}^{j} \boldsymbol{v}_{r,k}\, \boldsymbol{h}_k^e$$

Sentence 1: [political bargaining chip]

Sentence 2: [cut financial aid; uganda]

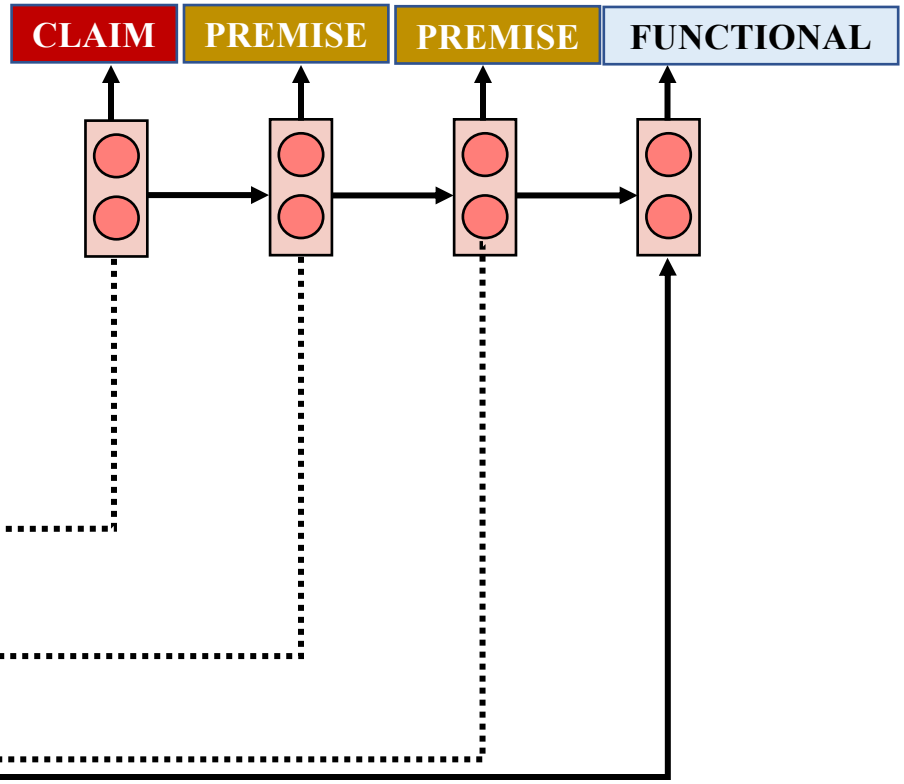
Keyphrase selection

$$P\left(\boldsymbol{v}_{j+1,k} = 1 | \boldsymbol{v}_{1:j}\right) = \sigma(\mathbf{w}_v^T \boldsymbol{s}_j + \boldsymbol{q}_j \mathbf{W}^c \boldsymbol{h}_k^e)$$

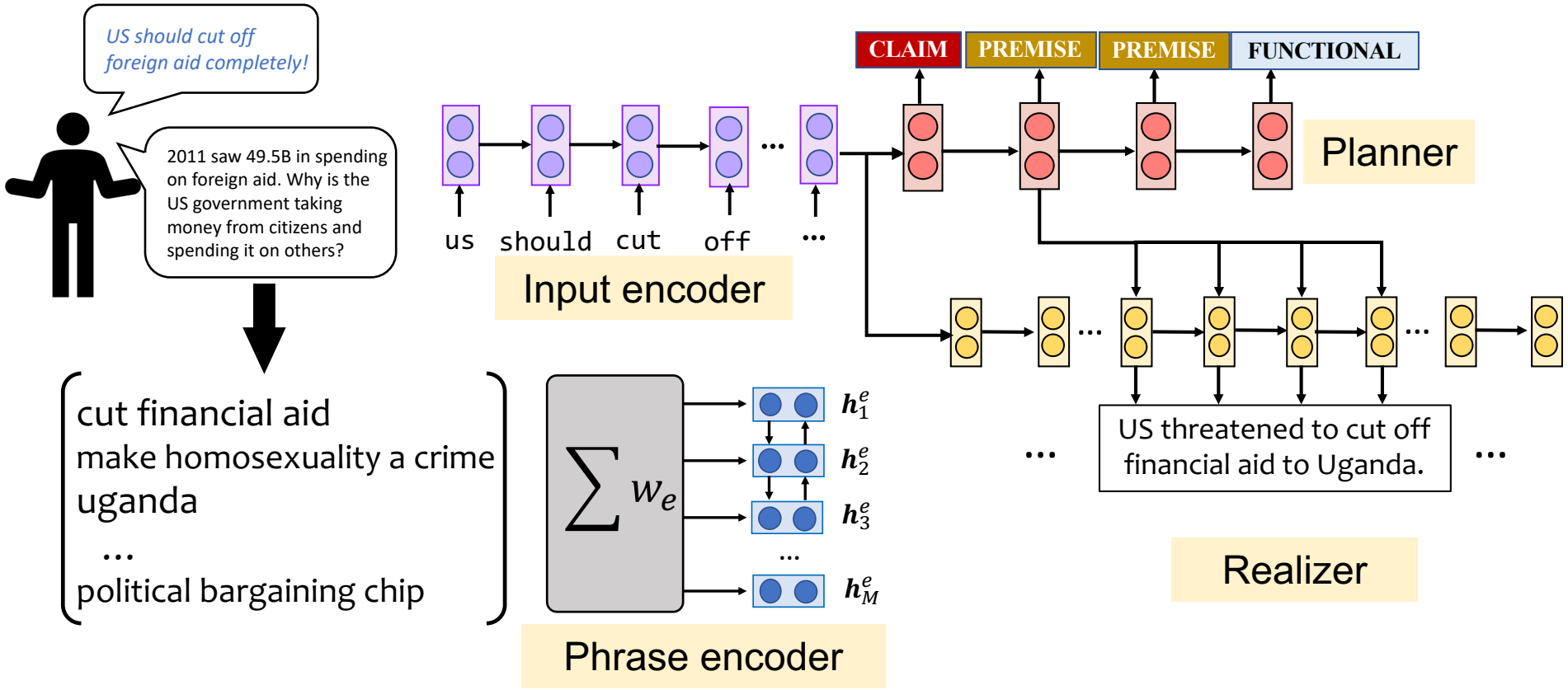
Selection history

Content Planning

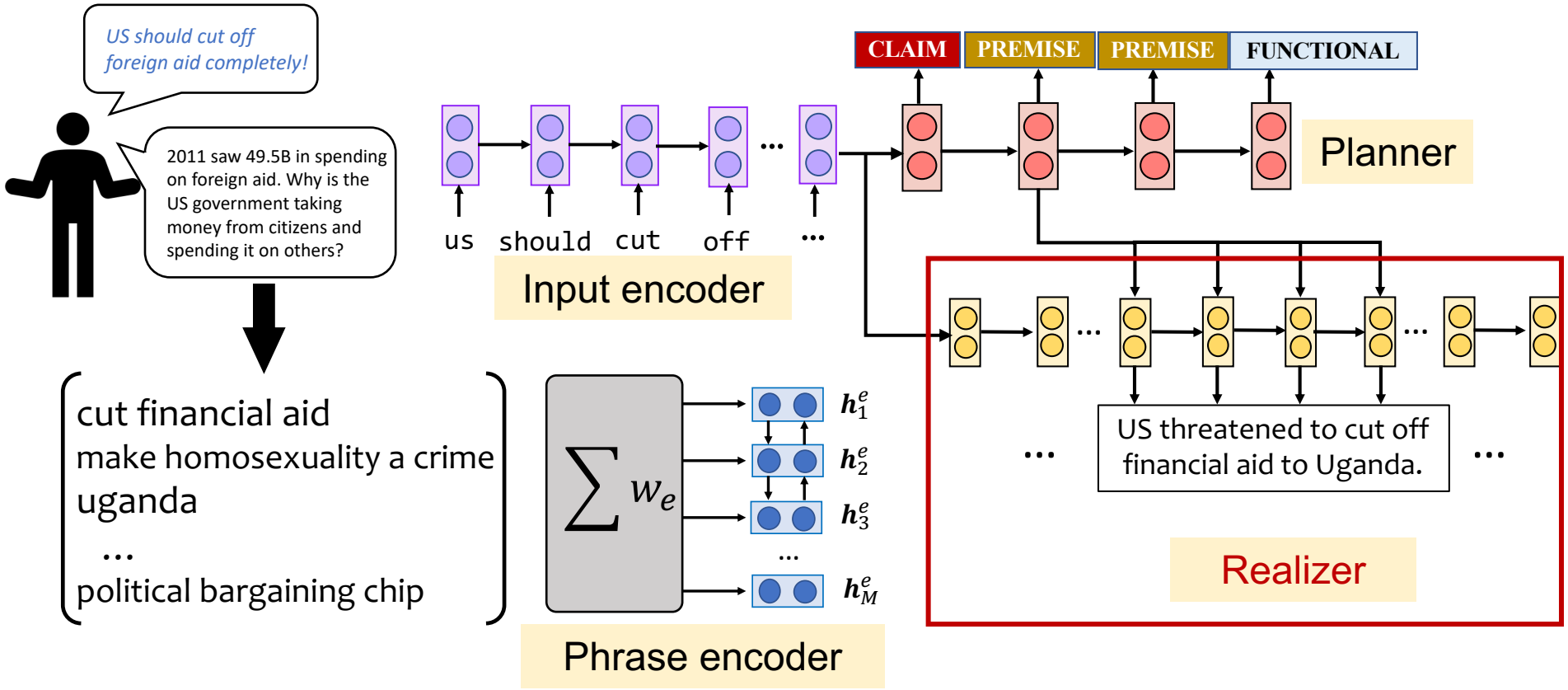
- Content selection decoding



Model


US should cut off foreign aid completely!
2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?
cut financial aid
make homosexuality a crime
uganda
...
political bargaining chip
us
should
cut
off
...
Input encoder
CLAIM
PREMISE
PREMISE
FUNCTIONAL
Planner
US threatened to cut off financial aid to Uganda.
Realizer
$\sum w_e$
h_1^e
h_2^e
h_3^e
h_M^e
Phrase encoder

Model


US should cut off foreign aid completely!
2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?
cut financial aid
make homosexuality a crime
uganda
...
political bargaining chip
us
should
cut
off
...
Input encoder
CLAIM
PREMISE
PREMISE
FUNCTIONAL
Planner
$\sum w_e$
h_1^e
h_2^e
h_3^e
...
h_M^e
Phrase encoder
US threatened to cut off financial aid to Uganda.
Realizer

Surface Realization

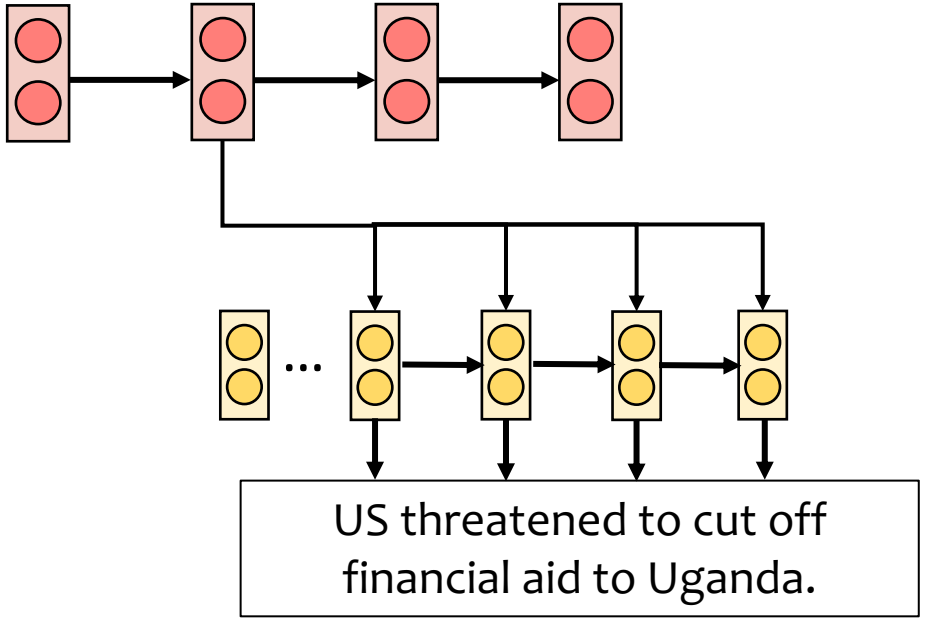
- Surface realization

Sentence 1: [political bargaining chip]

Sentence 2: [**cut financial aid; uganda**]

Sentence 3: [make homosexuality a crime]

Sentence 4: [NULL]



Surface Realization

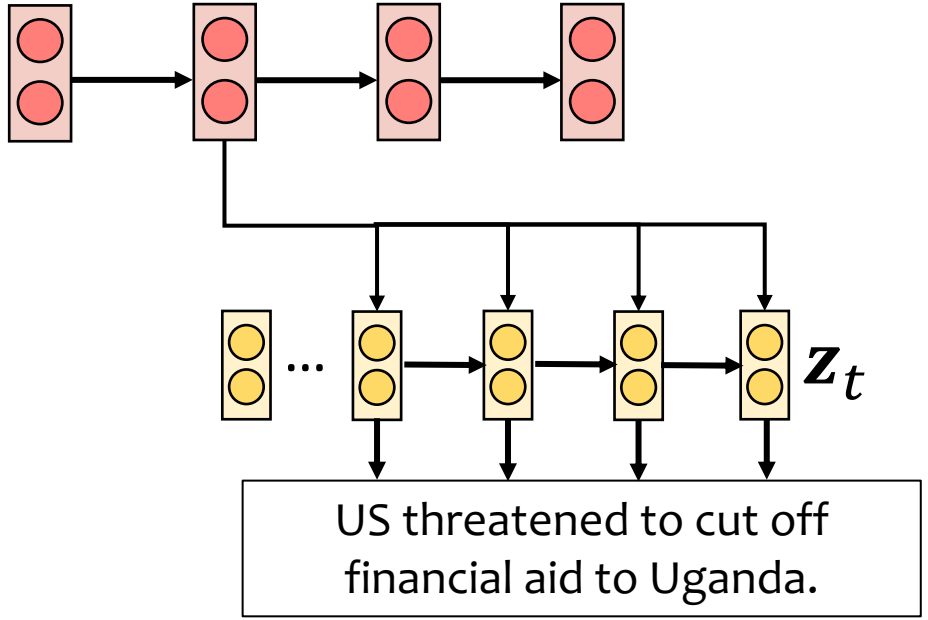
- Surface realization

Sentence 1: [political bargaining chip]

Sentence 2: [**cut financial aid; uganda**]

Sentence 3: [make homosexuality a crime]

Sentence 4: [NULL]



$$\boldsymbol{z}_t = LSTM(\boldsymbol{z}_{t-1}, \tanh(\mathbf{W}^{ws}\boldsymbol{s}_{J(t)} + \mathbf{W}^{ww}\boldsymbol{y}_{t-1}))$$

Surface Realization

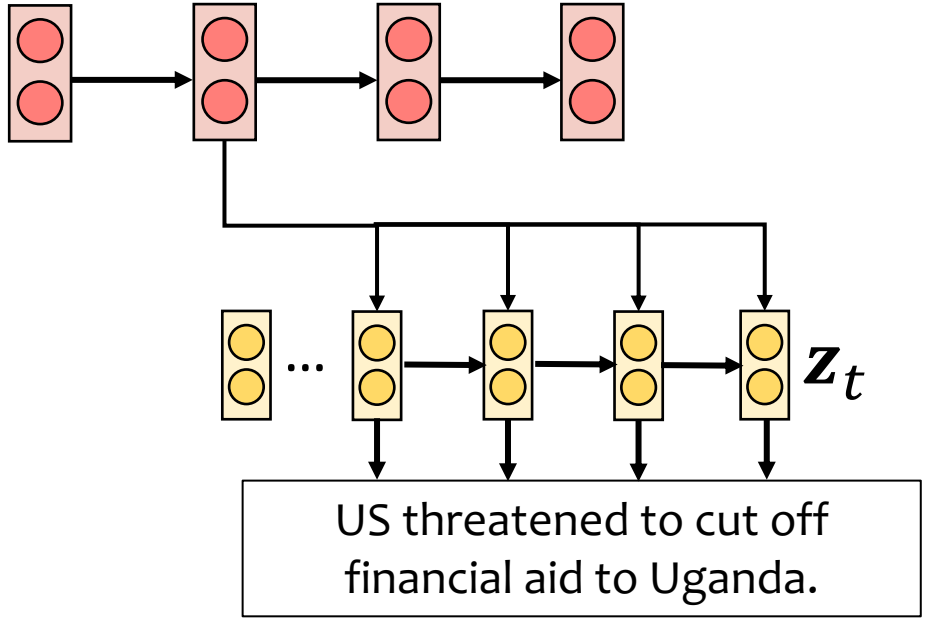
- Surface realization

Sentence 1: [political bargaining chip]

Sentence 2: [**cut financial aid; uganda**]

Sentence 3: [make homosexuality a crime]

Sentence 4: [NULL]



$$\boldsymbol{z}_t = LSTM(\boldsymbol{z}_{t-1}, \tanh(\mathbf{W}^{ws}\boldsymbol{s}_{J(t)} + \mathbf{W}^{ww}\boldsymbol{y}_{t-1}))$$

Content control

Surface Realization

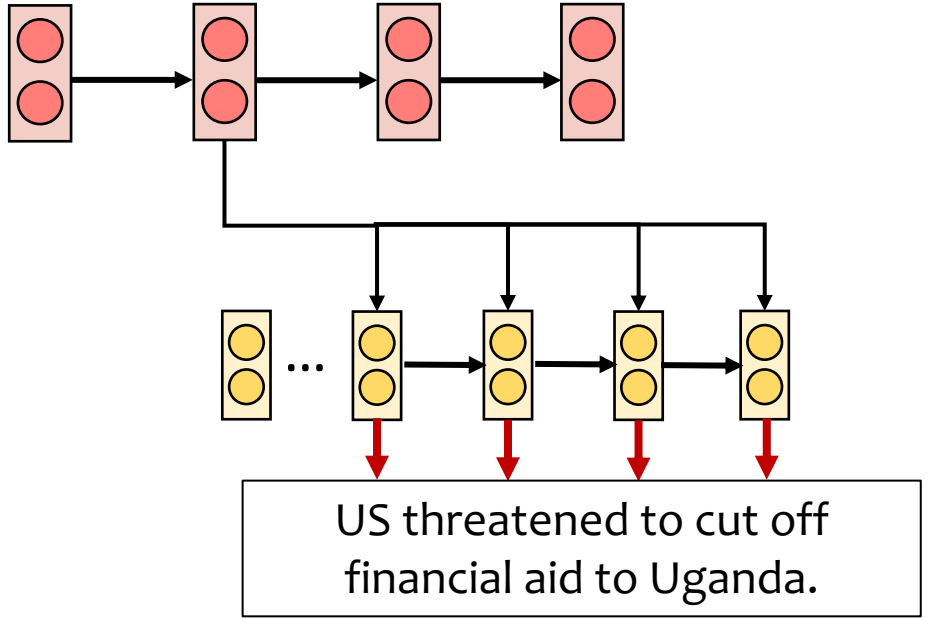
- Surface realization

Sentence 1: [political bargaining chip]

Sentence 2: [**cut financial aid; uganda**]

Sentence 3: [make homosexuality a crime]

Sentence 4: [NULL]



Output layer

$$P(y_t|y_{1:t-1}) = softmax(\tanh(\mathbf{W}^o[\boldsymbol{z}_t; \boldsymbol{c}_t; \boldsymbol{t}_{J(t)}]))$$

Style control

Model

- Training objective

$$L(\theta) = \sum -\log P(y^*|X;\theta) + \gamma L_{style}(\theta) + \eta L_{sel}(\theta)$$

Model

- Training objective

$$L(\theta) = \sum -\log P(y^*|X;\theta) + \gamma L_{style}(\theta) + \eta L_{sel}(\theta)$$

Token level cross-entropy

Style cross-entropy

Selection, binary cross-entropy

Roadmap

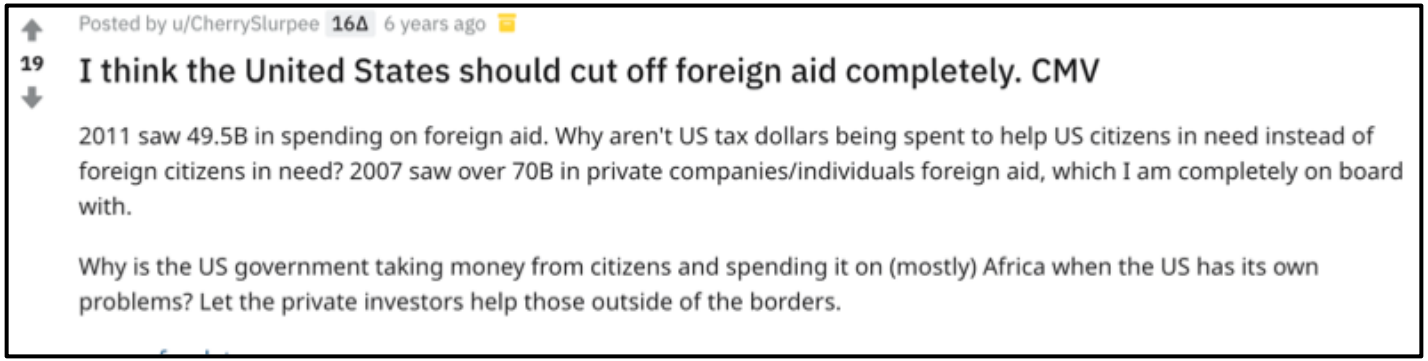
- Motivation
- Tasks
- Model
- Data
- Evaluation
- Conclusion

Data

- Argument generation
  - Data:  r/changemyview
  - Input statement: OP post
  - Output argument: “delta” awarded and well-upvoted replies

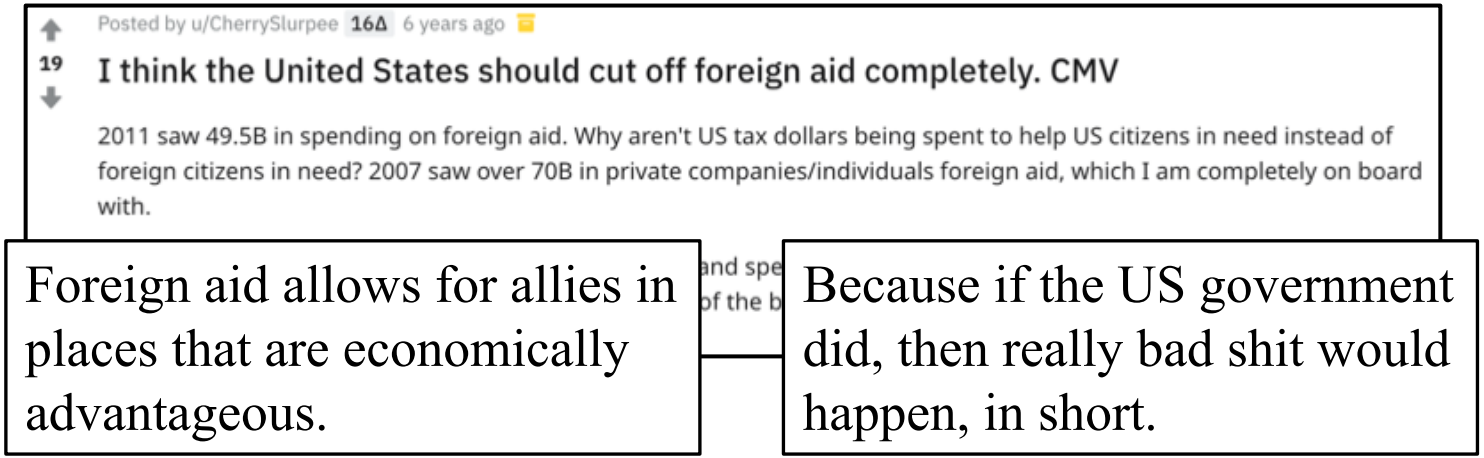
Data

- Argument generation
  - Data: r/changemyview
  - Input statement: OP post
  - Output argument: “delta” awarded and well-upvoted replies



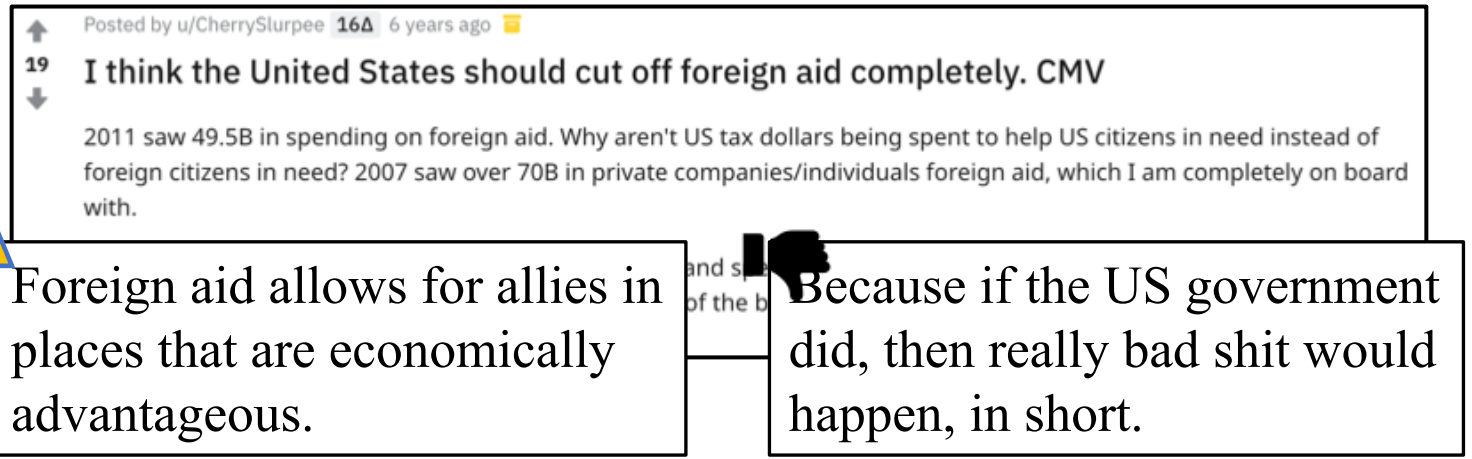
Data

- Argument generation
  - Data: r/changemyview
  - Input statement: OP post
  - Output argument: “delta” awarded and well-upvoted replies



Data

- Argument generation
  - Data: r/changemyview
  - Input statement: OP post
  - Output argument: “delta” awarded and well-upvoted replies



Data

- Argument generation
  - Data:  r/changemyview
  - Input statement: OP post
  - Output argument: “delta” awarded and well-upvoted replies
  - Keyphrases: noun/verb phrases from retrieved passages

Data

- Argument generation
  - Data: r/changemyview
  - Input statement: OP post
  - Output argument: “delta” awarded and well-upvoted replies
  - Keyphrases: noun/verb phrases from retrieved passages

Data

- Argument generation
  - Data: r/changemyview
  - Input statement: OP post
  - Output argument: “delta” awarded and well-upvoted replies
  - Keyphrases: noun/verb phrases from retrieved passages
    - Is at most 10 words
    - Contains at least one non-stopwords
    - Is a topic signature [Lin and Hovy, 2000] or a Wikipedia title

Data

- Argument generation
  - Data: r/changemyview
  - Input statement: OP post
  - Output argument: “delta” awarded and well-upvoted replies
  - Keyphrases: noun/verb phrases from retrieved passages
  - Sentence style: labeled by pattern matching and rules

Data

- Abstract generation for scientific papers
  - AGENDA dataset [Koncel-Kedziorski, Bekal, Luan, Lapata, and Hajishirzi, 2019]

Data

- Abstract generation for scientific papers

Title : Semantic Embeddings from Hashtags

Entities:

- short textual posts
- document recommendation task
- hastag prediction task
- …
- convolutional neural network

Abstract: We describe a convolutional neural network that learns feature representations for short textual posts using hashtags as a supervised signal. The proposed approach is …

Data

- Wikipedia paragraph generation
  - First paragraphs of Wikipedia articles

Data

- Wikipedia paragraph generation
  - First paragraphs of Wikipedia articles
  - Keyphrase as noun/verb chunks with content word(s)

Data

- Wikipedia paragraph generation
  - First paragraphs of Wikipedia articles
  - Keyphrase as noun/verb chunks with content word(s)

In computer science, **artificial intelligence** (**AI**), sometimes called **machine intelligence**, is intelligence demonstrated by machines, in contrast to the **natural intelligence** displayed by humans. Leading AI textbooks define the field as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals.[1] Colloquially, the term "artificial intelligence" is often used to describe machines (or computers) that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving".[2]

- computer science
- artificial intelligence
- machine intelligence
- …
- perceives its environment

Data

- Wikipedia paragraph generation
  - First paragraphs of Wikipedia articles
  - Keyphrase as noun/verb chunks with content word(s)
  - Sentence style as sentence complexities, proxied by length

Data

- Wikipedia paragraph generation
  - First paragraphs of Wikipedia articles
  - Keyphrase as noun/verb chunks with content word(s)
  - Sentence style as sentence complexities, proxied by length
  - Global style of simple vs. normal

In computer science, (…) any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals (…) that mimic "cognitive" functions that humans…

Artificial intelligence is the ability of a computer program or a machine to think and learn. (…) which tries to make computers "smart". (...) John McCarthy came up with the name (...)

Data

- Wikipedia paragraph generation
  - First paragraphs of Wikipedia articles
  - Keyphrase as noun/verb chunks with conte
  - Sentence style as sentence complexities, proxied by length
  - Global style of simple vs. normal

Model needs to capture the interplay between style and content.

In computer science, (…) any device that **perceives its environment** and takes actions that **maximize its chance of successfully achieving its goals** (…) that **mimic "cognitive" functions** that humans…

Artificial intelligence is the ability of a computer program or a machine **to think and learn**. (…) which tries to **make computers "smart".** (...) John McCarthy **came up with the name** (...)

Data

- Wikipedia paragraph generation
  - First paragraphs of Wikipedia articles
  - Keyphrase as noun/verb chunks with conte
  - Sentence style as sentence complexities, proxied by length
  - Global style of simple vs. normal

Model needs to capture the interplay between style and content.

We append one extra bit to the planner's input as global style control.

In computer science, (…) any device that **per**
and takes actions that **maximize its chance of successfully achieving its goals** (…) that **mimic "cognitive" functions** that humans…

Artificial intelligence is the ability of a computer program or a machine **to think and learn**. (…) which tries to **make computers "smart".** (...) John McCarthy **came up with the name** (...)

Roadmap

- Motivation
- Tasks
- Model
- Data
- Evaluation
- Conclusion

Experiment

- Comparisons for argument generation
  - RETRIEVAL: returns the highest ranked passage as output
  - SEQ2SEQ: encodes input prompt and keyphrase as tokens
  - Our 2018 ArgGen model (H&W 2018): generates keyphrases as auxiliary task

Experiment

- Comparisons for argument generation
- Comparisons for abstract generation
  - GRAPHWRITER [Koncel-Kedziorski et al, 2019]: state-of-the-art model with graph transformer that handles **both entities and relations**
  - SEQ2SEQ: encodes input prompt and keyphrase as tokens

Experiment

- Comparisons for argument generation
- Comparisons for abstract generation
- Comparisons for Wikipedia paragraph generation
  - RETRIEVAL: returns the most similar paragraph from training set
  - SEQ2SEQ: encodes input prompt and keyphrase as tokens
  - LOGREGSEL: predicts whether a phrase should be included, given global style

Automatic Evaluation

- Argument generation

	BLEU-2	ROUGE-L	METEOR	Length
RETRIEVAL	7.81	15.68	10.59	150.0
SEQ2SEQ	3.64	19.00	9.85	51.7
H&W (2018)	5.73	14.44	3.82	36.5

Automatic Evaluation

- Argument generation

	BLEU-2	ROUGE-L	METEOR	Length
RETRIEVAL	7.81	15.68	10.59	150.0
SEQ2SEQ	3.64	19.00	9.85	51.7
H&W (2018)	5.73	14.44	3.82	36.5
Ours	13.19	20.15	10.42	65.5

Automatic Evaluation

- Argument generation

	BLEU-2	ROUGE-L	METEOR	Length
RETRIEVAL	7.81	15.68	10.59	150.0
SEQ2SEQ	3.64	19.00	9.85	51.7
H&W (2018)	5.73	14.44	3.82	36.5
Ours	13.19	20.15	10.42	65.5
w/o Style	12.61	20.28	9.03	62.6

Automatic Evaluation

- Argument generation

BLEU, METEOR are improved with Style module.

	BLEU-2	ROUGE-L	METEOR	Length
RETRIEVAL	7.81	15.68	10.59	150.0
SEQ2SEQ	3.64	19.00	9.85	51.7
H&W (2018)	5.73	14.44	3.82	36.5
Ours	13.19	20.15	10.42	65.5
w/o Style	12.61	20.28	9.03	62.6

Automatic Evaluation

- Argument generation

	BLEU-2	ROUGE-L	METEOR	Length
RETRIEVAL	7.81	15.68	10.59	150.0
SEQ2SEQ	3.64	19.00	9.85	51.7
H&W (2018)	5.73	14.44	3.82	36.5
Ours	13.19	20.15	10.42	65.5
w/o Style	12.61	20.28	9.03	62.6
w/ Oracle Plan	**16.30**	**20.25**	**11.61**	65.5

Automatic Evaluation

- Argument generation

Oracle plan setup shows good planning leads to improved overall quality.

	BLEU-2	ROUGE-L	METEOR	Length
RETRIEVAL	7.81	15.68	10.59	150.0
SEQ2SEQ	3.64	19.00	9.85	51.7
H&W (2018)	5.73	14.44	3.82	36.5
Ours	13.19	20.15	10.42	65.5
w/o Style	12.61	20.28	9.03	62.6
w/ Oracle Plan	**16.30**	**20.25**	**11.61**	65.5

Automatic Evaluation

- Abstract generation

	BLEU-2	ROUGE-L	METEOR	Length
GRAPHWRITER	**29.95**	**28.56**	**19.90**	130.1
SEQ2SEQ	18.13	21.03	13.95	134.8
Ours	20.32	23.30	15.95	128.3
w/ Oracle Plan	25.03	26.18	19.21	125.8

Automatic Evaluation

- Abstract generation

With oracle plan, our system is competitive to the relation-aware SotA model.

	BLEU-2	ROUGE-L	METEOR	Length
GRAPHWRITER	**29.95**	**28.56**	**19.90**	130.1
SEQ2SEQ	18.13	21.03	13.95	134.8
Ours	20.32	23.30	15.95	128.3
w/ Oracle Plan	25.03	26.18	19.21	125.8

Automatic Evaluation

- Wikipedia paragraph generation (Normal)

	BLEU-2	ROUGE-L	METEOR	Length
RETRIEVAL	20.10	28.60	12.23	44.5
SEQ2SEQ	22.62	27.49	14.74	52.9
LOGREGSEL	29.28	28.65	**27.76**	34.4
Ours	**33.76**	**40.08**	25.70	65.4

Automatic Evaluation

- Wikipedia paragraph generation (Norma

Our model yields much better BLEU, ROUGE than comparisons.

	BLEU-2	ROUGE-L	METEOR	Length
RETRIEVAL	20.10	28.60	12.23	44.5
SEQ2SEQ	22.62	27.49	14.74	52.9
LOGREGSEL	29.28	28.65	**27.76**	34.4
Ours	**33.76**	**40.08**	25.70	65.4

Automatic Evaluation

- Wikipedia paragraph generation (Normal)

	BLEU-2	ROUGE-L	METEOR	Length
RETRIEVAL	20.10	28.60	12.23	44.5
SEQ2SEQ	22.62	27.49	14.74	52.9
LOGREGSEL	29.28	28.65	27.76	34.4
Ours	33.76	40.08	25.70	65.4
w/ Oracle Plan	**37.70**	**45.41**	**31.65**	79.8

Automatic Evaluation

- Wikipedia paragraph generation (Simple)

	BLEU-2	ROUGE-L	METEOR	Length
RETRIEVAL	21.99	33.44	12.97	34.7
SEQ2SEQ	21.98	29.36	16.94	52.8
LOGREGSEL	5.59	23.21	13.27	13.0
Ours	31.22	40.76	26.76	58.7
w/ Oracle Plan	**34.22**	**45.48**	**32.84**	70.5

Human Evaluation

- Human assessment
  - Grammaticality (1-5): fluency, free of grammar errors
  - Correctness (1-5): non-contradictory, right stance (argument)
  - Content richness (1-5): coverage of relevant points

Human Evaluation

- Human assessment
  - Grammaticality (1-5): fluency, free of grammar errors
  - Correctness (1-5): non-contradictory, right stance (argument)
  - Content richness (1-5): coverage of relevant points

Argument generation (30 samples)

	Gram.	**Corr.**	**Cont.**
HUMAN	4.81	3.90	3.48
Ours	3.99	2.78	2.61
w/o Style	3.03	2.26	2.03

Human Evaluation

- Human assessment
  - Grammaticality (1-5): fluency, free of grammar errors
  - Correctness (1-5): non-contradictory, right stance (argument)
  - Content richness (1-5): coverage of relevant points

Wikipedia paragraph generation (30 samples)

	Gram.	**Corr.**	**Cont.**
HUMAN	4.84	4.73	4.49
Ours	3.38	3.24	3.43
w/o Style	2.99	2.89	3.50

Human Evaluation

- Human assessment
  - Grammaticality (1-5): fluency, free of grammar errors
  - Correctness (1-5): non-contradictory, right
  - Content richness (1-5): coverage of releva

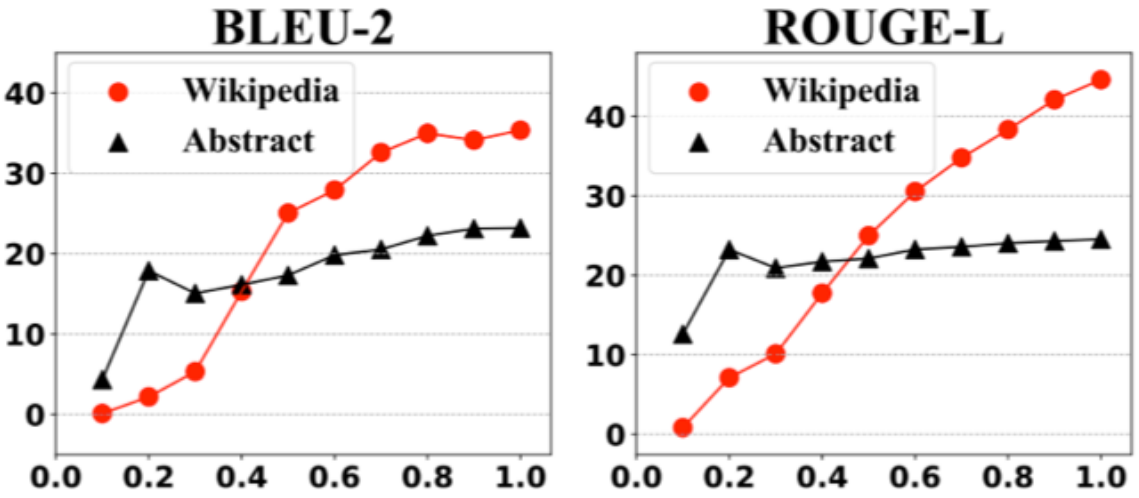
Our system is still far from human, yet style module helps significantly.

Wikipedia paragraph generation (30 samples)

	Gram.	Corr.	Cont.
HUMAN	4.84	4.73	4.49
Ours	3.38	3.24	3.43
w/o Style	2.99	2.89	3.50

Analysis

- Effect of keyphrase selection



F1 on Keyphrase Selection

Sample Argument Output

PROMPT: Aborting a fetus has some non-zero negative moral implications.

SEQ2SEQ: I'm not going to try to change your view here, but I don't want to change your position. I don't think it's fair to say that a fetus is not a person. It's not a matter of consciousness.

OURS: TL;DR: I agree with you, but I think it's important to note that fetuses are not fully developed. I don't know if this is the case, but it doesn't seem to be a compelling argument to me at all, so I'm not going to try to change your view by saying that it should be illegal to kill.

OURS w/o Style: Now, I don't think it's fair to say that it's okay to kill someone. If you don't want to die, you shouldn't be able to do that for the sake of the mother and the fetus.

Sample Wikipedia Output

TITLE: Breaking Bad

STYLE: Simple Wikipedia

SEQ2SEQ: Breaking is an American television soap opera that aired on CBS from September 26, 2007 to April 19, 2009. The series is set in main, New South Wales, in the United Kingdom. It is a spin-off of the ”breaking” series, and was produced by colloquialism productions and broadcast from the television series “ “.

OURS: Breaking bad is an American television series set in Albuquerque, New Mexico. It was started by January 2008 and ended on September 2013 after being cancelled shortly thereafter.

OURS w/o Style: Bad breaking is an American television series set in Albuquerque, New Mexico on cable channel from 2007 to 2013.

Sample Wikipedia Output

TITLE: Breaking Bad

STYLE: Simple Wikipedia

SEQ2SEQ: Breaking is an American television soap opera that aired on CBS from September 26, 2007 to April 19, 2009. The series is set in main, New South Wales, in the United Kingdom. It is a spin-off of the ”breaking” series, and was produced by colloquialism productions and broadcast from the television series “ “.

OURS: Breaking bad is an American television series set in Albuquerque, New Mexico. It was started by January 2008 and ended on September 2013 after being cancelled shortly thereafter.

OURS w/o Style: Bad breaking is an American television series set in Albuquerque, New Mexico on cable channel from 2007 to 2013.

Sample Wikipedia Output

TITLE: Breaking Bad

STYLE: Simple Wikipedia

SEQ2SEQ: Breaking is an American television soap opera that aired on CBS from September 26, 2007 to April 19, 2009. The series is set in main, New South Wales, in the United Kingdom. It is a spin-off of the ”breaking” series, and was produced by colloquialism productions and broadcast from the television series “ “.

SEQ2SEQ generates more factual errors.

OURS: Breaking bad is an American television series set in Albuquerque, New Mexico. It was started by January 2008 and ended on September 2013 after being cancelled shortly thereafter.

OURS w/o Style: Bad breaking is an American television series set in Albuquerque, New Mexico on cable channel from 2007 to 2013.

Conclusion

- We propose a generic text generation framework with sentence-level planning and realization.
- We utilize keyphrases as units for content selection, offering flexibility in different domains and tasks.
- Experiments on three tasks show the effectiveness of the model and the association between selection and final output quality.

Thanks!

hua.x@husky.neu.edu

https://github.com/XinyuHua/textgen-emnlp19

Sample Wikipedia Output

TITLE: Breaking Bad

STYLE: Normal Wikipedia

SEQ2SEQ: Bad breaking is an American crime drama television series created by David main. It ran from September 21, 2001 to January 19, 2009 on Fox. The series is set in New York City and follows the adventures of a group of in the fictional town of breaking series, after the events of the breaking series.

OURS: Bad breaking is an American neo-western crime drama television series an American neo-western crime drama television series drama series television written and produced by vince vince. The show aired on AMC for five seasons from January 20, 2011 to September 29, 2013.

OURS w/o Style: Bad breaking is an American an American neo-western crime drama television series that was produced by Vince Gilligan. The show ran for over five seasons from January 20 to September 29, 2011.

Sample Wikipedia Output

TITLE: Moon Jae-in

STYLE: Simple Wikipedia

SEQ2SEQ: Moon selection Park is a election politician who served as prime minister of Korea from 2007 to 2013. He was elected as a member of the house of democratic party in the Moon's the the Moon the first serving president of jae-in, in office since 2010.

OURS: Moon Jae-in is a South Korean politician and current president of South Korea from 2012 to 2017 and again from 2014 to 2017.

OURS w/o Style: Moon Moon Jae-in is a South Korean politician and current president of Singapore since 2016.

Sample Argument Output

PROMPT: All alleged crimes or infractions involving a police officer should be investigated by a higher-level agency.

SEQ2SEQ: If you look at the history of history, there was a time when the founding fathers didn't understand the history of the world. They didn't want to go to war. It wasn't be cause they didn't exist.

OURS: Jail sentences for example are a result of negligence. If you don't like it, you're going to have to deal with a lot of bad lawyers, and you'll have to prove your're guilty of a crime in the first place and you're not going to be able to sue for damages if you're in jail.

OURS w/o Style: The biggest problem with this is that this is a huge problem. The problem is that there are a lot of cases where the defendant is innocent, and the defendant has the power to prove that the defendant didn't want to go to trial, (…)

Backup Slides: Experiment Details

- Argument generation
  - Optionally append retrieved passages to input prompt.
  - Truncate input to 500 (OP), 400 (passages), and 70 (# keyphrases).
  - Pre-train encoder and realization decoder with more OP and replies.
- Wikipedia paragraph generation
  - Use sum of word embedding for title encoding.
  - Truncate keyphrase to at most 30.
- Abstract generation
  - Truncate keyphrase to at most 30.

Backup Slides: Experiment Details

- Overall
  - Two layers LSTM are used for both decoders.
  - Each layer has 512 dimensions, with dropout probabilities as 0.2.
  - Learning is driven by AdaGrad [Duchi et al., 2011] with 0.15 as learning rate and 0.1 as initial accumulartor.
  - Gradient clipped to 2.0.
  - Batch size is 64.
  - Beam search decoding of size 5, with trigram repetitions blocked.

Backup Slides: Data

- Overall statistics

	Argument generation	Abstract generation	Wikipedia paragraph generation (Nor/Sim)
# Tokens	54.87	141.34	70.57/48.60
# Sentences	2.48	5.59	3.15/3.20
# Keyphrase (candidates)	55.80	12.23	23.56
# Keyphrase (selected)	11.61	12.23	16.01/11.11
# Training data	272,147	38,720	125,136