

Neural Text Production: Encoding, Scaling and Generalising

Claire Gardent

CNRS/LORIA, Nancy, France
MSR Workshop, Hong Kong
3 October 2019



Joint work with

[Emilie Colin](#) and [Anastasia Shimorina](#)

PhD Student, Lorraine University, Nancy

[Angela Fan](#)

PhD Student, Lorraine University, Nancy and Facebook AI Research, Paris

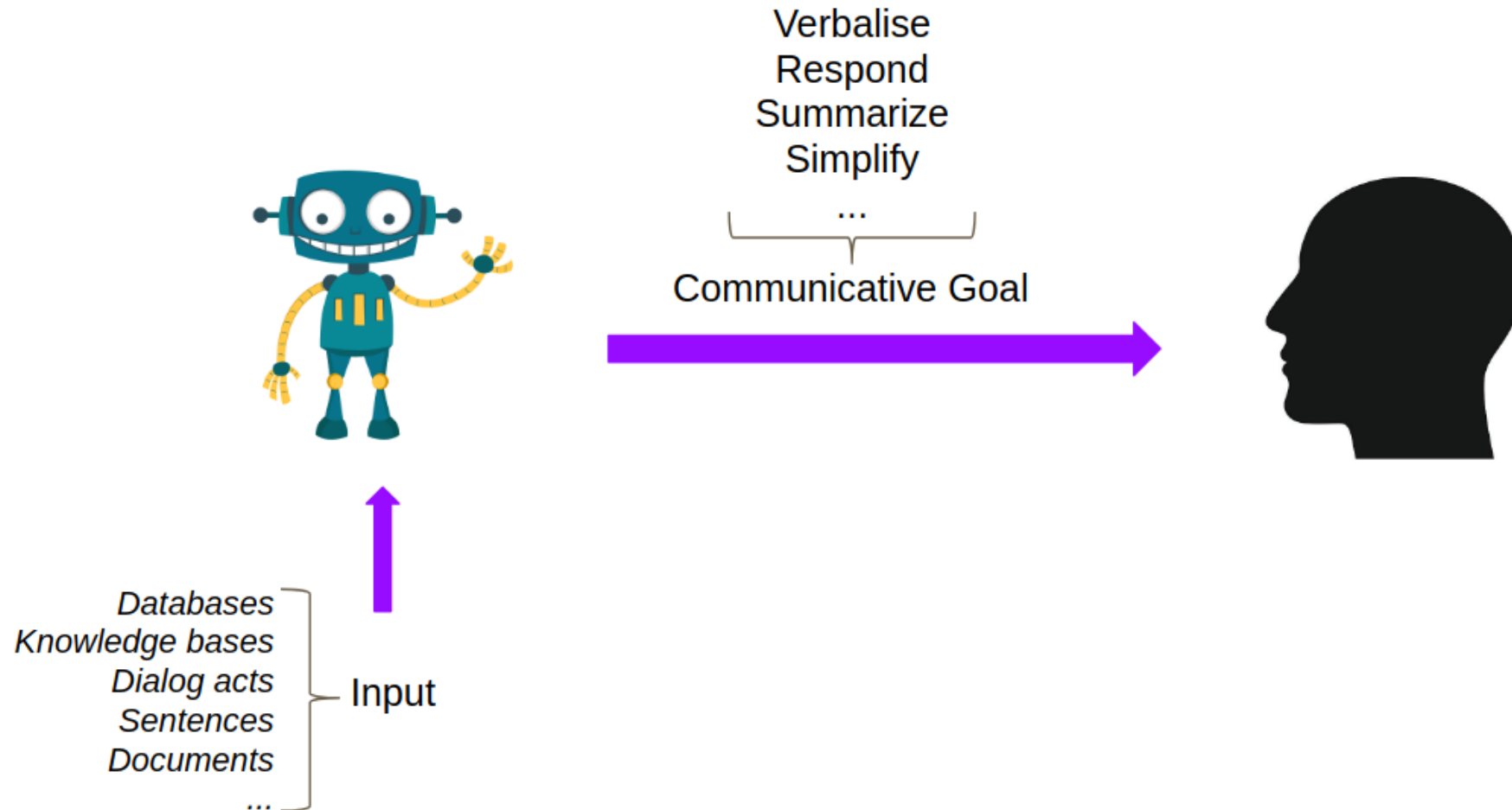
[Iryna Gurevych](#), [Yevgenyi Puzekov](#) and [Leonardo Ribeiro](#)

AIPHES and UKP Lab, Technische Universitat Darmstadt, Germany

[Ido Dagan](#)

Bar-Ilan University, Ramat-Gan, Israel

NLG: Many Inputs, Many Goals



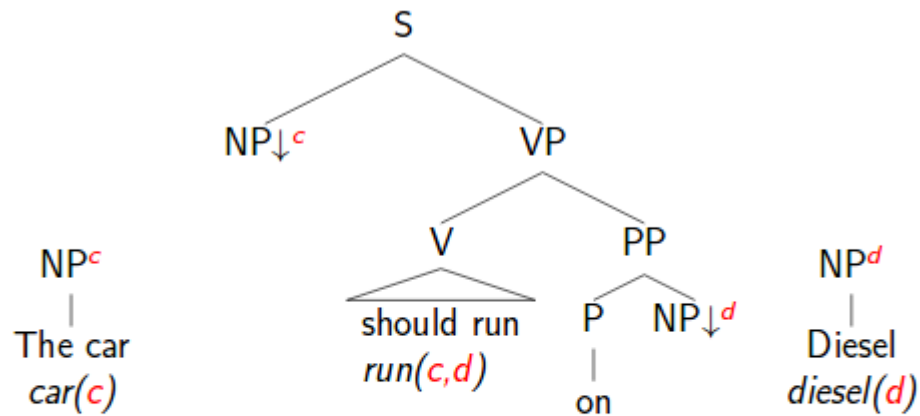
Generating from Data

The NLG Pipeline



Generating from Meaning Representations

Grammar



car(c), run(c,d), diesel(d)

Statistical modules

Language models

To choose between comparable intermediate results
the black cat/the cat black

Hypertaggers

To prune the initial search space

Rankers

To determine the best output

Generating from Text

In 1964 Peter Higgs published his second paper in Physical Review Letters describing Higgs mechanism [which] predicted a new massive spin-zero boson for the first time.

Peter Higgs wrote his paper explaining Higgs mechanism in 1964.

Higgs mechanism predicted a new elementary particle.

Split

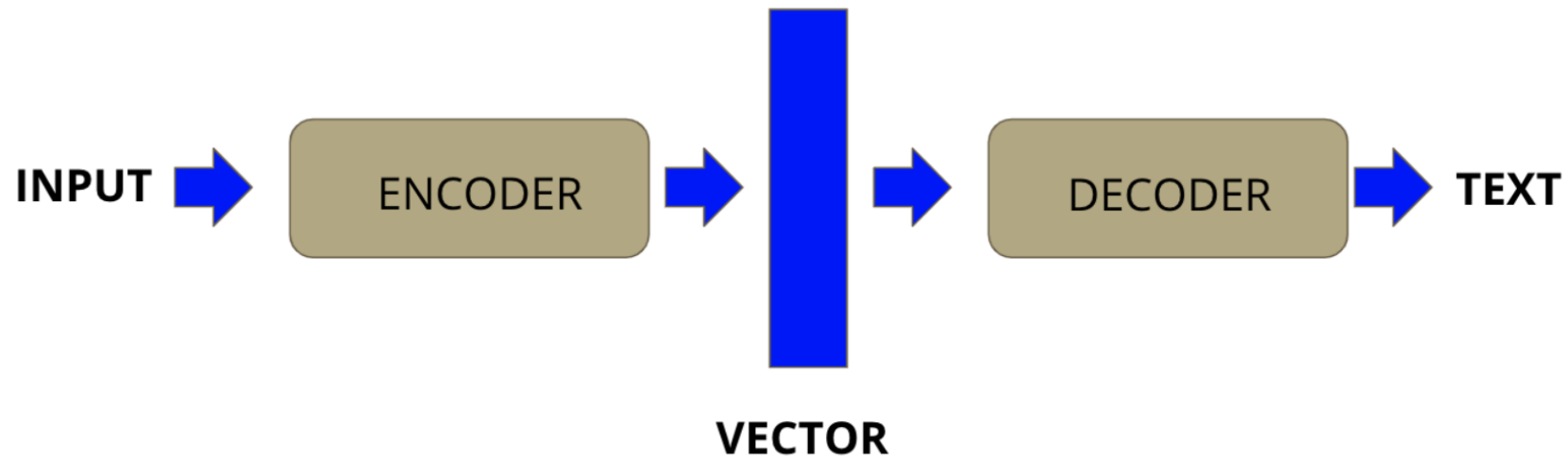
Rewrite

Reorder

Delete

Neural NLG

A Unifying Framework



Outline

Encoding

- Modelling Graph-Structured Input
- Local + Global Context

Scaling

- Encoding Large Quantities of Text

Generalising

- Delexicalisation

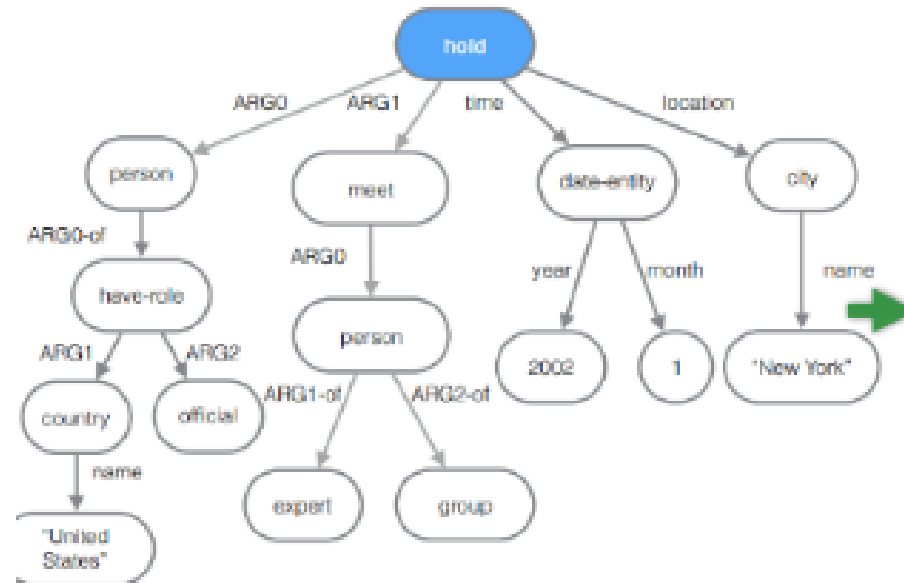
Training Data

- Making implicit knowledge explicit

Encoding

Modelling Graph-Structured Input

Generation from AMR 2017 Challenge



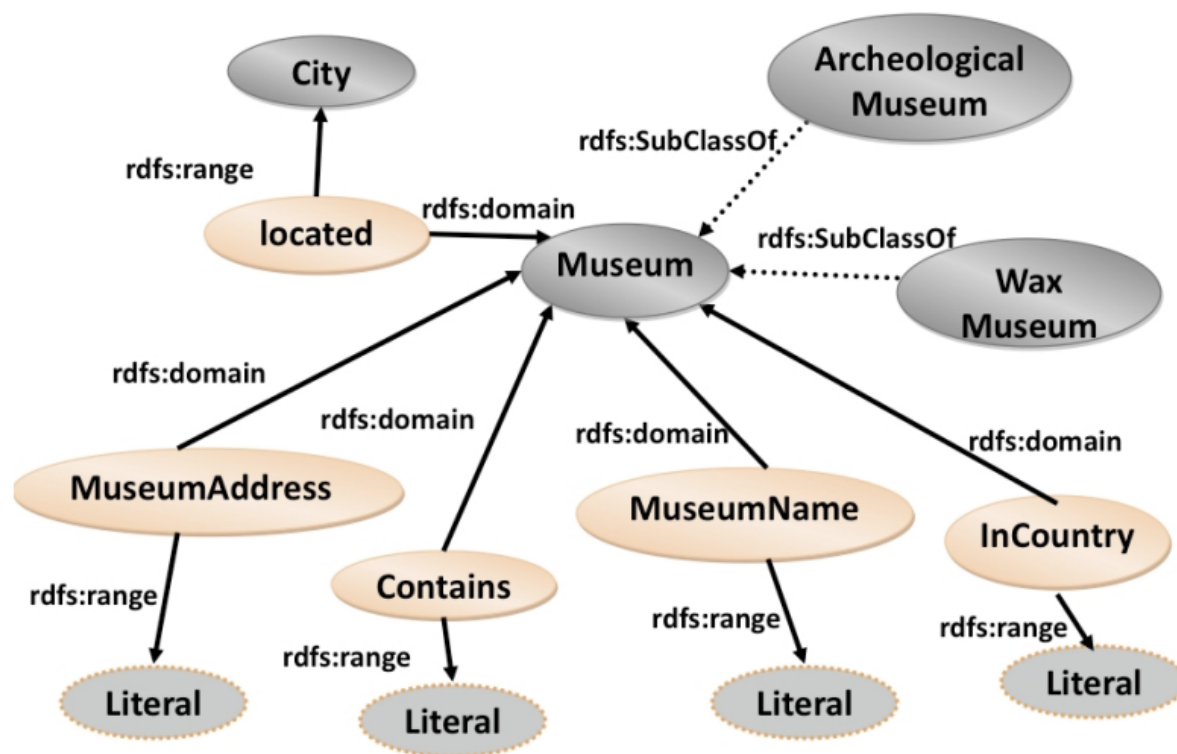
```

hold
  :ARG0 (person
        :ARG0-of (have-role
                  :ARG1 United_States
                  :ARG2 official)
        )
  :ARG1 (meet
        :ARG0 (person
              :ARG1-of expert
              :ARG2-of group)
        )
  :time (date-entity 2002 1)
  :location New_York
  
```

US officials held an expert group meeting in January 2022 in New York .

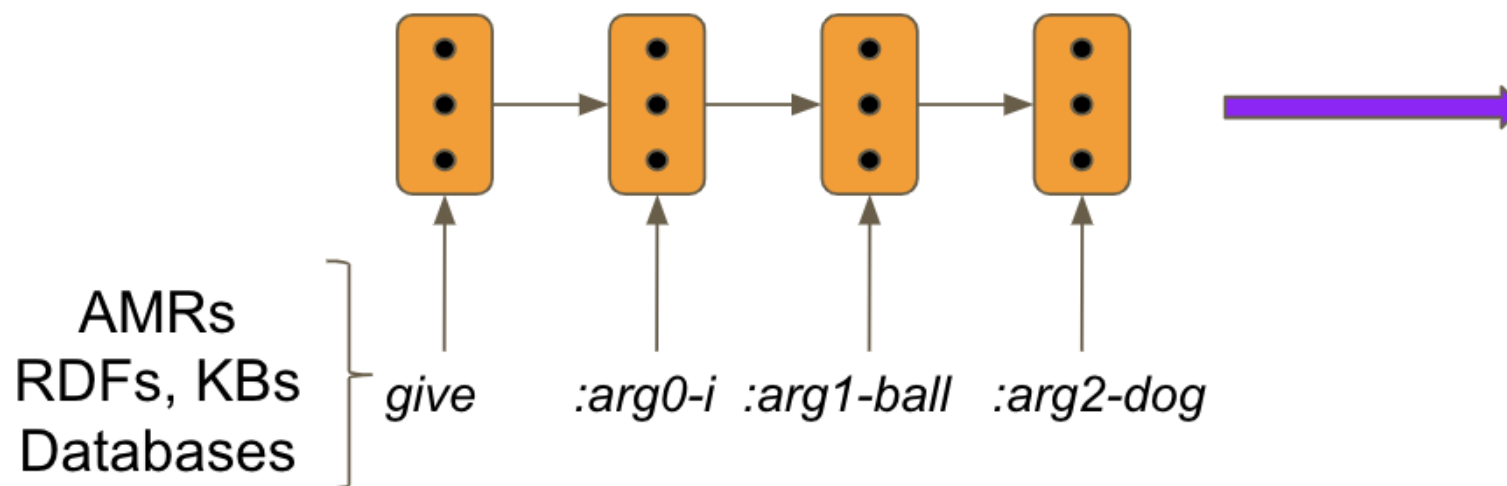
Generating from Sets of RDF Triples

WebNLG Challenge 2017



Graph Structure as a Sequence

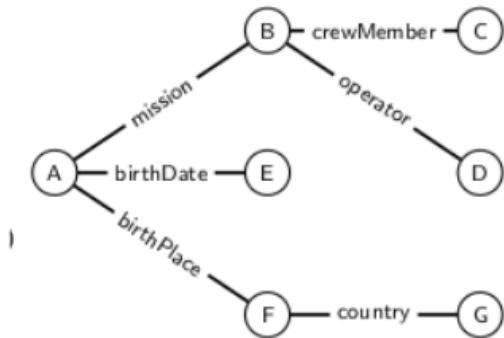
Early approaches to MR- or Data-to-Text generation encode the input structure as a sequence.



Problems with Graph Linearization

- Local dependencies available in the input turned into long-range dependencies
- RNNs often have trouble modeling long-range dependencies

D2T Generation (Data = RDF)

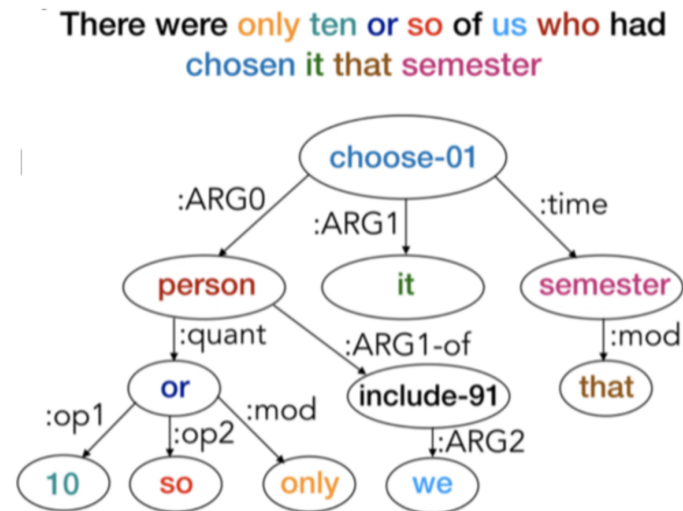


LINEARISATION

Alan_Bean **mission** Apollo_12 Apollo_12
crewMember Peter_Conrad Apollo_12
operator Nasa Alan_Bean birthDate
 1932-03-15 Alan_Bean birthPlace
 Wheeler_Texas Wheeler_Texas country
 USA

NLG with Graph Encoders

- (Damonte and Cohen, 2019): Graph Convolutional Networks
- (Song et al., 2018): Graph LSTM
- (Becker et al., 2018): GRU LSTM
- (Guo et al., TACL 2019): Densely Connected Graph Convolutional Network



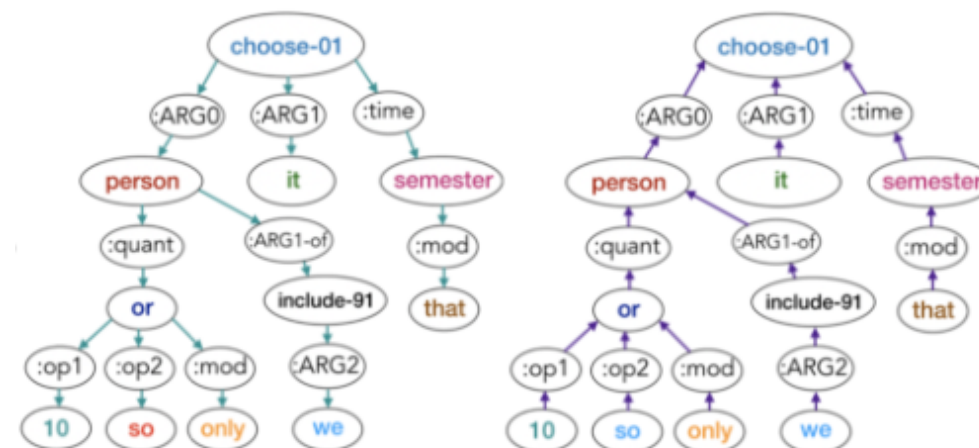
Dual Top-Down and Bottom-Up Graph Encoding

Ribeiro, Gardent, Gurevitch (EMNLP 2019)

$$\mathbf{h}_i^t = GE_t(\mathbf{n}_i, G_t)$$

$$\mathbf{h}_i^b = GE_b(\mathbf{n}_i, G_b),$$

$$\mathbf{n}_i = [\mathbf{h}_i^t \parallel \mathbf{h}_i^b \parallel \mathbf{l}_i]$$



Poster on Wednesday

Results

Cao & Clark 2019

- Delexicalisation
- Added Syntactic Information

Damonte et al. 2019

- Delexicalisation

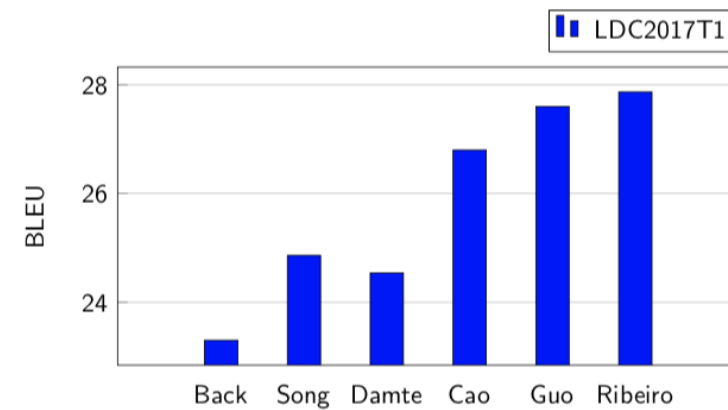
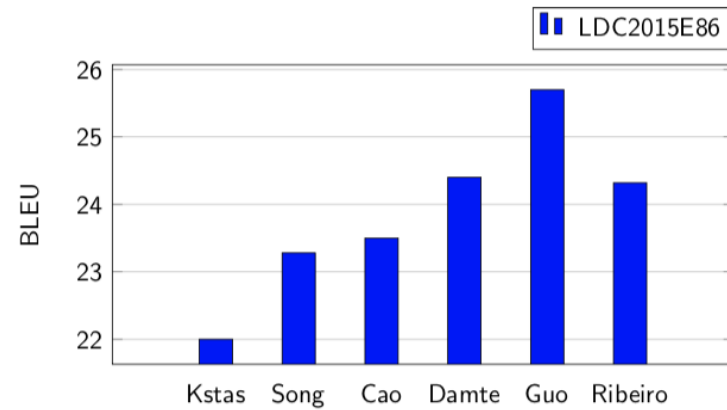
Guo et al. 2019

- Densely Connected Graph Convolutional Network

With data expansion

Model	External	BLEU
Konstas et al. (2017)	200K	27.40
Song et al. (2018)	200K	28.20
Guo et al. (2019)	200K	31.60
G2S-GGNN	200K	32.23

Table 3: Results on LDC2015E86 test set when models are trained with additional Gigaword data.



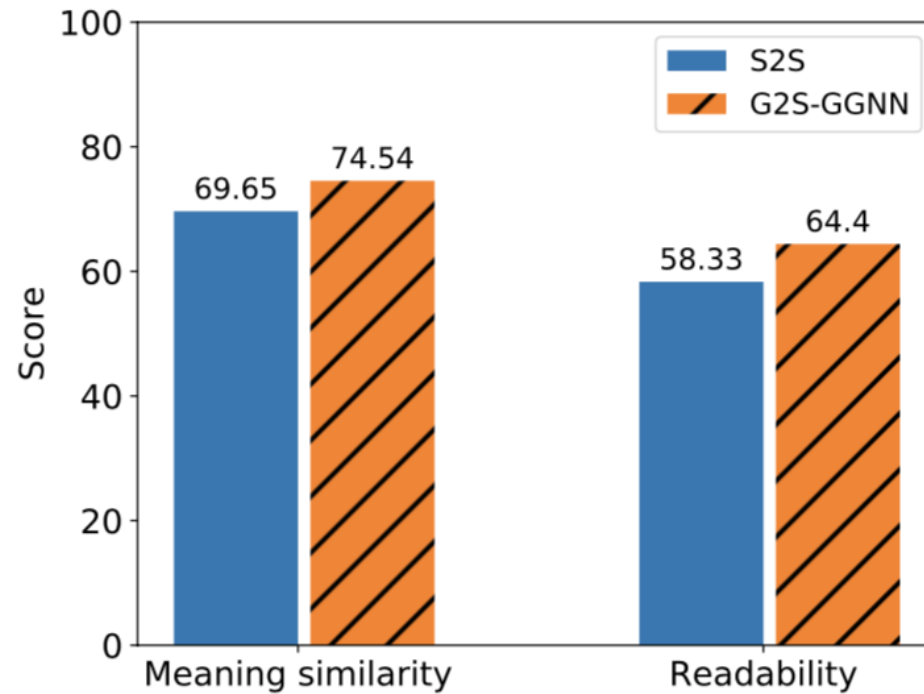
Semantic Adequacy

Does the reference entails the generated sentence and vice versa ?

REF \Rightarrow GEN			
Model	ENT	CON	NEU
S2S	38.45	11.17	50.38
G2S-GIN	49.78	9.80	40.42
G2S-GAT	49.48	8.09	42.43
G2S-GGNN	51.32	8.82	39.86
GEN \Rightarrow REF			
Model	ENT	CON	NEU
S2S	73.79	12.75	13.46
G2S-GIN	76.27	10.65	13.08
G2S-GAT	77.54	8.54	13.92
G2S-GGNN	77.64	9.64	12.72

Human Evaluation

Semantic adequacy and readability



Graph-Structure vs. N-Grams

GOLD

China and Kyrgyzstan **agreed in a joint communique that** terrorism, separatism and extremism still pose major threats to regional security and stability.

S2S

In the joint communique, China and Kyrgyzstan still **agreed to** threaten terrorism, separatism, extremism and regional stability.

Song et. al (2018)

In a joint communique, China and Kyrgyzstan have **agreed to** still be a major threat to regional security, and regional stability.

G2S-GGNN

At a joint communique, China and Kyrgyzstan **agreed that** terrorism, separatism and extremism are still a major threat to region security and stability.

Encoding

Combining Local and Global Information

Local + Global Encoding of AMR nodes

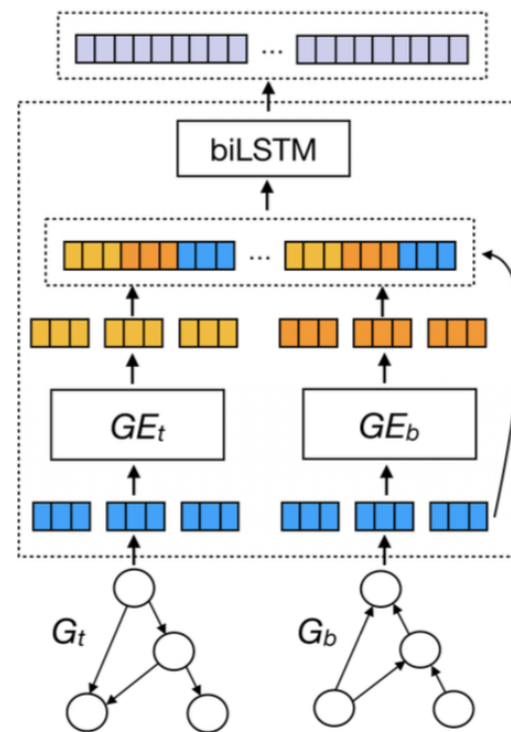
Ribeiro, Gardent, Gurevych (EMNLP 2019)

Local

- Node Embeddings (h_t, h_b)
- Label embedding (h_l)

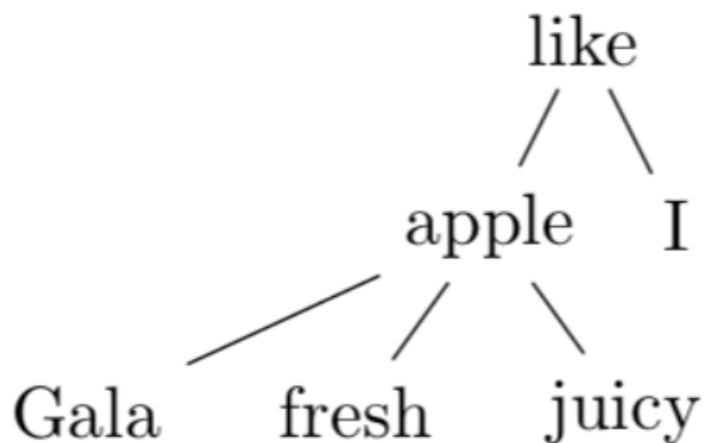
Global

- bi-LSTM Encoding of the whole graph



Surface Realisation Shared Task (Shallow Track)

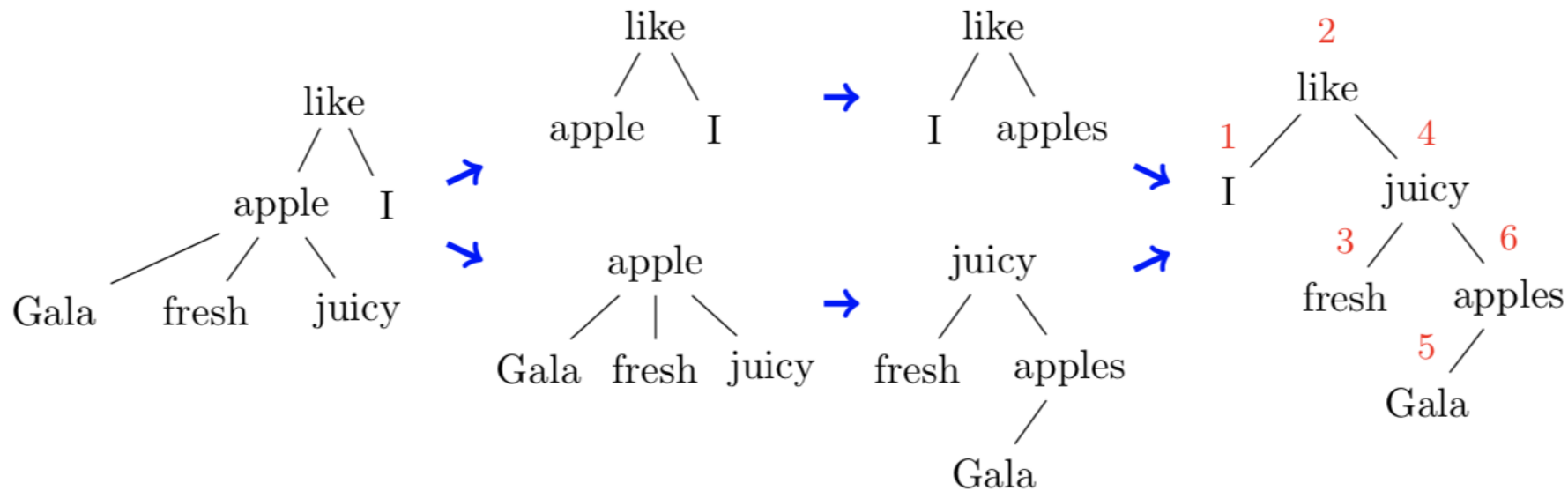
Unordered, Lemmatized Dependency Tree \Rightarrow Sentence



I like fresh juicy Gala apples

Surface Realisation Shared Task (Shallow Track)

Puzikov, Gardent, Dagan, Gurevych (INLG 2019)



Local + Global Encoding of Dependency Tree Nodes

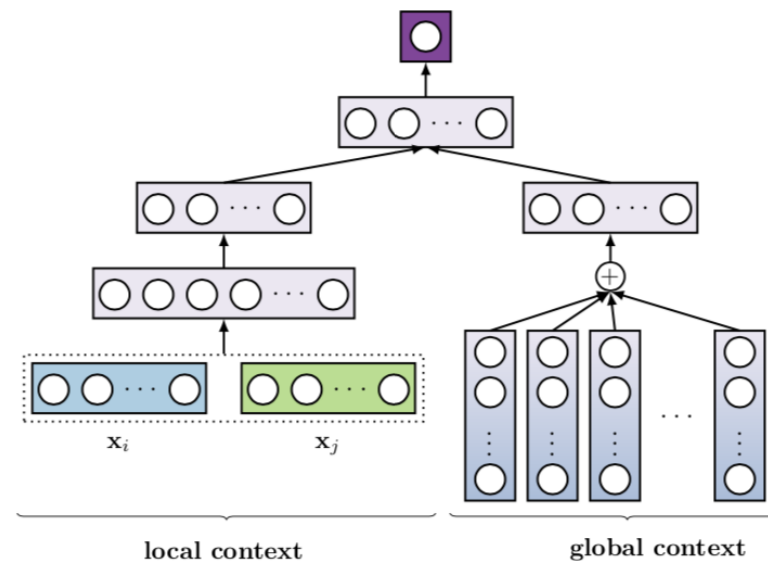
Puzikov, Gardent, Dagan, Gurevych (INLG 2019)

Local

- Dense feature representations of the two nodes
- Head node, Child node

Global

- All nodes of the subtree dominated by the head node
- Weighted sum of their feature representations



	BLEU	EDIST	NIST
BINLIN	24.92	35.91	9.55
+ data enrichment	48.47	62.04	10.72
+ new encoder	50.67	64.05	10.82

Encoding at Scale

Discrete Graph Encoding of Text

Encoding Large Quantities of Text

Question-Answering on Free Form Web Text

- ELI5 Dataset
- Query = NL question
- TASK: Generate answer from question + [web text \(200K words\)](#)

Multi-Document Summarisation

- Wikisum Dataset
- Query = WKP article title
- TASK: Generate lead paragraph from title + [web text \(200K words\)](#)

Question: Why consumers are still so terrified of genetically modified organisms (GMOs), yet there is little debate in the scientific community of whether they are safe or not. (Scientists are for GMOs)

Beginning of Web Search: The controversial safety of gmos and the skepticism of the use of gmos College paper Writing Service URL.0. diamond chemicals plc the merseyside. appendix f research question for. antisocial personality disorder affects family relations and interactions. The controversial safety of gmos and the skepticism of the use of gmos. Gmo facts what is a gmo genetically modified organisms the safety of gmos is unknown poll: skepticism of genetically modified foods abc news abc news network, 19 june 2015 web fernandez-cornejo, jorge, and seth james wechsler. The controversy over gmos, particularly in food continues: scientists are split pros and cons of gmo's september 28 and environmentalists and consumer groups remain unconvinced about the safety of gmos. The controversy around genetically modified foods on the surface, food evolution is a resetting of the controversial conversation around genetically modified organisms (gmos) we just ask people by a show of hands to tell us are they concerned about gmos for their own safety or the. When gmos are the movie star can documentaries on controversial science be entertaining and the message is that gmo food is safe to eat and that naysayers are. Genetically modified organisms (gmos) the top five anti-gmo tropes gmos are genetically modified organisms the evidence on gmo safety by ramez naam. Genetically modified organisms what are gmos with the use of gm technology, pure and safe equivalents can be produced using gmos and industrial scale. Here's a bullet-point summation of what nathanael johnson learned about gmos in 2013 20 gmo questions: animal, vegetable, controversy by pretty darn safe. The controversy surrounding genetically modified organisms what do we tolerate as far as detrimental this would be a profound service to scientific skepticism with regards to gmos current gmos are safe to eat- however [...]

Target Answer: There is little difference in essence between what is called GMO now and the techniques we have been using to domesticate and cultivate the food in the past. Its an arbitrary line that's been drawn in the sand and people fail to realize this often. That being said I think it is more then wrong the patenting of crops and again even more then wrong to genetically modify crops to not have viable seeds so that seed-washing can't be used to grow the next crop. So the real god damned issues are being masked and lost by this retarded polemic between GMO and more conventional genetic modification of organisms.

How to encode 200K words and generate from it ?

Previous Work

- TF-IDF-based information extraction
- Limits the input information to a few thousand words

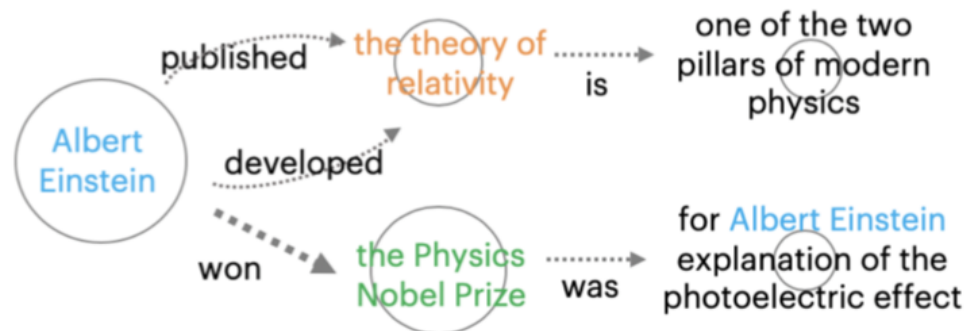
Knowledge Graphs

- Convert text to a graph
- Scales to 200K words



Non Neural Graph Encoding of Text: 200K \Rightarrow 10K Tokens

Q: Explain the theory of relativity



DOCUMENT 1

Albert Einstein, a German theoretical physicist, published the theory of relativity.

The theory of relativity is one of the two pillars of modern physics.

He won the physics Nobel Prize.

DOCUMENT 2

Albert Einstein (March 14, 1879 to April 18, 1955) developed the theory of relativity.

He won the Nobel Prize.

The great prize was for his explanation of the photoelectric effect.

Constructing the Knowledge Graph

QUERY: Can someone finally explain the theory of general relativity?

DOCUMENT SENTENCES with GRAPH OPERATIONS

① Albert Einstein, a German theoretical physicist, published the theory of relativity.

ADDED TO GRAPH

② The theory of relativity is one of the two pillars of modern physics.

MERGE OPERATION:

theory of relativity

EXISTS AS A NODE

NODE WEIGHT + 1

③ He won the physics Nobel Prize.

COREFERENCE:

he and Albert Einstein

MERGE OPERATION:

Albert Einstein

EXISTS AS A NODE

NODE WEIGHT + 1

④ Puppies are very cute.

FILTER OPERATION:

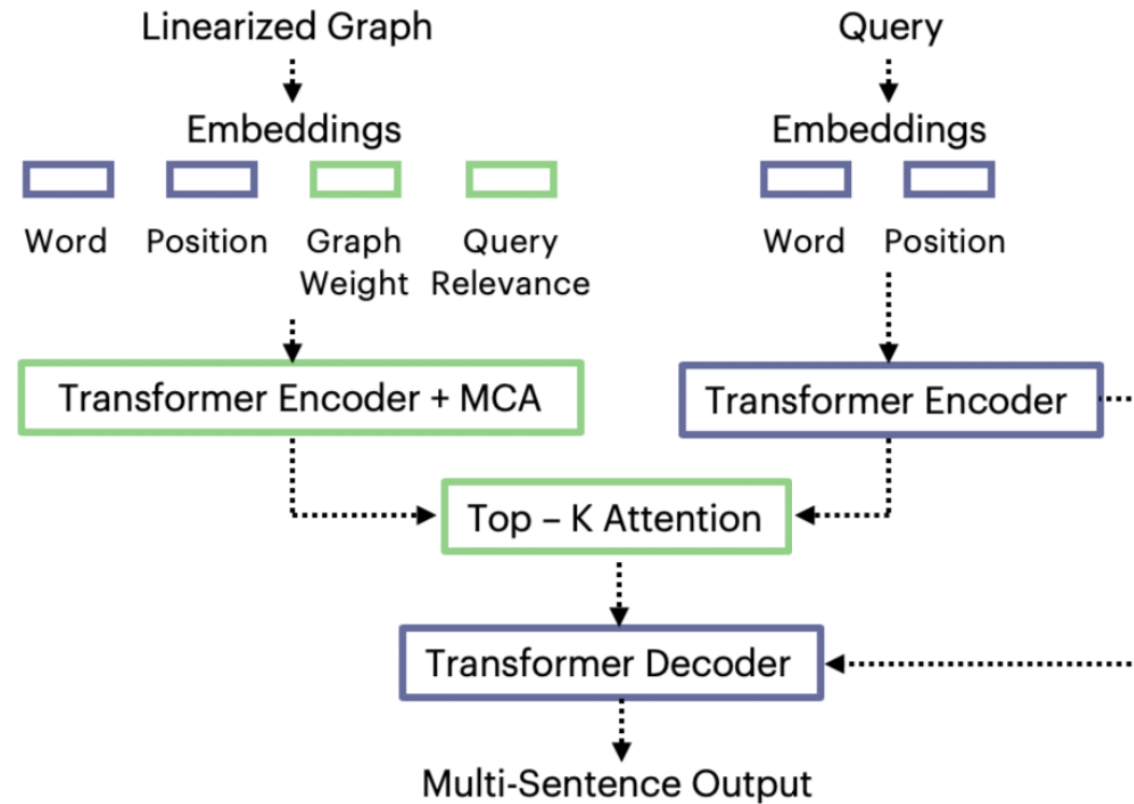
low TF-IDF overlap with query

NOT ADDED TO GRAPH

Rich Node Embeddings

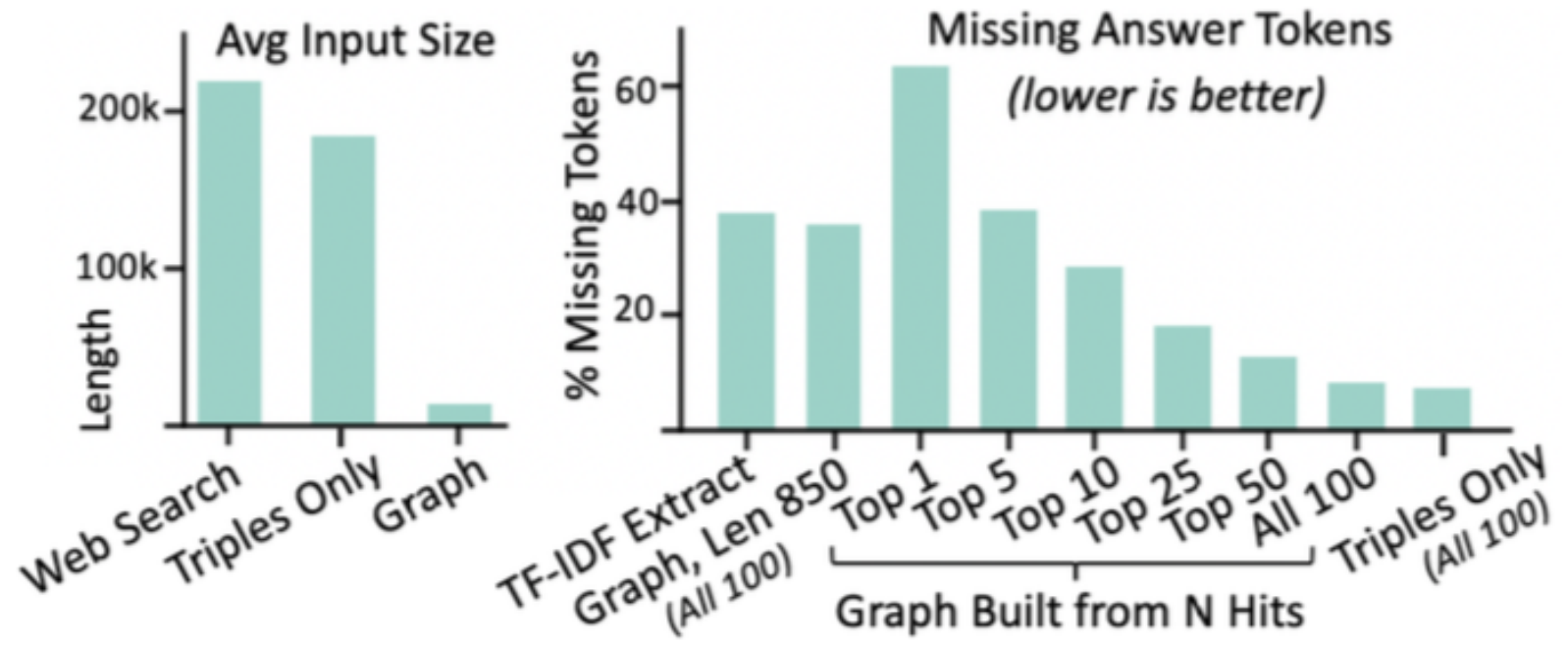
WORD EMBEDDING	<sub>	Albert	Einstein	<obj>	the	theory	of	relativity	<pred>	published	<s>	developed	<obj>	the	Physics	Nobel	Prize	<s>	won
POSITION EMBEDDING	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
GRAPH WEIGHT EMBEDDING	0	4	4	0	2	2	2	2	0	1	0	1	0	3	3	3	3	0	2
QUERY RELEVANCE EMBEDDING	0	1	1	0	1	1	1	1	0	1	0	2	0	1	1	1	1	0	1

Multi-Document Question Answering



Compression vs. Information Preservation

Graph construction drastically reduces the input size while preserving key information



Results

Eli5

Model	Input Length	ROUGE		
		1	2	L
Q + D to A*, TF-IDF	avg 850	28.3	5.1	22.8
Q + D to A, MMR	avg 850	28.1	5.0	22.9
Multi-task*	avg 850	28.9	5.4	23.1
Multi-task Triples	850	29.0	5.2	23.2
Multi-task Top20 Trip.	avg 570	28.8	5.3	23.2
Q + D to A Graph	850	28.8	5.3	23.3
Multi-task Graph	850	29.5	5.6	23.6
+ Top-100 Attention	850	29.7	5.7	23.8
+ E-MCA	11K	30.0	5.8	24.0

WikiSum

Model	Input Len	ROUGE-L
T + D to P*	500	34.2
Language Model + D-MCA*	11K	36.2
T + D to P	500	33.8
Multi-task	500	34.4
Multi-task Graph	500	34.9
+ Top-100 Attention	500	35.2
+ E-MCA	11K	36.5

Question: Why is touching microfiber towels such an uncomfortable feeling?

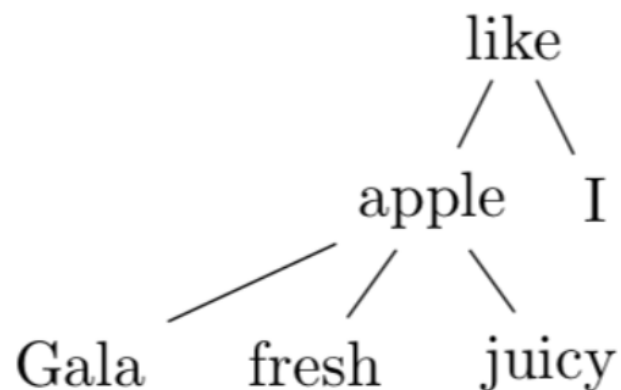
True Answer: Do you mean the kind of cloths used to clean glasses and lenses? I've never noticed any uncomfortable feeling myself, but I do find touching certain cleaning cloths can be quite uncomfortable. There's a brand called "e - cloth" which market themselves as not needing any cleaning supplies. Basically the threads have tiny little hooks and barbs in them, so dragging them over skin makes you feel like it 's catching against you.

Generation: The microfiber is made up of a bunch of tiny fibers that are attached to them. When you touch them, the fibers that make up the microfiber are attracted to each other. When they touch, they're attracted to the other end of the fiber, which is what makes them feel uncomfortable.

Encoding

Delexicalisation

SR Task, Shallow Track



I like fresh juicy Gala apples

Intuition

Word ordering mainly depends on syntax

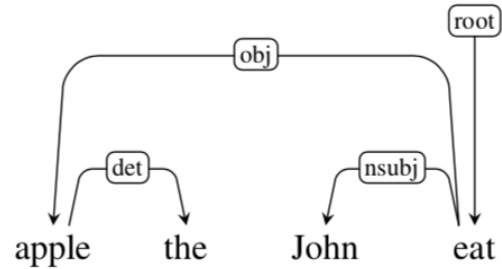
E.g., in English, the **subject** generally precedes the **verb**

Abstracting away from specific words

- reduces data sparsity
- helps handling rare/unknown words

Delexicalising

Shimorina, Gardent (EMNLP 2019)



]

Linearised, Delexicalised Tree

id1:verb:root:0 id2:noun:obj:1 id3:det:det:2 J
id4:pnoun:nsbj

Target Delexicalised Sequence

id4:pnoun:nsbj id1:verb:root:0 id3:det:det:2 J
id2:noun:obj:1

Output

John eat the apple

Results

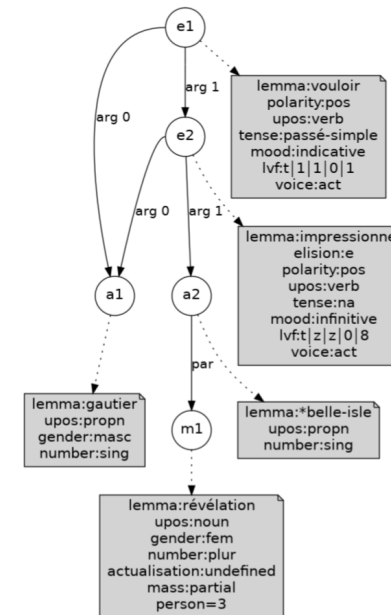
	ar	cs	en	es	fi	fr	it	nl	pt	ru
BL	29.6±1.39	48±0.7	53.57±0.15	46.5±0.78	27.2±0.4	46.4±0.5	49.07±0.9	36.6±0.7	44.3±0.26	58.1±0.46
WO	34.9±0.2	57.97±0.06	59.1±0.36	52.33±0.31	43.1±0.53	50.0±0.0	53.17±0.5	47.03±0.59	51.77±0.32	64.73±0.23
Δ	+5.3	+9.97	+5.53	+5.83	+15.9	+3.6	+4.1	+10.43	+7.47	+6.63

- BLEU score on lemmatised data
- Mean and Standard Deviation on 3 runs
- All results are statistically significant ($p < 0.05$)
- Baseline = Lexicalised Input

AMRs \Rightarrow Sentence

Colin, Gardent (INLG 2019)

- Factored, S2S Model with Attention
- Concatenation of 18 embeddings
- Each embedding represent a feature type (POS tag, gender, number etc.)



Lemma linearization:

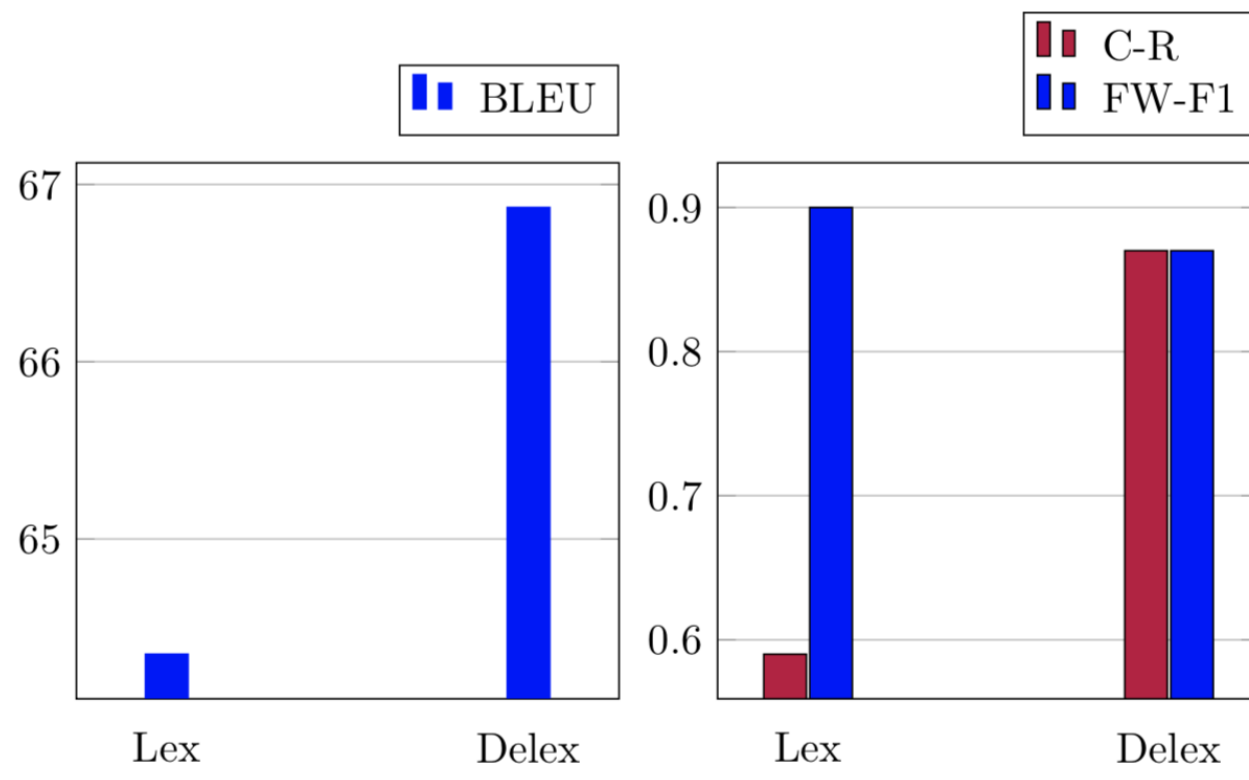
```
( vouloir|pos|verb|ps|... arg_0
  gautier|proprn|masc|sg )
( vouloir|pos|verb|ps|... arg_1
  impressionner|e|pos|verb|... )
( impressionner|e|pos|verb|... arg_0
  gautier|proprn|masc|sg )
( impressionner|e|pos|verb|... arg_1
  *belle-isle|proprn|sg )
par ( *belle-isle|proprn|sg <
  révélation|noun|fem|plur|... )
```

Anonymized linearization:

```
( e1|pos|verb|ps|... arg_0
  a1|proprn|masc|sg )
( e1 arg_1|pos|verb|ps|... ]
  e2|e|pos|verb|... )
( e2|e|pos|verb|... arg_0
  a1|proprn|masc|sg )
( e2|e|pos|verb|... arg_1
  a2|proprn|sg )
par ( a2|proprn|sg <
  m1|noun|fem|plur|... )
```

Gautier voulut impressionner Belles-Isle par des révélations.
(Gautier wanted to impress Belles-Isle by some revelations)

Results



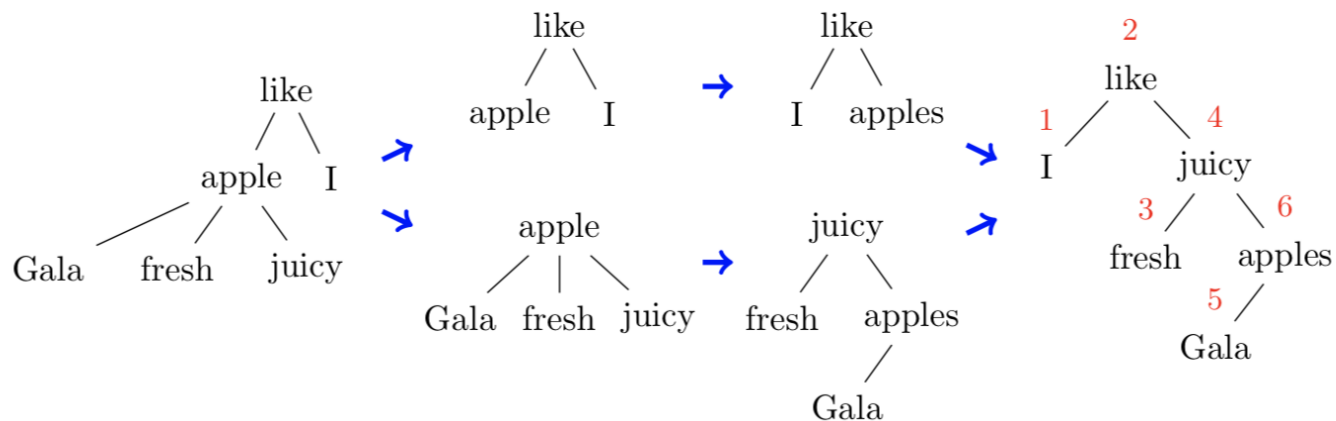
Recall (C-R) and F1 on function words
How well does the model reconstruct function words ?]

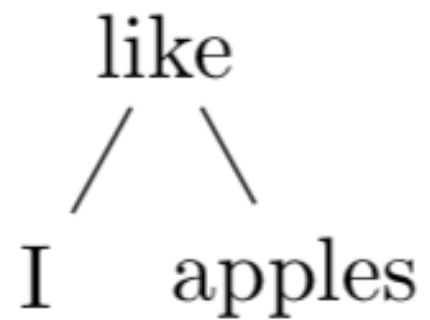
Data

Better Data

Word Ordering

- Convert input dependency into binary tree
- Use Multi-Layer Perceptron to
predict the relative order of head and child node





Training Example

("like", "I", "left")

Position of Child wrt Head

Added Training Example

("I", "like", "right")

Position of Head wrt Child

Results

Dual Head/Child, Child/Head encoding improves performance

	BLEU	EDIST	NIST
BINLIN	24.92	35.91	9.55
+ data enrichment	48.47	62.04	10.72

Further Key Issues

Multilingual Generation

- Surface Realisation Shared Task (10 languages, MR2T)
- From data and text ?

Interpretability

- Modular Approaches

Questions ?