Generating Text from various sources into multiple languages

Claire Gardent
CNRS/LORIA

Chaire IA xNLG “Generating from Multiple Sources into Multiple Languages”
Natural Language Processing

Analysis (Read)

Generating (Write)
Generating Text

Inputs

Tabular Data, Data Bases, Knowledge Bases

Text

Graph
Generating Text

And also ...

Images

Videos
Generating Text

What for?

Verbalising
• A graph
• A Data/Knowledge Base

Simplify, paraphrase, summarise
One or more text(s)
Conditional Text Generation

Input -> Text

ENCODER

DECODER

Input

Text
Conditional Text Generation

Input -> Text

ENCODER

DECODER

A Unifying Framework for Text Generation
Conditional Text Generation

From

• Meaning Representations
  Abstract Meaning Representations (AMR) -> 21 UE languages

• Knowledge Graphs
  Resource Description Format (RDF) -> Low Resource Languages (Breton, Welsh, Irish)

• Texts
  - Generating Wikipedia Biographies (Biais de données)
  - Simplifying Documents
Evaluating Generated Text

• Sentence Simplification
  - Sentence Level Estimate (SLE)

• KB-to-Text
  - Entity-Based Semantic Adequacy
  - A Semantic Approach
  - A Question-Based Approach
Verbalising Abstract Meaning Representations into 21 EU Languages

Angela Fan and Claire Gardent
“Multilingual AMR-to-Text Generation”
EMNLP 2020
**AMR Graph (Abstract Meaning Representation)**

- Acyclic Graph
- Nodes: concepts
- Edges: semantic roles

Ryan describes himself as a genius
US officials held an expert group meeting in January 2002 in New York.

Des responsables américains ont tenu une réunion d’un groupe d’experts en janvier 2002 à New York.

Funcionarios estadounidenses celebraron una reunión de un grupo de expertos en enero de 2002 en Nueva York.

Amerikáni predstavitelia usporiadali stretnutie expertnej skupiny v januári 2002 v New Yorku.

Amerikanska tjänstemän höll ett expertgruppsmöte i januari 2002 i New York.

Roman, Germanic, Slavic, Uralic
Challenges

• Graph $\rightarrow$ Sequence
Challenges

- Graph $\rightarrow$ Sequence
- Lack of Training/Test Data
Challenges

• Graph $\rightarrow$ Sequence

• Lack of Training/Test Data

• Underspecified Input
Challenges

- Graph $\rightarrow$ Sequence
- Lack of Training/Test Data
- Underspecified Input
- Multilingual: Generate texts with varied syntax and morphology
Encoder

Graph ➔ Vector
Encoder

- Linearisation
- Structural Embeddings
- Sub-words
- Pre-Training (MLM)
Linearisation

hold
:ARG0 person : ARG0-of have-org-role :ARG1 :op1 United :op2 States : ARG2 official
:ARG1 meet :ARG0 person : ARG1-of expert : ARG2-of group
:time date-entity : year 2002 : month 1
:location city : op1 New : op2 York
Structural Embeddings for Siblings and Branches
Pre-training

- Silver AMRs
- 30M sentences from CCNET parsed using JAMR
Decoder (Generation)

Vector → Text
Multilingual Generation

- Multilingual XLM Embeddings
- Pre-trained Language Model (30M sentences) for each language
- Multilingual Encoder-Decoder
XLM Multilingual Word Embeddings

curtains were

Transformer Model

<table>
<thead>
<tr>
<th>Token</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>/s</td>
<td>the</td>
<td>MASK</td>
<td>MASK</td>
<td>blue</td>
<td>/s</td>
<td>/s</td>
<td>MASK</td>
<td>rideaux</td>
<td>étaient</td>
<td>MASK</td>
<td>/s</td>
</tr>
<tr>
<td>Language</td>
<td>en</td>
<td>en</td>
<td>en</td>
<td>en</td>
<td>en</td>
<td>en</td>
<td>en</td>
<td>fr</td>
<td>fr</td>
<td>fr</td>
<td>fr</td>
<td>fr</td>
</tr>
</tbody>
</table>

Cross-lingual Language Model Pretraining
Guillaume Lample, Alexis Conneau
Des responsables américains ont tenu une réunion d'un groupe d'experts en janvier 2002 à New York.
Multilingual AMR-to-Text Model

Encoder
- Pre-trained AMRs
- Structural Embeddings

Decoder
- Pre-trained LMs
- XLM Vocabulary

Multilingual Encoder-Decoder
Des responsables américains ont tenu une réunion d'un groupe d'experts en janvier 2002 à New York.

Funcionarios estadounidenses celebraron una reunión de un grupo de expertos en enero de 2002 en Nueva York.

Americkí predstavitelia usporiadali stretnutie expertnej skupiny v 2002 v New Yorku.

Американските служители проведоха среща на експертна група през януари 2002 г. в Ню Йорк.

Американска tjänstemän höll ett expertgruppsmöte i januari 2002 i New York.
Data

Training
• Silver AMR, Europarl Text, 21 languages

Test
• Silver AMR, Europarl text, 21 languages
• Gold AMR, LDC Text, Espagnol, allemand, italien
Evaluation

Automatic Metrics (BLEU)
• Ablation
• Comparison with baselines
• Impact of related languages

Human-Based
  Word-Order, Morphology, Semantic adequacy, Paraphrasing
Ablation

Baseline (English, BLEU)                        32.5
+ Graph embeddings                       32.9
+ XLM vocabulary                            33.0
+ Pre-trained AMRs                         33.4
+ Pre-trained LMs                            33.8

https://github.com/jflanigan/jamr
Comparison: Monolingual vs. Multilingual

Monolingual

Des responsables américains ont tenu une réunion d'un groupe d'experts en janvier 2002 à New York.
Des responsables américains ont tenu une réunion d'un groupe d'experts en janvier 2002 à New York.
Results: Europarl

Monolingual: AMR -> X

Multilingual: AMR -> All

BLEU

en da de el es fi fr it nl pt sv bg cs et hu it lv pl ro sl sk

HRL

MRL
Résultats: LTC Data (Gold AMRs)

Monolingual: AMR -> X

<table>
<thead>
<tr>
<th>Language</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>en</td>
<td>25</td>
</tr>
<tr>
<td>es</td>
<td>22</td>
</tr>
<tr>
<td>it</td>
<td>19</td>
</tr>
<tr>
<td>de</td>
<td>16</td>
</tr>
</tbody>
</table>

Multilingual: AMR -> All

<table>
<thead>
<tr>
<th>Language</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>en</td>
<td>19</td>
</tr>
<tr>
<td>es</td>
<td>19</td>
</tr>
<tr>
<td>it</td>
<td>19</td>
</tr>
<tr>
<td>de</td>
<td>19</td>
</tr>
</tbody>
</table>
Comparison: NLG+Translation vs. End-to-End

Generation+Translation

Hold
:ARG0 person :ARG0-of have-org-role :ARG1 op1
United :op2 States :ARG2 official
:ARG1 meet :ARG0 person :ARG1-of expert :ARG2-of group
:time date-entity :year 2002 :month 1
:location city :op1 New :op2 York

US officials held an expert group meeting in January 2002 in New York.

English - X

Des responsables américains ont tenu une réunion d'un groupe d'experts en janvier 2002 à New York.

End-to-End

Hold
:ARG0 person :ARG0-of have-org-role :ARG1 op1
United :op2 States :ARG2 official
:ARG1 meet :ARG0 person :ARG1-of expert :ARG2-of group
:time date-entity :year 2002 :month 1
:location city :op1 New :op2 York

Des responsables américains ont tenu une réunion d’un groupe d’experts en janvier 2002 à New York.
Comparison: NLG+Translation vs. End-to-End

Comparison of BLEU scores for different languages:
- AMR -> En -> X
- AMR -> X

Languages: de, es, fr, it, nl, pt, sv

Scores for AMR -> En -> X:
- de: 19
- es: 21
- fr: 19
- it: 18
- nl: 19
- pt: 20
- sv: 25

Scores for AMR -> X:
- de: 17
- es: 19
- fr: 19
- it: 18
- nl: 20
- pt: 20
- sv: 25
# Impact of Training Languages

<table>
<thead>
<tr>
<th></th>
<th>Da</th>
<th>De</th>
<th>Ni</th>
<th>Sv</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Language</td>
<td>21.3</td>
<td>17.0</td>
<td>18.5</td>
<td>18.7</td>
</tr>
<tr>
<td><strong>ALL Germanic</strong></td>
<td>21.8</td>
<td>21.9</td>
<td>19.6</td>
<td>19.3</td>
</tr>
<tr>
<td>Languages <strong>21</strong></td>
<td>21.9</td>
<td>17.5</td>
<td>19.4</td>
<td>19.5</td>
</tr>
</tbody>
</table>
Impact of Training Languages

Bilingual Model: *Pairing with the closest language yields the best results*

Romance: Spanish/Portuguese
Germanic: Swedish/Danish
Uralic: Finnish/Estonian
Slavic: Czech/Slovak
Human-Based Evaluation

• Semantic:
  • Does the generated text convey the AMR meaning?

• Morphology:
  • Are agreement and inflection constraints respected?

• Word Order
  • Is the word order natural?
Semantic

2 means minor differences

2.4, 2.3, 2.2, 2.2, 2.1, 2.3, 2.3, 2.4, 2.3, 2.2, 2.3, 2.2, 2.1, 2.1, 1.6, 1.9
This point will certainly be the subject of subsequent further debates in the council.

This is a point that will undoubtedly be discussed later in the council.

Je ne suis pas favorable à des exceptions à cette règle.

A mon avis, il n’est pas bon de faire des exceptions à cette règle.
Human-Based Evaluation

Results are good across the board for morphology, semantics and word order.

A multilingual model generalises well to our set of target languages.
Des responsables américains ont tenu une réunion d'un groupe d'experts en janvier 2002 à New York.
Generating from Knowledge Graphs

Gardent et al. ACL 2017, Castro-Ferreira et al. 2020, Cripwell et al. 2023
W. Soto-Martinez, Y. Parmentier and C. Gardent AACL 2023
Alan Bean graduated from UT Austin in 1955 with a Bachelor of Science degree. He was hired by NASA in 1963 and served as a test pilot. Apollo 12’s backup pilot was Alfred Worden and was commanded by David Scott.
The WebNLG Shared Task

Gardent et al. ACL 2017, Castro-Ferreira et al. 2020, Cripwell et al. 2023
WebNLG 2017: RDF ==> English

<table>
<thead>
<tr>
<th></th>
<th>Train+Dev</th>
<th>Test (Seen Category)</th>
<th>Test (Unseen Category)</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td># (Graph,Text)</td>
<td>20,370</td>
<td>2,495</td>
<td>2,413</td>
<td>25,298</td>
</tr>
<tr>
<td># Graphs</td>
<td>7,812</td>
<td>971</td>
<td>891</td>
<td>9,674</td>
</tr>
</tbody>
</table>

- DBPedia Graphs with root entities from different DBPedia Categories
- Text is crowd-sourced (human written)

Gardent et al. ACL 2017
10 seen categories (seen at training time)
• Astronaut, University, Monument, Construction, Comics, Food, Airport, SportTeam, Town, WrittenWork

5 unseen categories
• CelestialBody, MeanOfTransportation, City, Athlete, Politician, Artist.

Gardent et al. ACL 2017
WebNLG 2017: RDF $\rightarrow$ English

6 participants, 10 systems
- 3 symbolic
- 1 statistical
- 5 neural

Gardent et al. ACL 2017
WebNLG 2017: RDF ==> English

All
7.07 - 45.13

Seen
19.87 - 60.54

Unseen
5.13 - 35.7
WebNLG 2020

Generation

• RDF ==> English
Generation

- RDF ==> English
- RDF ==> Russian

Castro-Ferreira et al. INLG 2020
WebNLG 2020

Generation
• RDF ==> English
• RDF ==> Russian

Semantic Parsing
• English ==> RDF
• Russian ==> RDF
### WebNLG 2020

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Dev</th>
<th>Test NLG/SP</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td># (Graph,Text)</td>
<td>35,426</td>
<td>4,664</td>
<td>5,150</td>
<td>47,395</td>
</tr>
<tr>
<td># Graphs</td>
<td>13,211</td>
<td>1,667</td>
<td>1,779</td>
<td>17,409</td>
</tr>
</tbody>
</table>

- 16 seen categories
- 3 unseen categories
- Unseen entities
# WebNLG 2020: Participation

<table>
<thead>
<tr>
<th>System</th>
<th>Affiliation</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>MED</td>
<td>Sber AI Lab</td>
<td>Russia</td>
</tr>
<tr>
<td>RALI-UMONTRÉAL</td>
<td>Université de Montréal</td>
<td>Canada</td>
</tr>
<tr>
<td>ORANGE-NLG</td>
<td>Orange Labs</td>
<td>France</td>
</tr>
<tr>
<td>CUNI-UFAL</td>
<td>Charles University</td>
<td>Czechia</td>
</tr>
<tr>
<td>TGen</td>
<td>AIST</td>
<td>Japan</td>
</tr>
<tr>
<td>BT5</td>
<td>Google</td>
<td>US</td>
</tr>
<tr>
<td>UPC-POE</td>
<td>Universitat Politècnica de Catalunya</td>
<td>Spain</td>
</tr>
<tr>
<td>DANGNT-SGU</td>
<td>Saigon University</td>
<td>Vietnam</td>
</tr>
<tr>
<td>HUAWEI</td>
<td>Huawei Noah’s Ark Lab</td>
<td>UK</td>
</tr>
<tr>
<td>AMAZONAI</td>
<td>Amazon AI (Shanghai)</td>
<td>China</td>
</tr>
<tr>
<td>NILC</td>
<td>University of São Paulo</td>
<td>Brazil</td>
</tr>
<tr>
<td>NUIG-DSI</td>
<td>National University of Ireland</td>
<td>Ireland</td>
</tr>
<tr>
<td>CYCLEGT</td>
<td>Amazon</td>
<td>China</td>
</tr>
<tr>
<td>OSU NEURAL NLG</td>
<td>The Ohio State University</td>
<td>US</td>
</tr>
<tr>
<td>FBConvAI</td>
<td>Facebook</td>
<td>US</td>
</tr>
</tbody>
</table>

17 participants
WebNLG 2020: Results

Min and Max BLEU Scores per Category of Test Data

- Russian
- Unseen Entity
- English (All)
- Unseen Category
- Seen Category

Min BLEU:
- Russian: 0.25
- Unseen Entity: 0.39
- English (All): 0.35
- Unseen Category: 0.23

Max BLEU:
- Russian: 0.51
- Unseen Entity: 0.61
- English (All): 0.54
- Unseen Category: 0.52

 BLEU
### WebNLG 2023: Low Resource Languages

<table>
<thead>
<tr>
<th>Language</th>
<th>Silver Train</th>
<th>Dev</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breton</td>
<td>13,211</td>
<td>1,399</td>
<td>1,778</td>
</tr>
<tr>
<td>Welsh</td>
<td>13,211</td>
<td>1,665</td>
<td>1,778</td>
</tr>
<tr>
<td>Irish</td>
<td>13,211</td>
<td>1,665</td>
<td>1,778</td>
</tr>
<tr>
<td>Maltese</td>
<td>13,211</td>
<td>1,665</td>
<td>1,778</td>
</tr>
</tbody>
</table>
### WebNLG 2023: Generation + Translation

<table>
<thead>
<tr>
<th>Team</th>
<th>Affiliation</th>
<th>Country</th>
<th>Breton</th>
<th>Welsh</th>
<th>Irish</th>
<th>Maltese</th>
<th>Russian</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUNI-Wue</td>
<td>Charles University</td>
<td>Czechia</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DCU/TCD-FORGe</td>
<td>ADAPT/DCU/Trinity College</td>
<td>Ireland</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Interno</td>
<td>Pulkovo Observatory</td>
<td>Russia</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>IREL</td>
<td>IIT Hyderabad</td>
<td>India</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DCU-NLG-PBN</td>
<td>ADAPT/DCU</td>
<td>Ireland</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
</tbody>
</table>

**RDF ==> English**
- T5 or mT5 fine-tuned on English WebNLG data
- GPT3-5, in context learning

**English ==> LRL**
- Machine Translation: NLLB or Google Translate
WebNLG 2023 : Résultats

![Min and Max BLEU Scores per Language](image)

- English 2020
- Irish
- Maltese
- Welsh
- Breton
**NLG+MT Pipeline vs. End-to-End**

**NLG+MT**
- Only possible for Languages with MT models
- For Breton, Machine Translation is very poorly

❌ Generation + Translation
**NLG+MT Pipeline vs. End-to-End**

**NLG+MT**
- Only possible for Languages with MT models
- For Breton, Machine Translation is very poor

**Generation + Translation**

**Fine-Tuning mT5 on WebNLG Dev data**
(BLEU : 0.10)
**NLG+MT Pipeline vs. End-to-End**

**NLG+MT**
- Only possible for Languages with MT models
- For Breton, Machine Translation is very pool

- Generation + Translation
- Fine-Tuning (BLEU : 0.10)

- Soft-Prompt Fine-tuning
WebNLG 2023: Soft Prompt Fine Tuning

- Fine tune mT5 on WebNLG Dev set
- Structured soft-prompt
WebNLG 2023: Structured Soft Prompt

Soft Prompt

<table>
<thead>
<tr>
<th>50 Tokens Task</th>
<th>15 Tokens Source Family</th>
<th>15 Tokens Source Genus</th>
<th>15 Tokens Source Language</th>
<th>15 Tokens Target Family</th>
<th>15 Tokens Target Genus</th>
<th>15 Tokens Target Language</th>
<th>n Tokens Input Sequence</th>
</tr>
</thead>
</table>
## Pre-training

### Step 1. Self-supervised pre-training (Language Models)

**Monolingual data**

<table>
<thead>
<tr>
<th>Task</th>
<th>Source</th>
<th>Target</th>
<th>Original Input Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masked LM</td>
<td>Germanic</td>
<td>West Germanic</td>
<td>RDF</td>
</tr>
<tr>
<td>Prefix LM</td>
<td>Germanic</td>
<td>West Germanic</td>
<td>English</td>
</tr>
<tr>
<td>Suffix LM</td>
<td>Celtc</td>
<td>Britonic</td>
<td>Welsh</td>
</tr>
<tr>
<td>Deshuffling</td>
<td>Celtc</td>
<td>Britonic</td>
<td>Breton</td>
</tr>
<tr>
<td>Generate</td>
<td>Celtc</td>
<td>Goidelic</td>
<td>Irish</td>
</tr>
</tbody>
</table>

- **Masked LM**: Masked Language Model
- **Prefix LM**: Prefix Language Model
- **Suffix LM**: Suffix Language Model
- **Deshuffling**: Deshuffling Language Model
- **Generate**: Generate Language Model
## Pre-training

### Step 1. Self supervised Pre-training (Language Models)

**Monolingual Data**

<table>
<thead>
<tr>
<th>Task</th>
<th>Source</th>
<th>Target</th>
<th>Original Input Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masked LM</td>
<td>Germanic</td>
<td>West Germanic</td>
<td>RDF</td>
</tr>
<tr>
<td>Prefix LM</td>
<td>Germanic</td>
<td>West Germanic</td>
<td>English</td>
</tr>
<tr>
<td>Suffix LM</td>
<td>Celtic</td>
<td>Britonic</td>
<td>Welsh</td>
</tr>
<tr>
<td>Deshuffling</td>
<td>Celtc</td>
<td>Britonic</td>
<td>Breton</td>
</tr>
<tr>
<td>Generate</td>
<td>Celtc</td>
<td>Goidelic</td>
<td>Irish</td>
</tr>
</tbody>
</table>

### Step 2. Fine-tuning on RDF-Text data
Pre-training

**Step 1.** Self supervised Pre-training (Language Models)

*Monolingual Data*

<table>
<thead>
<tr>
<th>Task</th>
<th>Source</th>
<th>Target</th>
<th>Original Input Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masked LM</td>
<td>Germanic</td>
<td>RDF</td>
<td>&lt;S&gt; Einstein &lt;P&gt; &lt;mask&gt; &lt;P&gt; Poland</td>
</tr>
<tr>
<td>Prefix LM</td>
<td>Germanic</td>
<td>English</td>
<td>Thank you for &lt;mask&gt; &lt;pad&gt; &lt;pad&gt;</td>
</tr>
<tr>
<td>Suffix LM</td>
<td>Celtic</td>
<td>Welsh</td>
<td>&lt;mask&gt; honno ? &lt;pad&gt; &lt;pad&gt; &lt;pad&gt;</td>
</tr>
<tr>
<td>Deshuffling</td>
<td>Celtc</td>
<td>Breton</td>
<td>skuizh ? out Ha &lt;pad&gt; &lt;pad&gt; &lt;pad&gt;</td>
</tr>
<tr>
<td>Generate</td>
<td>Celtc</td>
<td>Irish</td>
<td>Seo &lt;mask&gt; &lt;pad&gt; &lt;pad&gt; &lt;pad&gt; &lt;pad&gt;</td>
</tr>
</tbody>
</table>

**Step 2.** Fine-tuning on RDF-Text data

**Inference.** Using the target language soft prompt
WebNLG 2023 : Résultats

![Bar chart showing BLEU scores for Breton, Welsh, and Irish languages. The chart compares Control Prefixes, Vanilla Full Fine-Tuning, and PI-TST models.]
Key Takeaways

• **Pre-Training:** 2017 vs 2020, Pre-training improve results

• **Target Language:** Results are better for English than for Russian

• **Generalisation:** Models underperform on unseen data

• **LRL:** results are poor for LRL

• **LRL Models:** Pipeline models are used but depends on having MT models for the LRL and the output quality largely depends on the quality of the models

• **PEFT:** Soft-prompt fine tuning helps improve results
  
  
  **BLEU Breton:** 10 (NLG+MT) $\rightarrow$ 18.15 (PEFT)
Generation from Knowledge Graphs remains limited

- Poor performance on unseen data
- Few languages
- There is a need for a multilingual graph/text similarity metrics
  - To filter noisy parallel data
  - To guide generation
  - To generalise to other languages and other domains
Generation into LRL is an open problem

- Lack of data
- Pipeline models and fine tuning yield poor results
Generating from Text
Generating Wikipedia Biographies

Angela Fan and Claire Gardent
“Generating Full Length Wikipedia Biographies. The Impact of Gender Bias on the Retrieval-Based Generation of Women Biographies.”
ACL 2022
Generating Wikipedia Biographie

Retrieval Augmented Generation (RAG)
Generation + Information Retrieval

WIKIPEDIA

Joan Paton

Joan Burton Paton AM née Cleland (1916–April 2000) was an Australian teacher, naturalist, environmentalist and ornithologist. One of the first women to become a member of the exclusive Adelaide Ornithologists Club, of which she was elected President 1991–1993, she also served as president of the South Australian Ornithological Association (1979–1982). Her father was Professor Sir John Burton Cleland, a notable microscopist and pathologist who strongly encouraged her early interest in natural history.

Contents
Early life and education
Career
Legacy and honours
References
External References

Early life and education

Joan Burton Paton was born in Sydney, New South Wales, the daughter of John Burton Cleland (1878–1971) and his wife, Dora Isabel Paton (1889–1955).[1] She had three sisters, Dr Margaret Burton Cleland, Elizabeth Robson Cleland and Barbara Burton Cleland; and a brother, William Paton 'Biff' Cleland, who became a surgeon. The father encouraged his children's interest in science. Joan Paton was educated at the University of Adelaide, where she majored in organic chemistry and biochemistry. In 1951 she married Ernest Norman Paton (1922–1985), son of Adolph Ernest Paton and Isla Marie Poynton. Their son is Prof David Cleland Paton.[2]

Career

In 1967 Paton became a lecturer on ornithology in South Australia's Workers Educational Association.[3] Among those she inspired to work in ornithology and environmental conservation was Margaret Cameron, who became the President of the Royal Australian Ornithologists Union (RAOU).[4]

Paton was active in the RAOU, as well as in the South Australian Ornithological Association (SAAO), of which she was elected Vice-President 1974–1978, and President 1979–1982. She was one of the first women to become a member of the exclusive Adelaide Ornithologists Club, of which she was elected president (1991–1993).[5]

Legacy and honours

- 1990: she was made an Honorary Member of the SAAO.
- 1996: she was made an Honorary Member of the Adelaide Ornithologists Club.
Challenges

- Gather relevant evidence
- Generate structured Text
- Factuality
BIOGRAPHY OF KATHERINE JOHNSON

INTRO PARAGRAPH

EARLY LIFE

CAREER

INPUT WEB EVIDENCE

DOC 1
What Was Katherine Johnson's Early Life Like?

As a young girl, Katherine loved to count.

She counted everything.

She would count the number of steps she took to the road.

She counted the steps into church.

DOC 2

RETRIEVAL MODULE

RETRIEVAL OUTPUT

What Was Katherine Johnson's Early Life Like?

She counted everything.

QUERY

SUBJECT: Katherine Johnson, Mathematician, Early Life

OCCUPATION

SECTION

GENERATION ENCODER

CACHE: PREVIOUS SECTIONS

GENERATION DECODER

CITATION MODULE

Katherine Johnson was born as Grzelda Katherine Coleman on August 26, 1918, in White Sulphur Springs, West Virginia, to Joylette Roberts (Lowe) and Joshua McKinley Coleman. She was the youngest of four children. Johnson showed strong math abilities from an early age, CAREER [1][2]

BACKPROP THROUGH RETRIEVAL MODULE

75
Extraction

QUERY
Katherine Johnson
Mathematician
Early Life
Katherine Johnson
Mathematician
Early Life

20 first documents segmented in sentences
Extraction

Requête

Katherine Johnson
Mathematician
Early Life

Web Search
20 first documents segmented in sentences

Semantic Filtering
40 most similar sentences (1,000 words)
Katherine Johnson
Mathematician
Early Life

Katherine Johnson was born as Creola Katherine Coleman on August 26, 1918, in White Sulphur Springs, West Virginia, to Joylette Roberta (Lowe) and Joshua McKinley Coleman. She was the youngest of four children. Johnson showed strong math abilities from an early age. CAREER [1][2]
Cache Transformer-XL

Each section predicts the next, to write a full biography

- The hidden states of preceding sections are stored
- Used as memory to generate the next section
Ablation

<table>
<thead>
<tr>
<th>Model</th>
<th>ROUGE-L</th>
<th>Entailment</th>
<th>Named Entity Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>BART Pretraining + Finetuning</td>
<td>17.4</td>
<td>15.8</td>
<td>21.9</td>
</tr>
<tr>
<td>+ Retrieval Module</td>
<td>18.8</td>
<td>17.2</td>
<td>23.1</td>
</tr>
<tr>
<td>+ Caching Mechanism</td>
<td>19.3</td>
<td>17.9</td>
<td>23.4</td>
</tr>
</tbody>
</table>

IR and the cache mechanism allow for statistically significant improvement.
Evaluation par l’humain de la factualité

- 71% of information in the reference text is not in the generated text.
Evaluation par l’humain de la factualité

- 71% of information in the reference text is not in the generated text.
- 68% of the information in the generated sections is not present in the reference text.
Evaluation par l’humain de la factualité

- 71% of information in the reference text is not in the generated text.
- 68% of the information in the generated sections is not present in the reference text.
- 17% of the added information is validated by examining the web evidence — some information added by the generative model is valid.
## Data Bias

### Wikisum: Wikipedia biographies

### Our dataset: Woman Wikipedia biographies

<table>
<thead>
<tr>
<th></th>
<th>WikiSum Evaluation Dataset</th>
<th>Our Evaluation Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Number of Sections</td>
<td>7.2</td>
<td>5.8</td>
</tr>
<tr>
<td>Average Length of a Section</td>
<td>151.0</td>
<td>132.3</td>
</tr>
<tr>
<td>Average Length of Total Article</td>
<td>892.3</td>
<td>765.9</td>
</tr>
<tr>
<td>Avg overlap of Web Hits and Biography</td>
<td>39.8%</td>
<td>24.9%</td>
</tr>
</tbody>
</table>

- **WikiSum Evaluation Dataset**
- **Our Evaluation Dataset**
**Biais dans les données Data Bias**

- Woman Biographies are shorter
  - Fewer sections
  - Shower sections
  - Fewer words

<table>
<thead>
<tr>
<th>WikiSum Evaluation Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Number of Sections</td>
</tr>
<tr>
<td>Average Length of a Section</td>
</tr>
<tr>
<td>Average Length of Total Article</td>
</tr>
<tr>
<td>Avg overlap of Web Hits and Biography</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Our Evaluation Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Number of Sections</td>
</tr>
<tr>
<td>Average Length of a Section</td>
</tr>
<tr>
<td>Average Length of Total Article</td>
</tr>
<tr>
<td>Avg Number of Web Hits (max 20)</td>
</tr>
<tr>
<td>Avg overlap of Web Hits and Biography</td>
</tr>
</tbody>
</table>
**Data Bias**

Less Web evidence for women

- Fewer documents: 18 in average

<table>
<thead>
<tr>
<th>WikiSum Evaluation Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Number of Sections</td>
</tr>
<tr>
<td>Average Length of a Section</td>
</tr>
<tr>
<td>Average Length of Total Article</td>
</tr>
<tr>
<td>Avg overlap of Web Hits and Biography</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Our Evaluation Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Number of Sections</td>
</tr>
<tr>
<td>Average Length of a Section</td>
</tr>
<tr>
<td>Average Length of Total Article</td>
</tr>
<tr>
<td>Avg Number of Web Hits (max 20)</td>
</tr>
<tr>
<td>Avg overlap of Web Hits and Biography</td>
</tr>
</tbody>
</table>
## Data Bias

Less web evidence for woman

Smaller overlap between web evidence and Wikipedia biographies

<table>
<thead>
<tr>
<th>WikiSum Evaluation Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Number of Sections</td>
</tr>
<tr>
<td>Average Length of a Section</td>
</tr>
<tr>
<td>Average Length of Total Article</td>
</tr>
<tr>
<td>Avg overlap of Web Hits and Biography</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Our Evaluation Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Number of Sections</td>
</tr>
<tr>
<td>Average Length of a Section</td>
</tr>
<tr>
<td>Average Length of Total Article</td>
</tr>
<tr>
<td>Avg Number of Web Hits (max 20)</td>
</tr>
<tr>
<td>Avg overlap of Web Hits and Biography</td>
</tr>
</tbody>
</table>
Fewer Evidence, Lower Quality Output

<table>
<thead>
<tr>
<th>Model</th>
<th>WikiSum Test</th>
<th>Women</th>
<th>Scientists</th>
<th>Women in Asia</th>
<th>Women in Africa</th>
</tr>
</thead>
<tbody>
<tr>
<td>BART Pretraining</td>
<td>19.0</td>
<td>17.4</td>
<td>18.2</td>
<td>16.7</td>
<td>16.4</td>
</tr>
<tr>
<td>+ Retrieval</td>
<td>21.4</td>
<td>18.8</td>
<td>19.3</td>
<td>17.9</td>
<td>17.1</td>
</tr>
<tr>
<td>+ Caching</td>
<td>21.8</td>
<td>19.3</td>
<td>19.7</td>
<td>18.4</td>
<td>17.3</td>
</tr>
</tbody>
</table>

- Performance decreases for woman biographies particularly, Asian and African
Example Output

toplevel ashley mckenzie is a canadian film director, screenwriter and producer. she is the winner of the stella artois jay scott prize for emerging talent at the 2016 toronto international film festival. her first feature film, werewolf, premiered at the toronto film festival in 2016. she has also directed short films for the national film board of canada and the canadian screen actors guild. she was born in montreal, quebec, canada, and grew up in ottawa, ontario. [1,3,11,13,14]

=personal life= mckenzie was born in london, england. she is the daughter of alexander mckenzie, who was a member of the british rock band the beatles. she has a younger sister, jessica, who is also a singer. she was educated at st mary ’s college, oxford, where she graduated with a bachelor of arts degree in english literature. she also studied at the university of london. she married fellow x factor contestant andrew davies in september 2006. they have two children, a son and a daughter. [3,4,7,8,10,11]

career mckenzie was a contestant on the third series of the x - factor in 2006. she was eliminated in the first week of the competition. in 2007, mckenzie released her debut single "don ’t pretend you hadn’ t, now..." which peaked at no .160; 2 on the uk singles chart. she also released a second single ," i ’m not afraid ", in 2008. in 2009, she released her third single ," don’ t pretend you haven ’t, now ". in 2010, she was a judge on the x factor uk. [2]
ChatGPT does not know Ashley McKenzie, winner of the 2016 Toronto international film festival.
Henri Poincaré (29 April 1854 – 17 July 1912) was a French mathematician, theoretical physicist, and philosopher of science. Renowned for his exceptional contributions to mathematics and theoretical physics, Poincaré made groundbreaking advances in various fields, earning him a place among the most influential intellectuals of his time.

**Early Life and Education:**
Henri Poincaré was born in Nancy, France, into an academic family. His father, Léon Poincaré, was a professor of medicine, and his mother, Eugénie Launois, came from a family of distinguished scholars. Henri showed an early aptitude for mathematics, displaying remarkable talent during his formative years. He attended the Lycée Impérial in Nancy and later pursued higher education at the École Polytechnique in Paris.

**Mathematical Contributions:**
Poincaré's work in mathematics covered a broad spectrum of topics, including celestial mechanics, algebraic topology, number theory, and the three-body problem in physics. He made significant advancements in the field of algebraic topology, introducing the concept of homology, which laid the foundation for later developments in the understanding of space and shape.
Human evaluation shows that multilingual techniques generalize across languages
Human evaluation shows that multilingual techniques generalize across languages

Multilingual models benefit from increased training data and performs better than monilingual ones
Human evaluation shows that multilingual techniques generalize across languages.

Multilingual models benefit from increased training data and perform better than monilingual ones.

Using English-Centric AMR, we can decode into languages with varied syntax and morphology.
Conclusion
Thanks !