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

Verbalise
Respond
Summarize
Simplify

Communicative Goal

Input

Databases
Knowledge bases
Dialog acts
Sentences
Documents
...
Generating from Data

The NLG Pipeline

- **Document Planning**: content planning, document outline
- **Microplanning**: referring expressions, word choice, aggregation
- **Realisation**: converting specifications to a real text
Generating from Meaning Representations

Grammar

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

\[ \text{car}(c), \text{run}(c,d), \text{diesel}(d) \]
In 1964, Peter Higgs published his second paper in Physical Review Letters describing the Higgs mechanism [which] predicted a new massive spin-zero boson for the first time.

Peter Higgs wrote his paper explaining the Higgs mechanism in 1964. The Higgs mechanism predicted a new elementary particle.
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

US officials held an expert group meeting in January 2002 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)

\[ h^t_i = GE_t(n_i, G_t) \]

\[ h^b_i = GE_b(n_i, G_b), \]

\[ n_i = [h^t_i \| h^b_i \| l_i] \]
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

<table>
<thead>
<tr>
<th>Model</th>
<th>External</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Konstas et al. (2017)</td>
<td>200K</td>
<td>27.40</td>
</tr>
<tr>
<td>Song et al. (2018)</td>
<td>200K</td>
<td>28.20</td>
</tr>
<tr>
<td>Guo et al. (2019)</td>
<td>200K</td>
<td>31.60</td>
</tr>
<tr>
<td>G2S–GGNN</td>
<td>200K</td>
<td><strong>32.23</strong></td>
</tr>
</tbody>
</table>

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?

<table>
<thead>
<tr>
<th>Model</th>
<th>ENT</th>
<th>CON</th>
<th>NEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>S2S</td>
<td>38.45</td>
<td>11.17</td>
<td>50.38</td>
</tr>
<tr>
<td>G2S–GIN</td>
<td>49.78</td>
<td>9.80</td>
<td>40.42</td>
</tr>
<tr>
<td>G2S–GAT</td>
<td>49.48</td>
<td>8.09</td>
<td>42.43</td>
</tr>
<tr>
<td>G2S–GGNN</td>
<td>51.32</td>
<td>8.82</td>
<td>39.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>ENT</th>
<th>CON</th>
<th>NEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>S2S</td>
<td>73.79</td>
<td>12.75</td>
<td>13.46</td>
</tr>
<tr>
<td>G2S–GIN</td>
<td>76.27</td>
<td>10.65</td>
<td>13.08</td>
</tr>
<tr>
<td>G2S–GAT</td>
<td>77.54</td>
<td>8.54</td>
<td>13.92</td>
</tr>
<tr>
<td>G2S–GGNN</td>
<td>77.64</td>
<td>9.64</td>
<td>12.72</td>
</tr>
</tbody>
</table>
Human Evaluation

Semantic adequacy and readability

![Bar chart showing comparison between S2S and G2S-GGNN for Meaning similarity and Readability scores.](chart.png)
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 ($h_l$, $h_b$)
- 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
```

```
like
  \   \  
apple  I
    \  \  
  Gala  fresh  juicy
```
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
<table>
<thead>
<tr>
<th></th>
<th>BLEU</th>
<th>EDIST</th>
<th>NIST</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BINLIN</strong></td>
<td>24.92</td>
<td>35.91</td>
<td>9.55</td>
</tr>
<tr>
<td>+ data enrichment</td>
<td>48.47</td>
<td>62.04</td>
<td>10.72</td>
</tr>
<tr>
<td>+ new encoder</td>
<td>50.67</td>
<td>64.05</td>
<td>10.82</td>
</tr>
</tbody>
</table>
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
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 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**

3. He won the physics Nobel Prize.

**COREFERENCE:**
he and Albert Einstein

**MERGE OPERATION:**
Albert Einstein
EXISTS AS A NODE
NODE WEIGHT + 1

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

4. Puppies are very cute.

**FILTER OPERATION:**
low TF-IDF overlap with query
NOT ADDED TO GRAPH
Rich Node Embeddings

<table>
<thead>
<tr>
<th>WORD EMBEDDING</th>
<th>Albert Einstein</th>
<th>the theory of relativity</th>
<th>published</th>
<th>developed</th>
<th>the Physics Nobel Prize</th>
<th>won</th>
</tr>
</thead>
<tbody>
<tr>
<td>POSITION EMBEDDING</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>GRAPH WEIGHT EMBEDDING</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>QUERY RELEVANCE EMBEDDING</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Multi-Document Question Answering

Linearized Graph

Embeddings

- Word
- Position
- Graph Weight
- Query Relevance

Transformer Encoder + MCA

Top - K Attention

Transformer Decoder

Multi-Sentence Output

Query

Embeddings

- Word
- Position
Compression vs. Information Preservation

Graph construction drastically reduces the input size while preserving key information.
# Results

## Eli5

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

## WikiSum

<table>
<thead>
<tr>
<th>Model</th>
<th>Input Len</th>
<th>ROUGE-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>T + D to P*</td>
<td>500</td>
<td>34.2</td>
</tr>
<tr>
<td>Language Model + D-MCA*</td>
<td>11K</td>
<td>36.2</td>
</tr>
<tr>
<td>T + D to P</td>
<td>500</td>
<td>33.8</td>
</tr>
<tr>
<td>Multi-task</td>
<td>500</td>
<td>34.4</td>
</tr>
<tr>
<td>Multi-task Graph</td>
<td>500</td>
<td>34.9</td>
</tr>
<tr>
<td>+ Top-100 Attention</td>
<td>500</td>
<td>35.2</td>
</tr>
<tr>
<td>+ E-MCA</td>
<td>11K</td>
<td>36.5</td>
</tr>
</tbody>
</table>
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


Target Delexicalised Sequence


Output

John eat the apple
Results

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

- 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:

\[
\begin{align*}
&\text{vouloir|pos|verb|ps|... arg_0} \\
&\text{gautier|propn|masc|ag} \\
&\text{vouloir|pos|verb|ps|... arg_1} \\
&\text{impressionner|e|pos|verb|...} \\
&\text{impressionner|e|pos|verb|... arg_0} \\
&\text{gautier|propn|masc|ag} \\
&\text{impressionner|e|pos|verb|... arg_1} \\
&\text{+belle-isle|propn|sg} \\
&\text{par ( +belle-isle|propn|sg <} \\
&\text{révélation|noun|fem|plur|...})
\end{align*}
\]

Anonymized linearization:

\[
\begin{align*}
&\text{e1|pos|verb|ps|... arg_0} \\
&\text{a1|propn|masc|ag} \\
&\text{e1 arg_1|pos|verb|ps|...} \\
&\text{e2|e|pos|verb|...} \\
&\text{e2|e|pos|verb|... arg_0} \\
&\text{a1|propn|masc|ag} \\
&\text{a2|propn|a1} \\
&\text{par ( a2|propn|ag <} \\
&\text{ml|noun|fem|plur|...})
\end{align*}
\]

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

<table>
<thead>
<tr>
<th></th>
<th>BLEU</th>
<th>EDIST</th>
<th>NIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>BINLIN</td>
<td>24.92</td>
<td>35.91</td>
<td>9.55</td>
</tr>
<tr>
<td>+ data enrichment</td>
<td>48.47</td>
<td>62.04</td>
<td>10.72</td>
</tr>
</tbody>
</table>
Further Key Issues

Multilingual Generation

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

Interpretability

- Modular Approaches
Questions?