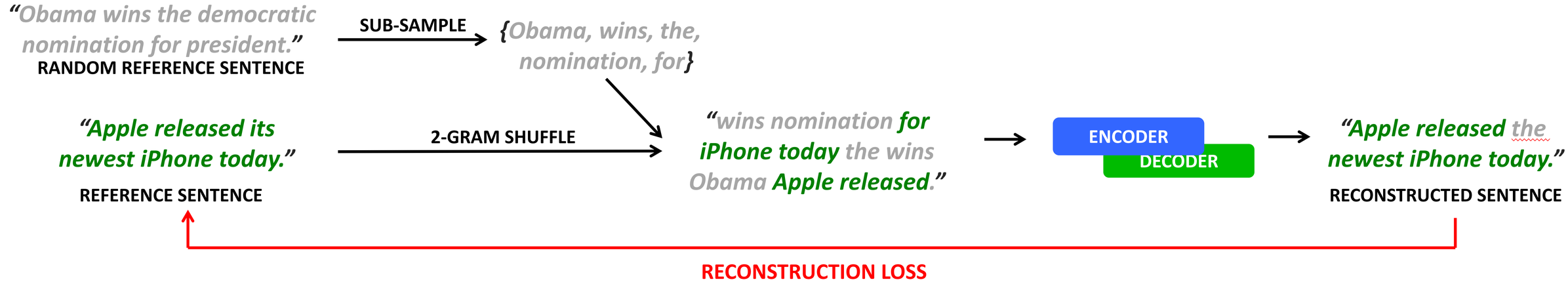
Unsupervised Sentence Compression using Denoising Auto-Encoders

Thibault Févry and Jason Phang

- **Motivation:** Sentence compression and text summarization models often require strong linguistic heuristics or paired corpora to train
- **Strategy**: Synthesize longer "sentences" by sampling words from another sentence and shuffling
  → Train a denoising auto-encoder to recover the original sentence
- **Performance**: In human evaluation, our model is competitive with supervised baselines in grammatical correctness and retention of meaning

Training Scheme



Length Countdown

- Decoder is fed numerical countdown to desired output length at each time-step
- Target half the length of input sentence in inference

$$h_t^{dec} = RNN(h_{t-1}^{dec}, x_t, T_{dec} - t)$$

Could semantic conditioning help?

- Conditioning decoder on sentence semantics by incorporating pretrained InferSent embeddings of input sentence into initial hidden state

$$h_0^{dec} = f(h_{T_{enc}}^{enc}, s)$$

Training / Evaluation method

- Trained on Annotated Gigaword using only reference sentences
- At evaluation: target half the length of input sentence

Human Evaluation

Models	Grammar	Meaning
2-g shuf	**3.53** (±0.18)	2.53 (±0.16)
2-g shuf + InferSent	2.87 (±0.16)	2.13 (±0.13)
Seq2seq	3.43 (±0.18)	**2.60** (±0.17)
Ground Truth	4.07 (±0.13)	3.87 (±0.16)

ROUGE Evaluation

	R-1	R-2	R-L	*Len*
Baselines				
All Text	28.91	10.22	25.08	31.3
First-8 words	26.90	9.65	25.19	8
Unsupervised (Ours)				
2-g shuf	27.72	7.55	23.43	15.4
2-g shuf + InferSent	28.42	7.82	24.95	15.6
Supervised				
Seq2seq	**35.50**	15.54	32.45	15.4
Nallapati et al. (2016)	34.97	**17.17**	**32.70**	-

Example #1

Input: Nearly ### of the released hostages remain in hospital, and more than ### of them are in very serious condition, Russian medical authorities said Sunday.
2-g shuf: More than ### hostages are in serious condition, Russian medical authorities said.
2-g shuf + InferSent: Nearly ### hostages of the nearly released in serious medical condition, said.
Ground-Truth: Nearly ### people still hospitalized more than ### in critical condition

Summaries are highly extractive (by construction), but still grammatically correct

Example #2

Input: French president Jacques Chirac arrived here Friday at the start of a <unk> during which he is expected to hold talks with Romanian leaders on Bucharest's application to join NATO.
2-g shuf: French president Jacques Chirac arrived here Friday to hold talks with Romanian leaders on NATO.
2-g shuf + InferSent: French president Jacques Chirac arrived here Friday at the start of talks to join NATO.
Ground-Truth: Chirac arrives in Romania

Long strings of names and entities are recovered by the model, likely due to the dataset (news headlines)

Varying Desired Length

Input: Three convicted serial killers have been hanged in Tehran's Evin prison, the Khorasan newspaper reported Sunday.
L=11: Three convicted serial killers have been hanged in prison Sunday.
L=13: Three convicted serial killers have been hanged, a newspaper reported Sunday.
L=15: Three convicted serial killers have been hanged in Tehran, a newspaper reported Sunday.
L=17: Three convicted serial killers have been hanged in Tehran, the Tehran's newspaper reported Sunday.

We can control the length of our summaries thanks to length-countdown

Takeaways

- Seq2seq denoising auto-encoders are surprisingly capable of recovering sentences from shuffled sets of words
- Naïve conditioning of sentence generation on InferSent improves ROUGE score but hurts human evaluation
- Need more inductive biases for unsupervised summarization!