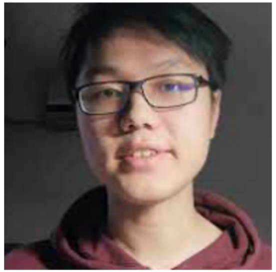
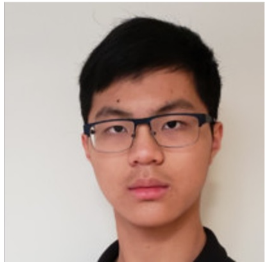
InFillmore: Frame-Guided Language Generation with Bidirectional Context




Jiefu Ou   Nathaniel Weir   Anton Belyy   Felix Yu   Ben Van Durme

JOHNS HOPKINS UNIVERSITY

Motivation

- State-of-the-art **text infilling models**: high-quality human-like text generation, core components of **automatic and interactive story generation**.

- Lack of control mechanism: explicitly **control for underlying semantic content.**

- Propose a way for humans (or content planning models) to **specify discrete semantic content** while conditioning on **surrounding textual context.**

FrameNet

Research

One of 1221 frames

Definition:

A Researcher (an individual a group or institution) attempts to answer a Question by means of consulting literature, observation, or conducting experiments in a particular Field pertinent to the Question. The Question may be underspecified in the form of a Topic.

Simmons and Nadel did RESEARCH pertaining to Ardea purpurea and Egretta garzetta in 1998

(Core) Frame Elements:

Question [que]	A significant unanswered open proposition which the Researcher is attempting to resolve. When we have RESEARCHED why Brazilian mothers have stopped breastfeeding, principal reasons given are 'poor milk' and 'insufficient milk' .
Researcher [res] **Semantic Type:** Sentient	The people engaged in a research endeavor. Before he became an astronaut, Grunsfeld *did* RESEARCH in physics.
Topic [top]	A phenomenon or idea involved in a Question. Would you take part in RESEARCH on the best place to have a baby?

Lexical Units:

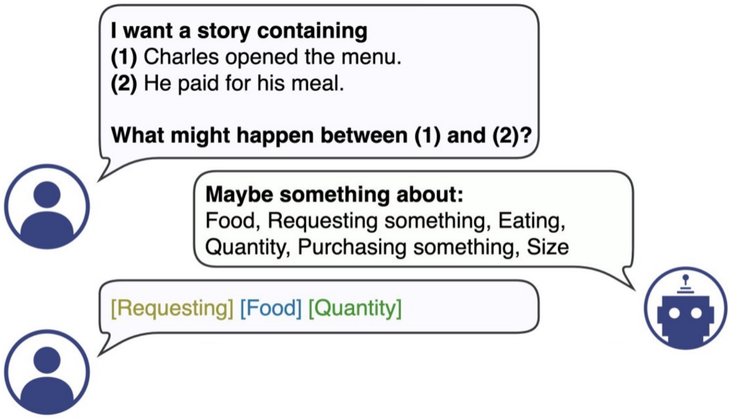
investigate.v, research.n, research.v

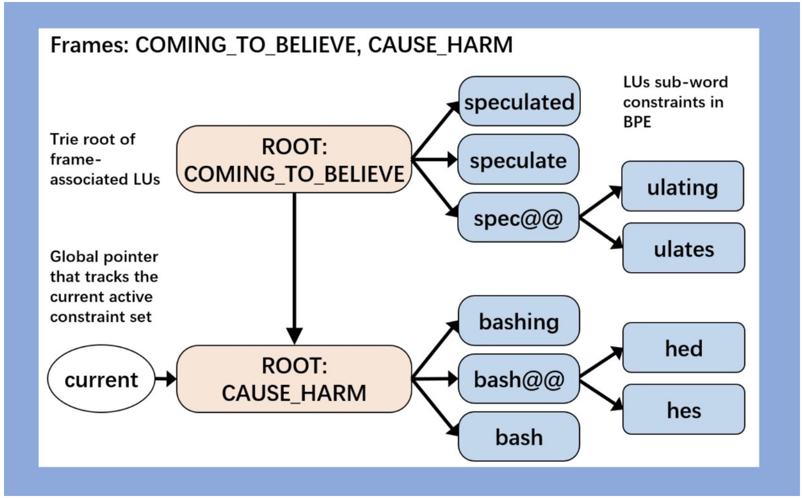
Created by JKR on 06/29/2004 10:55:02 PDT Tue

Lexical Unit	LU Status	Lexical Entry Report
investigate.v	**Finished_Initial**	Lexical entry
research.n	**Finished_Initial**	Lexical entry
research.v	**Finished_Initial**	Lexical entry

Baker, Collins et al. The Berkeley FrameNet Project. ACL 1998

Frame-Guided Generation





Story	Charles went shopping. He bought fruit. Then he left.
ILM	Charles went shopping. [blank] Then he left. *[sep] He bought fruit.*
S-FFL	[sep] [Food] He bought fruit.
A-FFL	[sep] [Commerce_buy] [Food] He bought fruit.

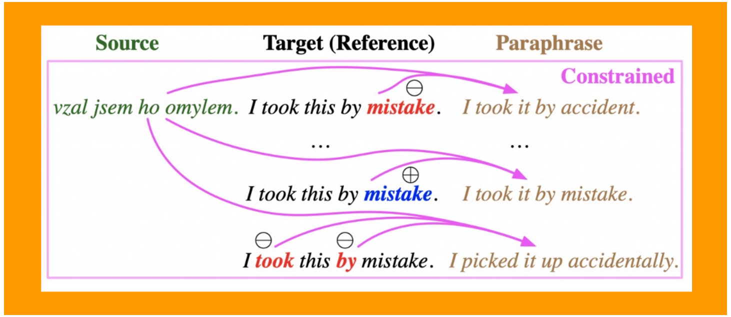
We introduce two approaches to FGG:

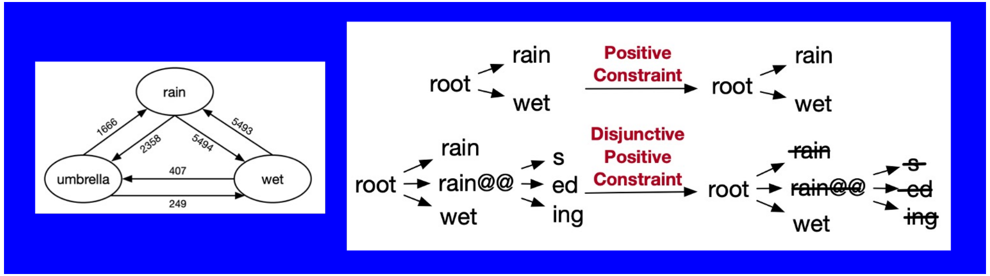
1. Lexically constraining decoding (LCD) to include a frame's lexical units (LUs)
[Bringing] -> {carry, bring, convey, haul,...}

2. Fine-tuning a "framefilling" (FFL) model to generate from control codes that specify frame content

Lexically Constrained Decoding (LCD)

- A modification to beam search that forces the inclusion of pre-specified words and phrases in the output.
- Applications: Machine Translation, **Paraphrasing**, **Causal Generation**





Post, Matt, and David Vilar. "Fast Lexically Constrained Decoding with Dynamic Beam Allocation for Neural Machine Translation." *NAACL 2018*

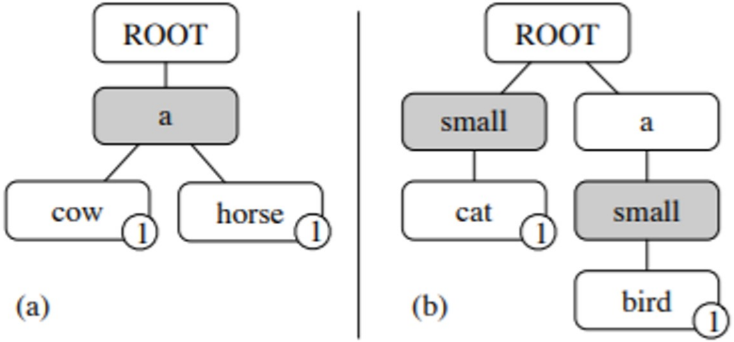
Hu, J. Edward, et al. "ParaBank: Monolingual bitext generation and sentential paraphrasing via lexically-constrained neural machine translation." AAAI 2019.

Li, Zhongyang, et al. "Guided Generation of Cause and Effect." *IJCAI* 2020.

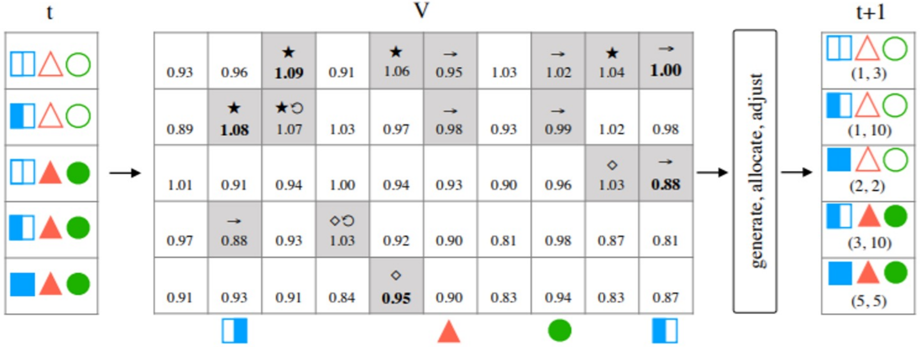
Lexically Constrained Decoding (LCD)

- A modification to beam search that forces the inclusion of pre-specified words and phrases in the output.
- Applications: Machine Translation, Paraphrasing, Casual Generation

Constraints Progress Bookkeeping



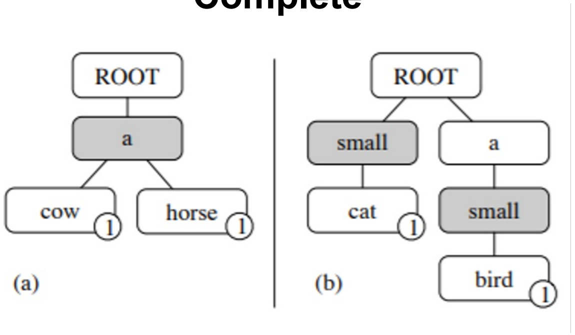
(Dynamic) Beam Allocation



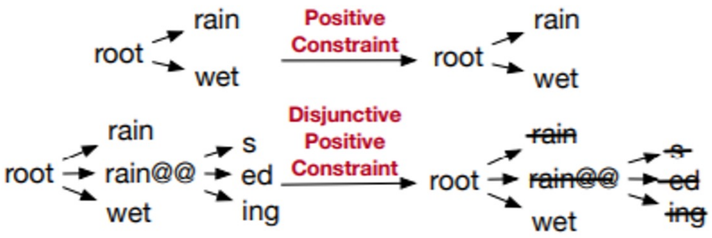
Lexically Constrained Decoding (Cont'd)

- **Extending the type of constraints**

Complete



Single Disjunctive Set

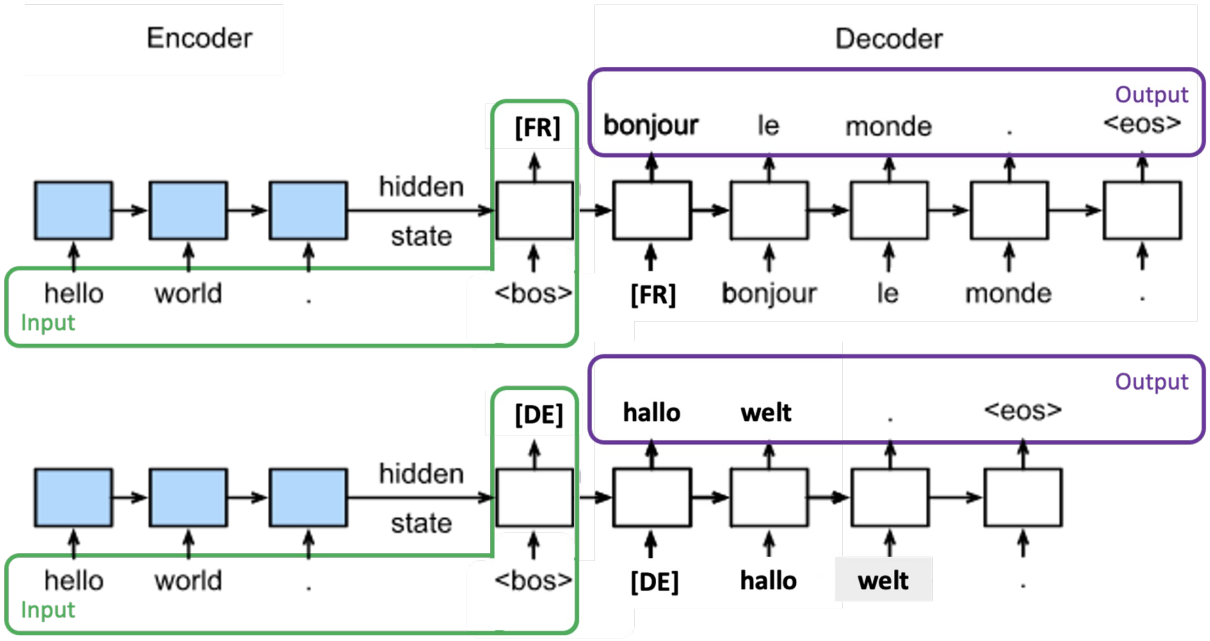


Multiple Disjunctive Sets (Our)

Frames: COMING_TO_BELIEVE, CAUSE_HARM

Fine-tuned “Framefilling” with Control Codes

A **control code** is a special vocabulary token used as a **soft specification content** to a LM



Used previously for

- Multilingual machine translation [1]
- Arbitrary style + content features [2]
- Semantically diverse NLG [3]
- …and many more

[1] Ha, Thanh-Le, Jan Niehues, and Alex Waibel. "Toward Multilingual Neural Machine Translation with Universal Encoder and Decoder." IWSLT 2016
[2] Keskar, Nitish Shirish, et al. "Ctrl: A conditional transformer language model for controllable generation." 2019.
[3] Weir, Nathaniel, João Sedoc, and Benjamin Van Durme. "COD3S: Diverse Generation with Discrete Semantic Signatures." *EMNLP* 2020.

Fine-tuned “Framefilling” with Control Codes

We add frame IDs as control codes to an “Infilling Language Model” fine-tuned for bidirectional-context generation

Original Story:
(1) Charles opened the menu.
(2) *He ordered a few dishes.*
(3) *The waiter brought him the food.*
(4) He paid for his meal.

Splice and reorder for infilling

Infilling Language Model (ILM; Donahue et al. 2020):
Charles opened the menu. **[blank]. [blank].** He paid for his meal. **[sep]** He ordered a few dishes. **[answer]** the water brought him the food. **[answer]**

Add frame control codes

“Framefilling” Language Model (FFL; ours):
Charles opened the menu. **[blank]. [blank].** He paid for his meal. **[sep]**
[Requesting] [Food] [Quantity] He ordered a few dishes. **[answer]**
[Bringing] [Profession] the water brought him the food. **[answer]**

Evaluation Task: Story Cloze with ROCStories

Story
Ari spends $20 a day on pickles.
He decides to make his own to save money.
He puts the pickles in brine.
[blank]
Ari opens the jar to find perfect pickles.

Gold
Ari waits 2 weeks for his pickles to get sour.

ILM Baseline
He puts the pickles in a jar.

FFL	ILM+LCD
Single Frame: [Transition_to_State]	
He ends up with a jar full of pickles.	He gets the pickles and puts them in jars.
Multiple Frames: [Cardinal_Numbers] [Transition_to_State]	
He ends up with 5 jars of pickles.	He puts one in the jar and opens it to get a drink.
All Frames: [Cardinal_Numbers] [Measure_duration] [Transition_to_State][Chemical-sense_description]	
He waits for a week for the pickles to get sour.	He waits for the pickles to thaw out of the jar to thaw one day he gets the pickles and eats them delicious.

Mostafazadeh, Nasrin et al. A Corpus and Evaluation Framework for Deeper Understanding of Commonsense Stories. NAACL 2016

Automatic Evaluation

Fidelity ↑	Recall			Perfect Recall	
# Frames	Single	Multi	All	Multi	All
ILM (no guidance)	.169	.166	.165	.091	.026
ILM + LCD	**.584**	.595	.610	.418	.232
ILM + LCD-ord	–	**.598**	.626	**.427**	.255
FFL	.518	.559	.640	.381	.259
FFL (rand sample)	.461	.511	.601	.338	.224
FFL-ord	–	.585	**.669**	.415	**.298**

FFL and **LCD** are substantially better at capturing the underlying gold semantic content compared to the unconstrained **ILM**

Human Evaluation

Confusion rate (%) ↑	# Frames		
	Single	Multi	All
ILM (no guidance)	**41**	**41**	**41**
ILM+LCD	35	31*	20*
FFL	33*	39	38
FFL-ordered	33*	38	37

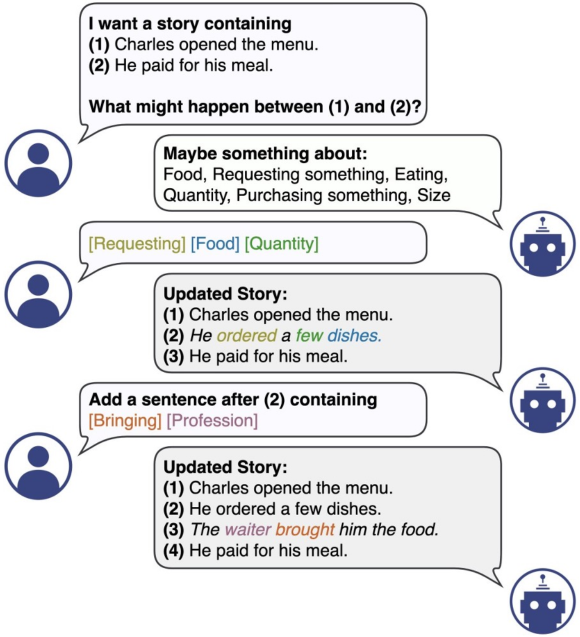
Average rank (1..10) ↓	# Frames		
	Single	Multi	All
ILM (no guidance)	**5.48**	**5.48**	5.48
ILM+LCD	5.85*	6.38*	7.50*
FFL	5.88*	5.57	5.11
FFL-ordered	5.88*	5.53	**5.02***

- Most FFL variants are competitive with ILM on generation quality

Conclusions

- We introduce the frame-guided generation framework that uses FrameNet frames to guide semantic content in an infilling model.
- We propose two extensions to an infilling model to efficiently utilize frame-semantic information during text generation.
- Experiments on the story cloze task show that both our extensions enable high semantic frame fidelity with competitive generation quality.

Thank you!



https://nlp.jhu.edu/demos/infillmore