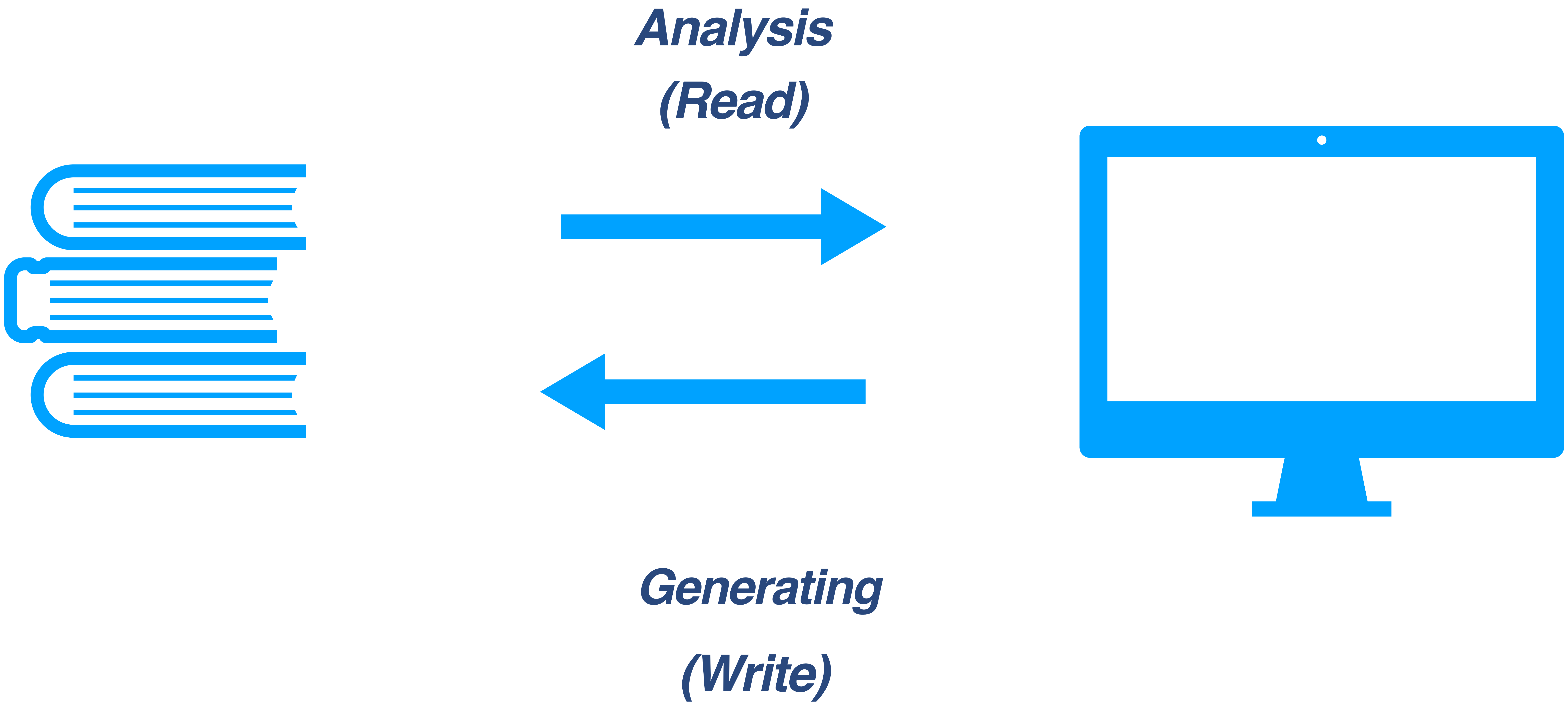
Generating Text from various sources into multiple languages

Claire Gardent
CNRS/LORIA

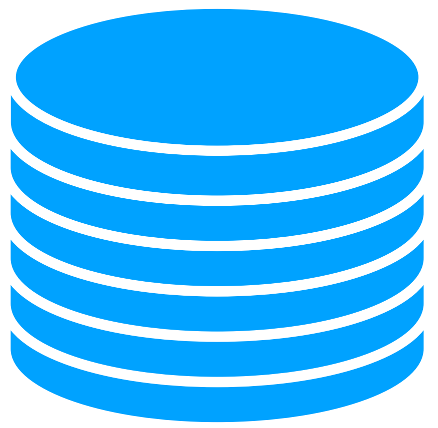
xNLG AI Chair “*Generating from Multiple Sources into Multiple Languages*”

Natural Language Processing

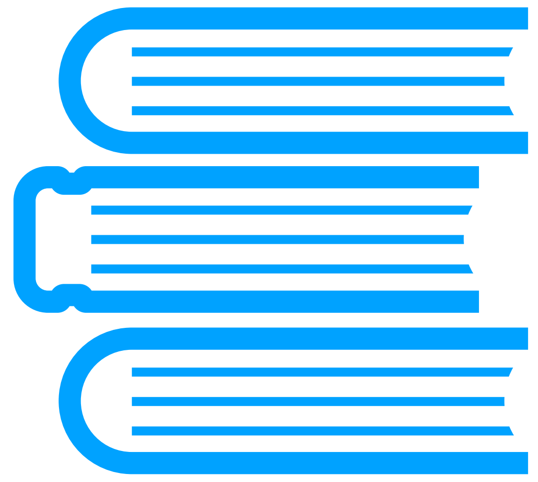


Generating Text

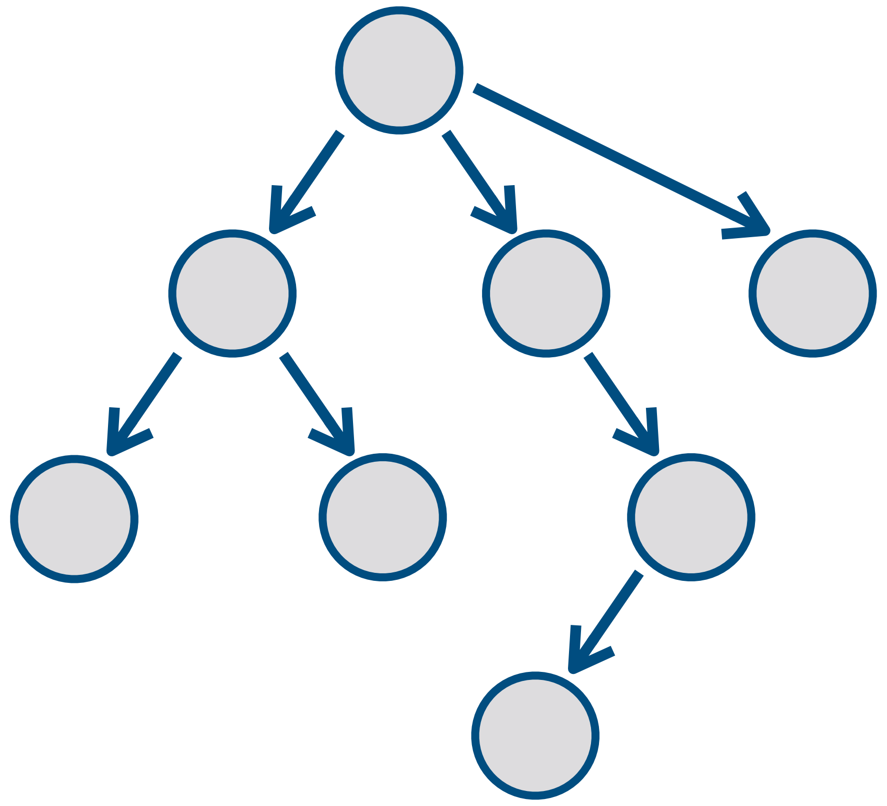
Inputs



Tabular Data, Data Bases, Knowledge Bases



Text



Graph

Generating Text

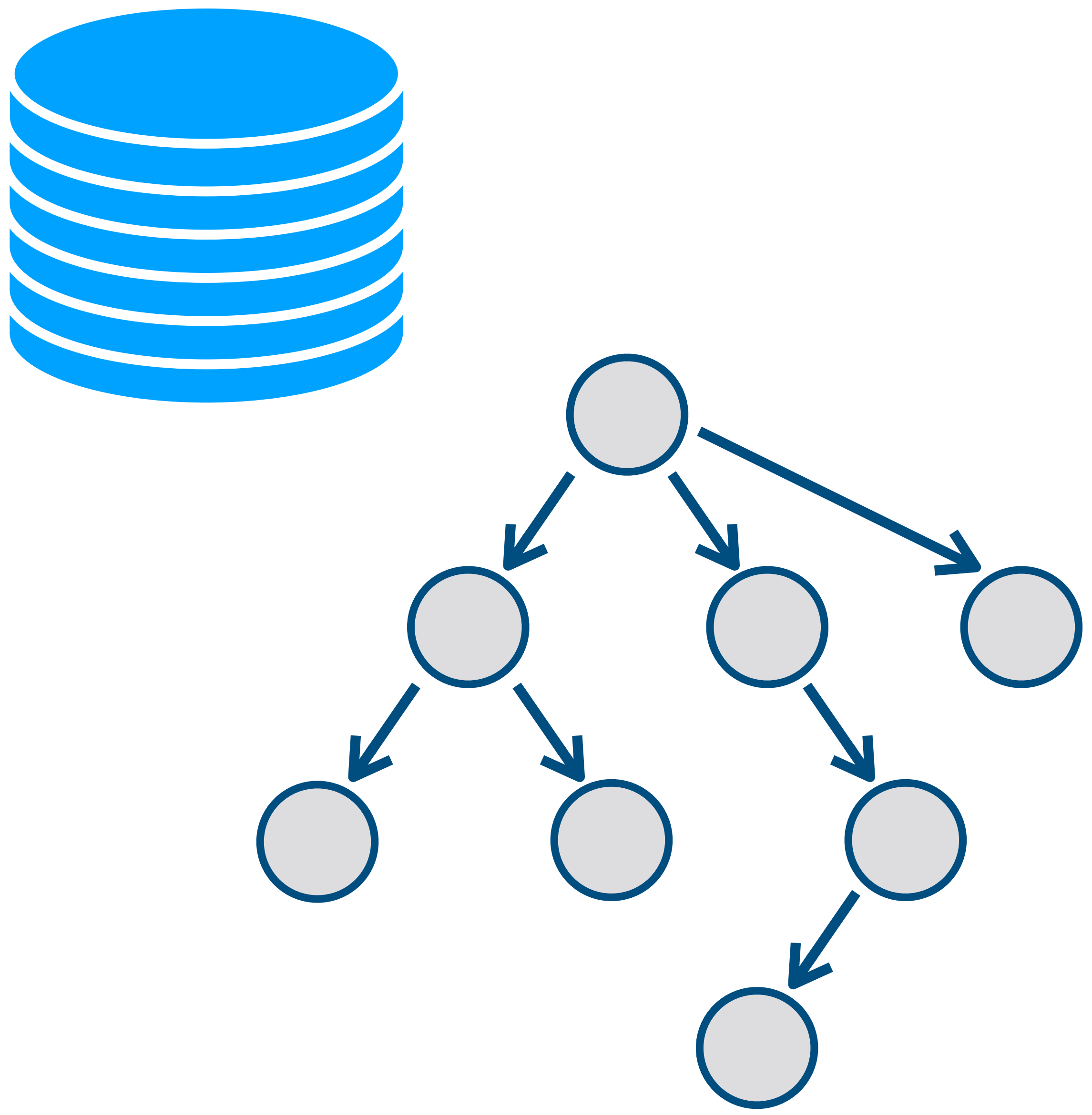
And also …

Images

Videos

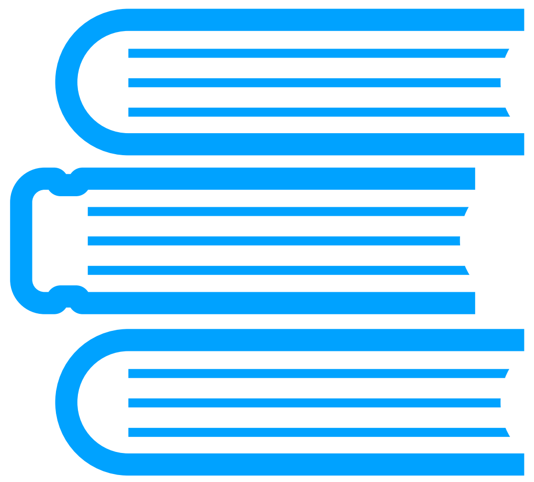
Generating Text

What for ?



Verbalising

- A graph
- A Data/Knowledge Base

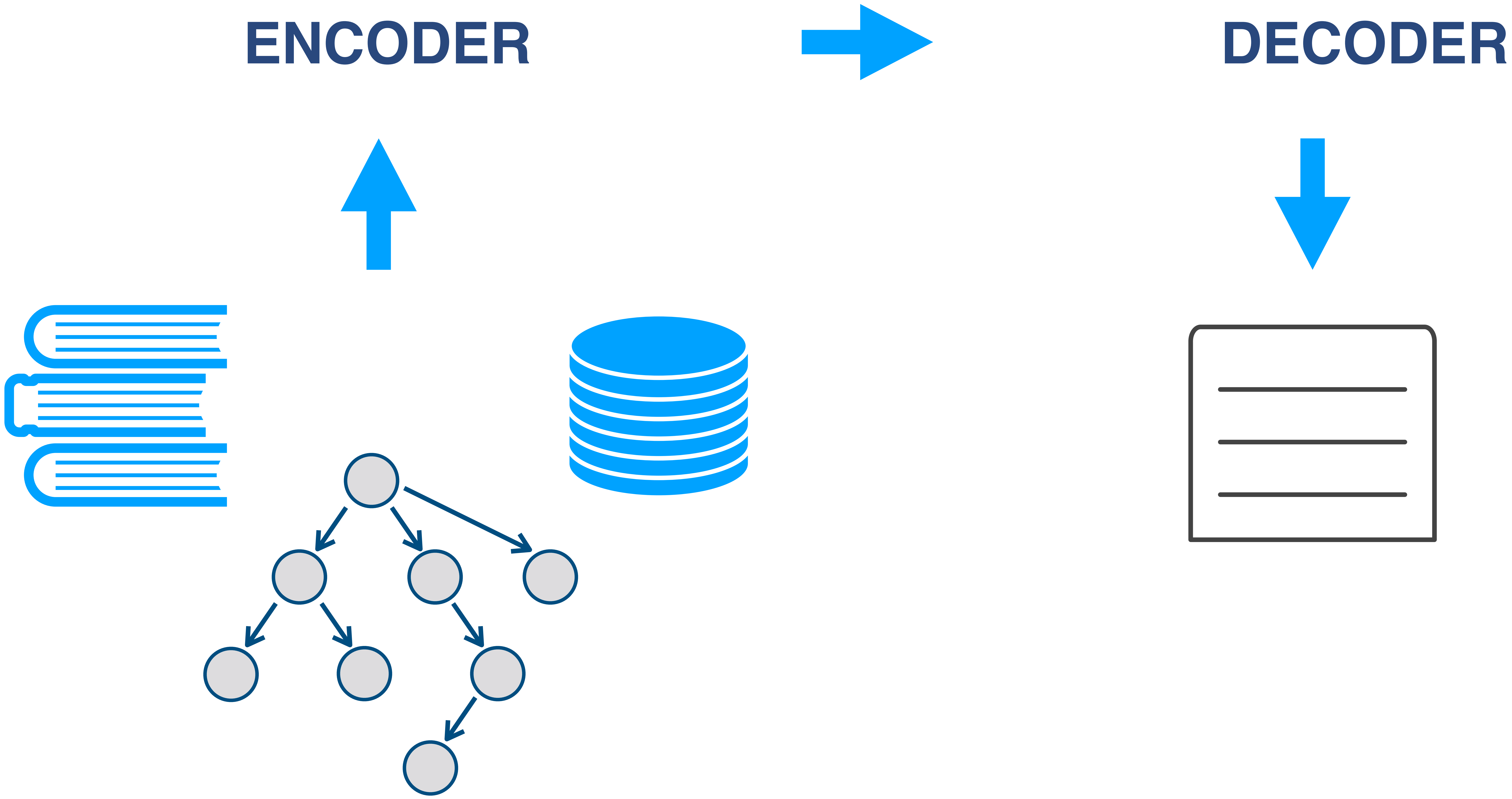


Simplify, paraphrase, summarise

One or more text(s)

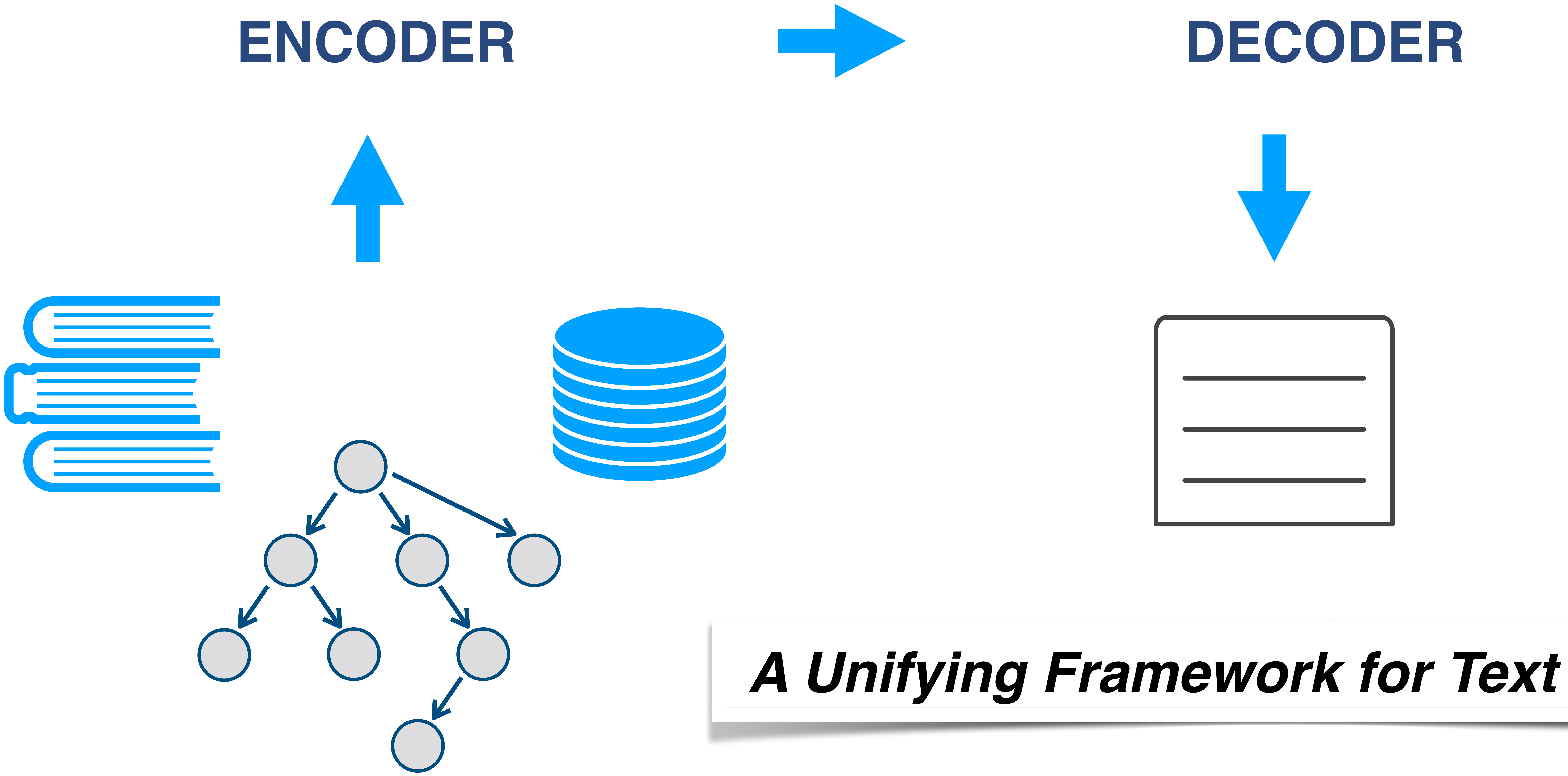
Conditional Text Generation

Input -> Text



Conditional Text Generation

Input -> Text



A Unifying Framework for Text Generation

Conditional Text Generation from …

- Meaning Representations

  *Abstract Meaning Representations (AMR) -> **21** EU languages*

- Knowledge Graphs

  *Resource Description Format (RDF) -> **Low Resource Languages** (Breton, Welsh, Irish)*

- Texts

  *- Generating Wikipedia Biographies (**Data Bias)***

  - Simplifying Documents

Evaluating Generated Text

- Sentence Simplification

  - *Sentence Level Estimate (SLE)*

- KB-to-Text

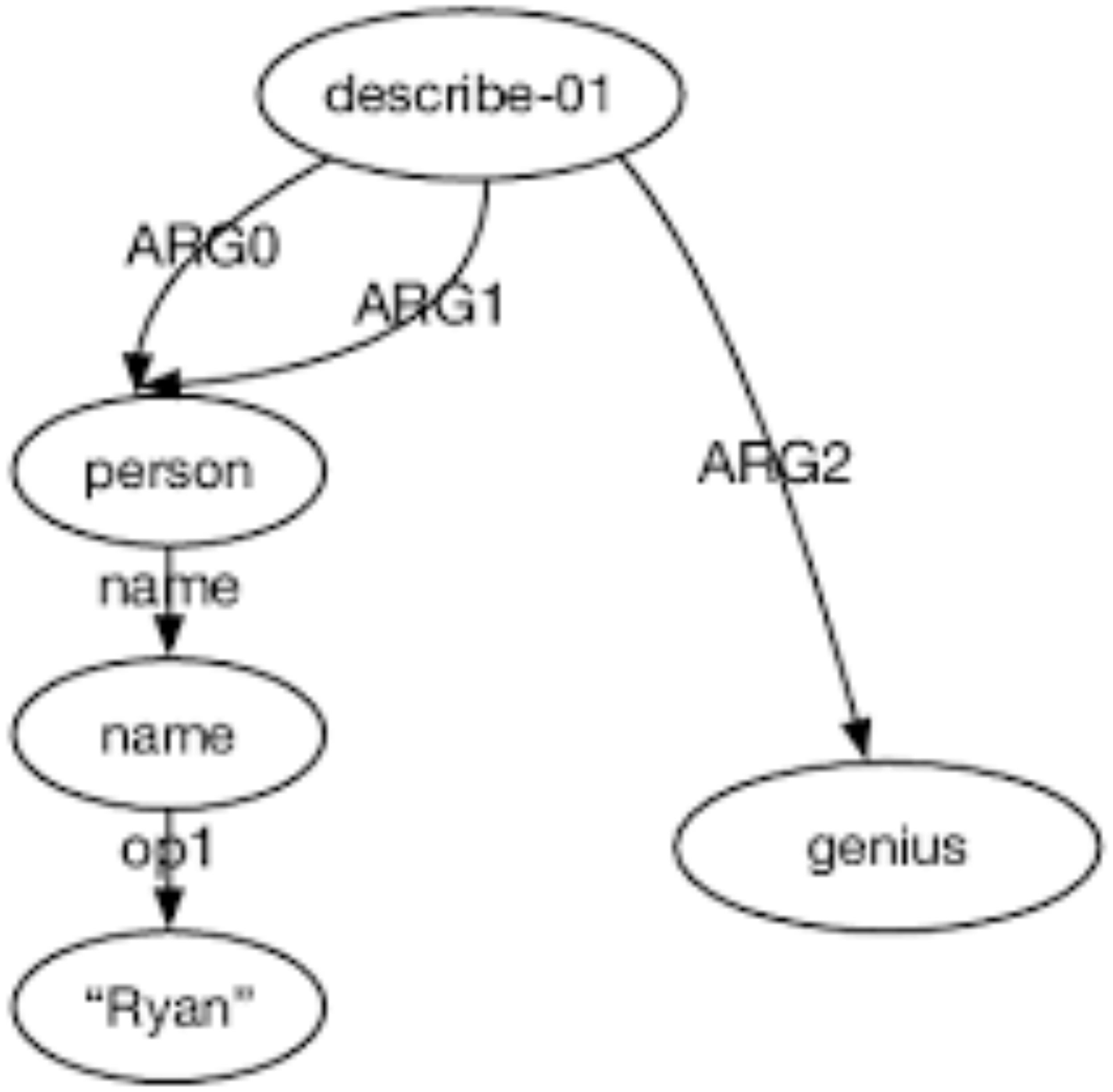
  - *Entity-Based Semantic Adequacy (ESA)*
  - *A Semantic Approach (EREDAT)*

Verbalising Abstract Meaning Representations into 21 EU Languages

Angela Fan and Claire Gardent

“Multilingual AMR-to-Text Generation”
EMNLP 2020.

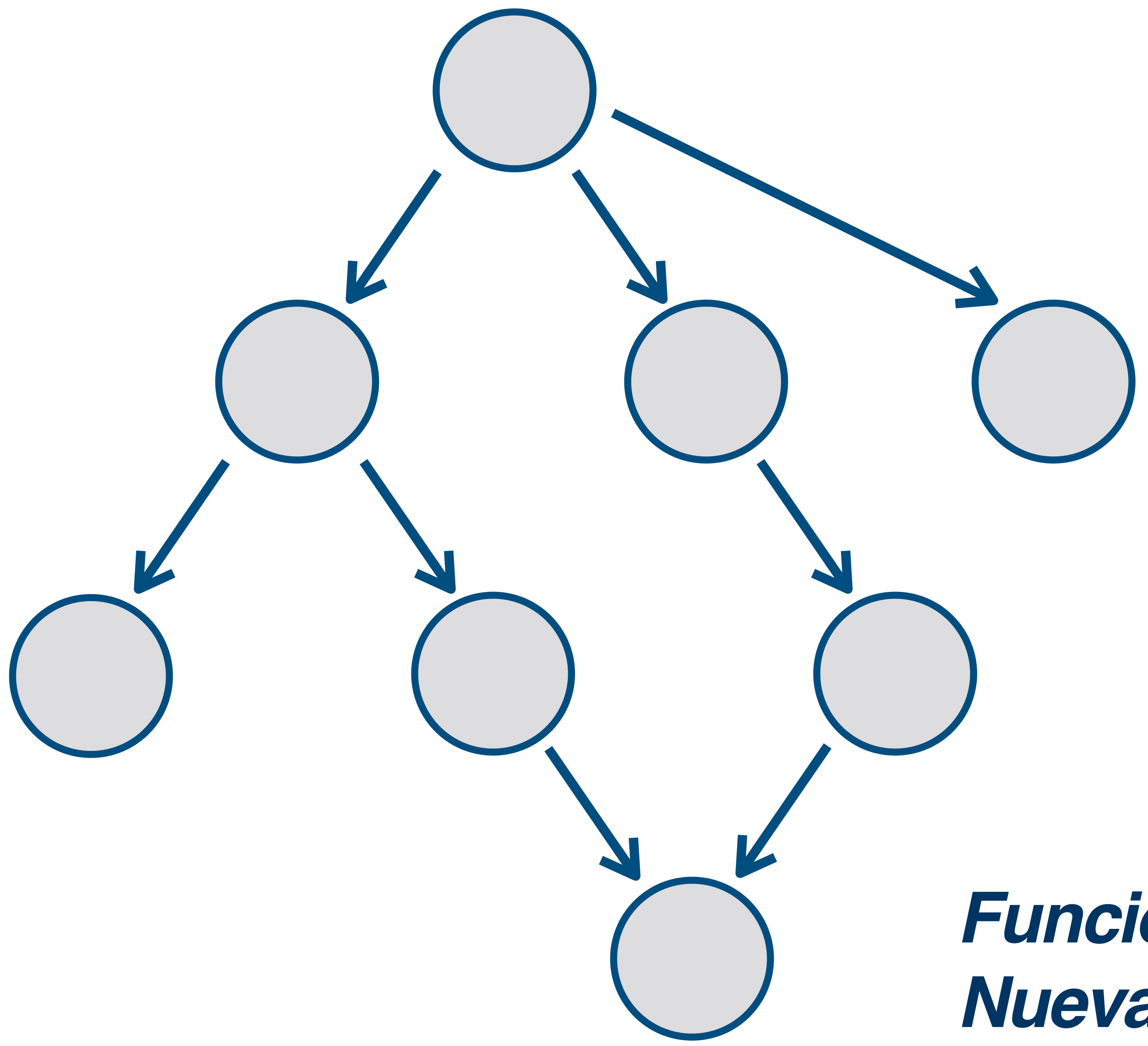
AMR Graph (Abstract Meaning Representation)



Ryan describes himself as a genius

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

AMR Graph —> 21 Languages



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

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

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.

Roman, Germanic, Slavic, Uralic

Challenges

- Graph —> Sequence

Challenges

- Graph —> Sequence
- Lack of Training/Test Data

Challenges

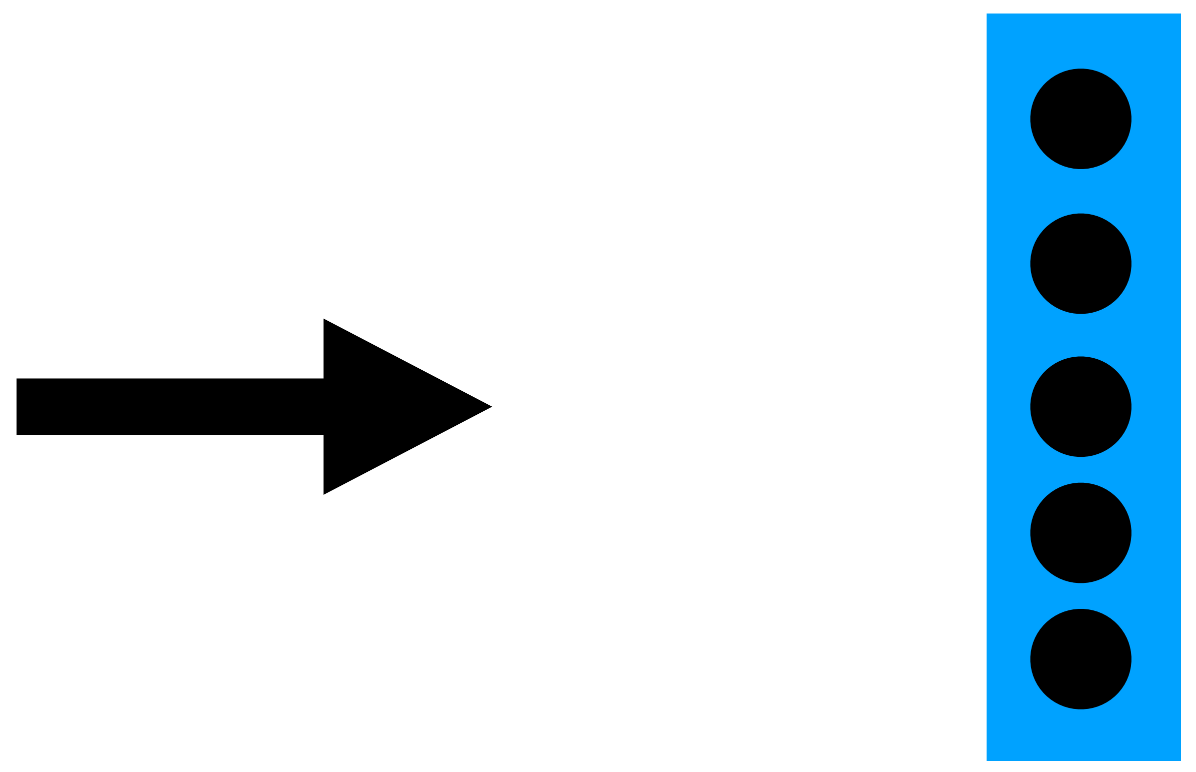
- Graph —> Sequence
- Lack of Training/Test Data
- Underspecified Input

Challenges

- Graph —> Sequence
- Lack of Training/Test Data
- Underspecified Input
- Multilingual : Generate texts with varied syntax and morphology

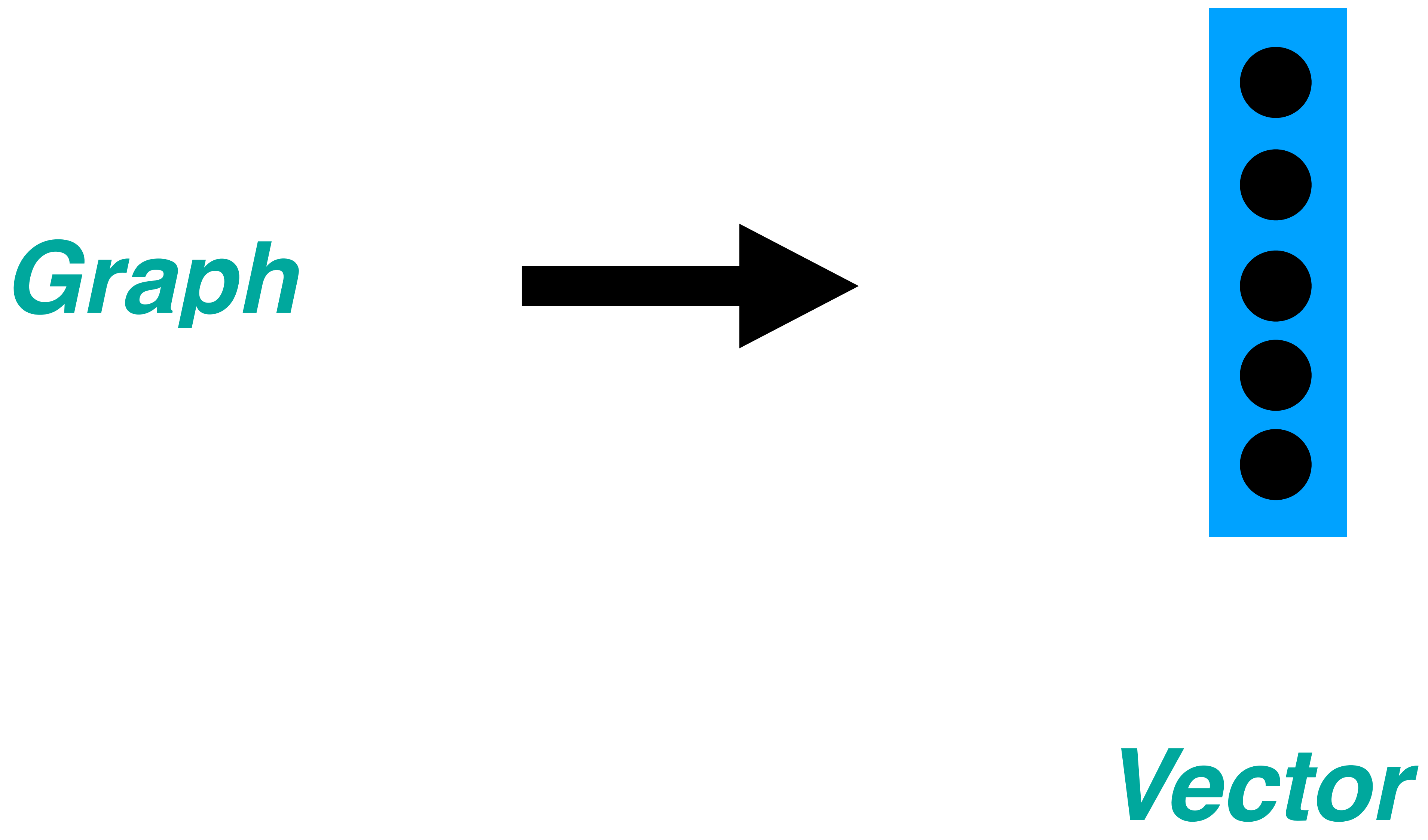
Encoder

Graph



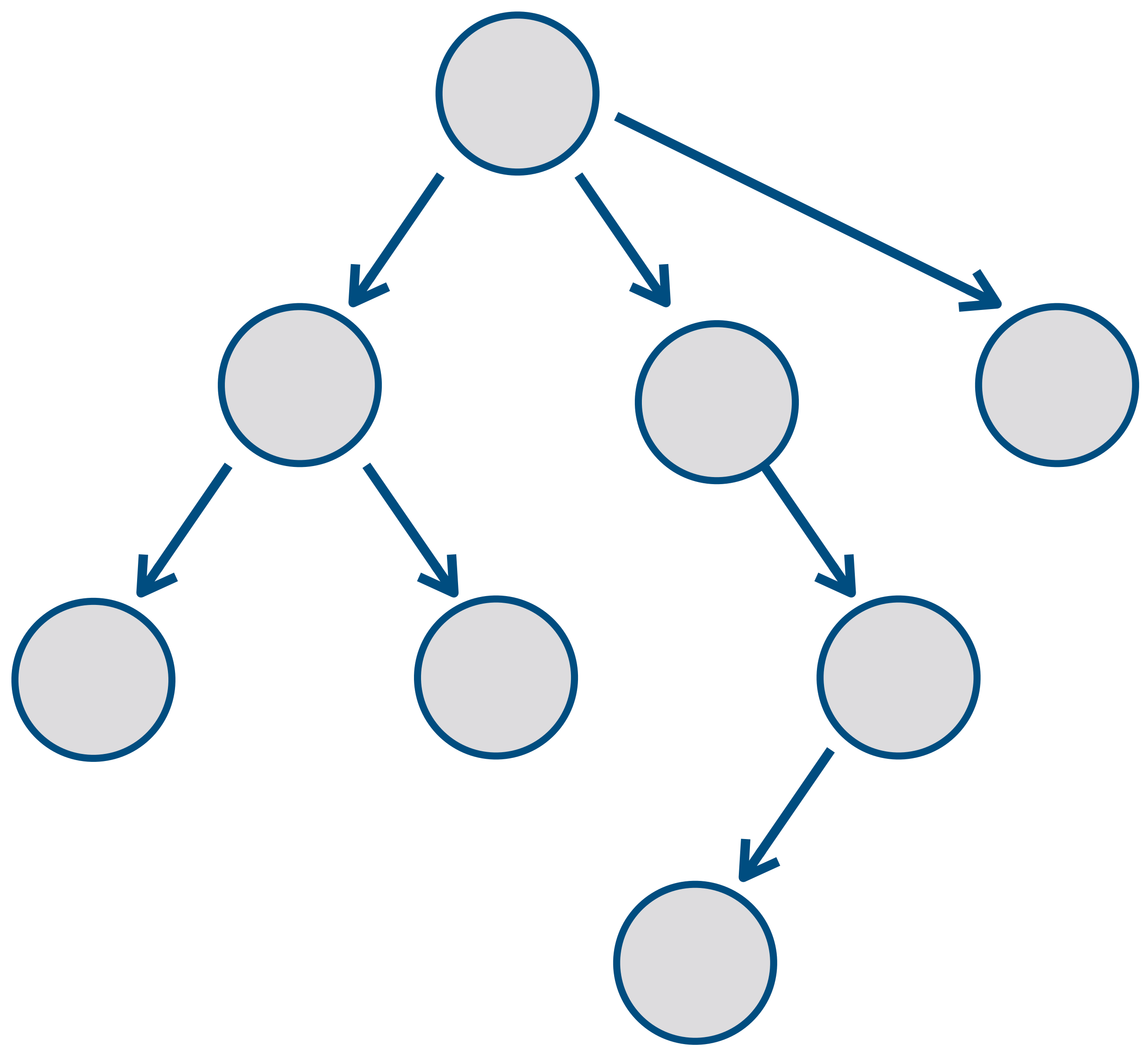
Vector

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

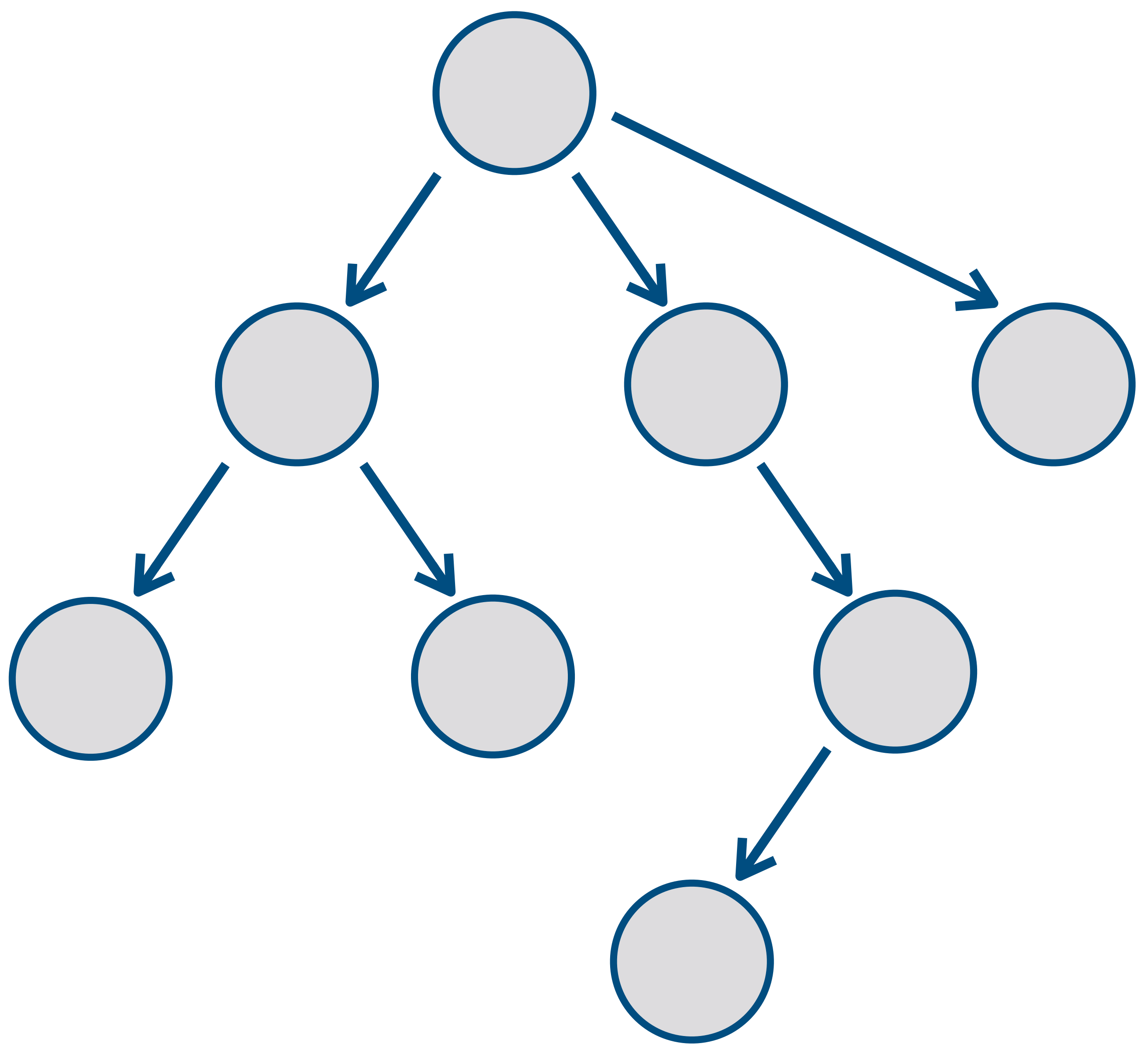
Structural Embeddings for Siblings and Branches

Subwords

- Remove variables and instance-of relation

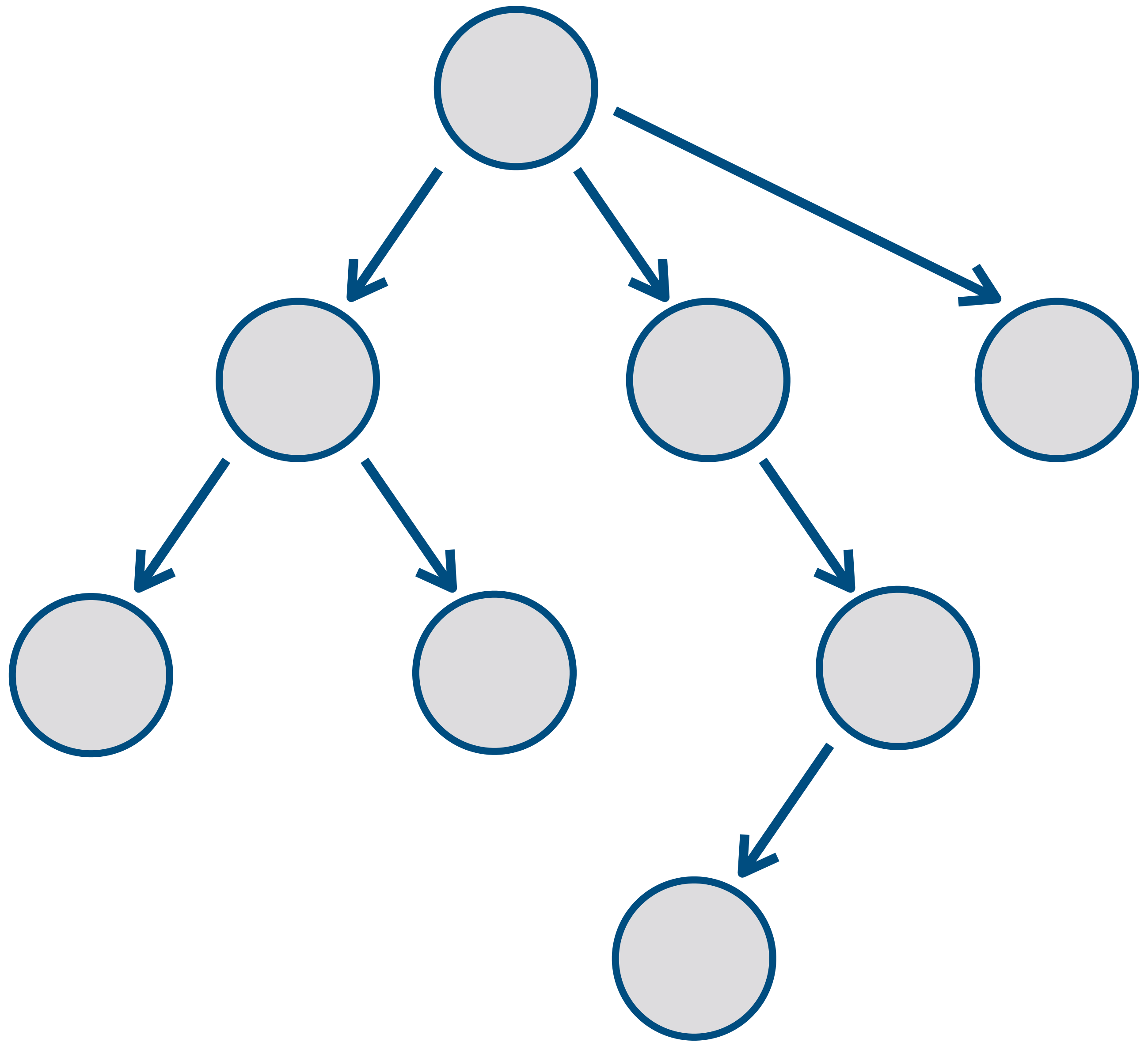
- No delexicalisation

- Sentence piece Vocabulary 32K operations

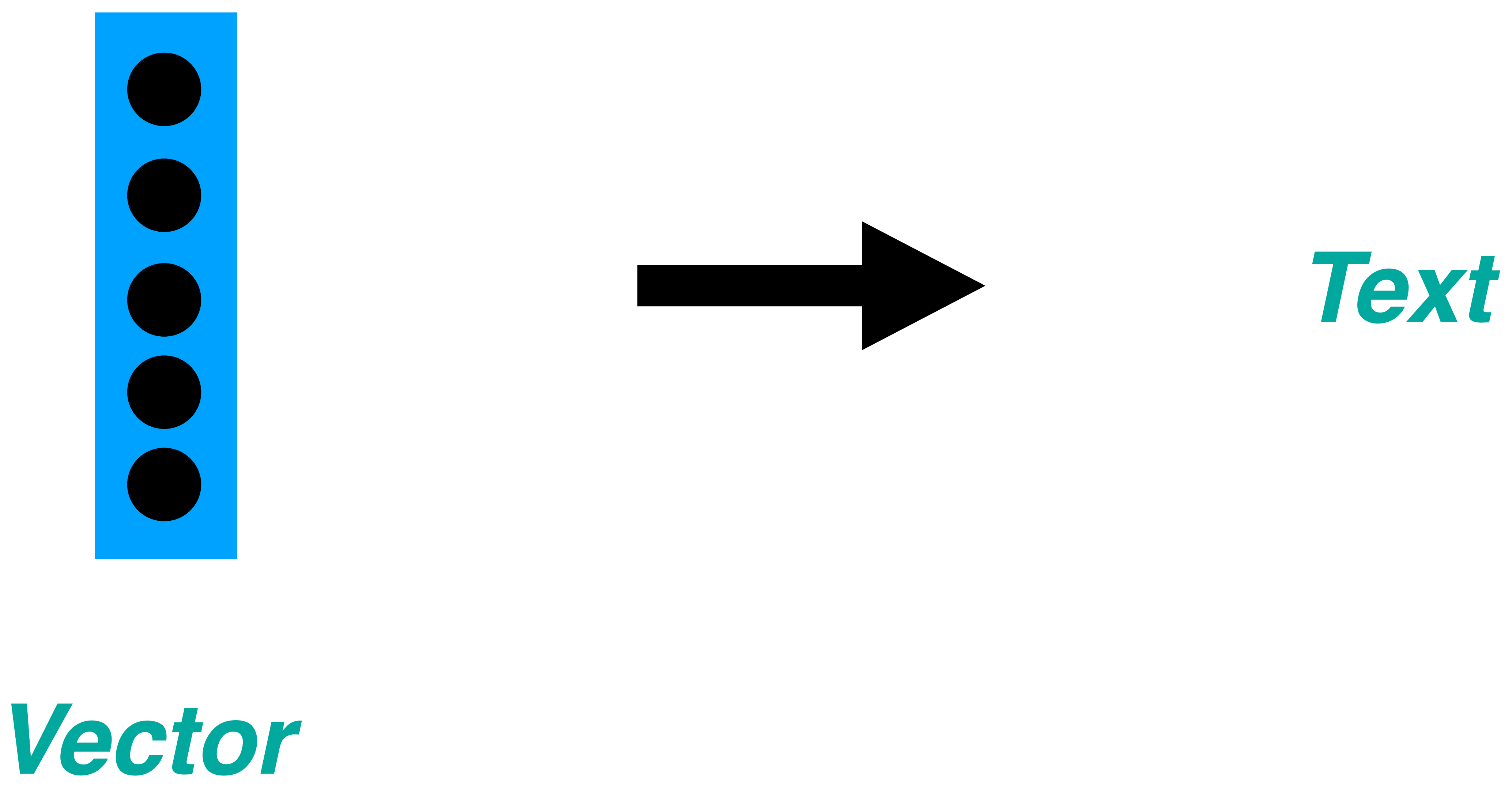


Pre-training

- Silver AMRs
- 30M sentences from CCNET parsed using JAMR



Decoder
(Generation)



Multilingual Generation

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

French

Des responsables américains

….

Spanish

Funcionarios estadounidenses

….

Slovak

Americkí predstavitelia

….

Bulgarian

Американските служители

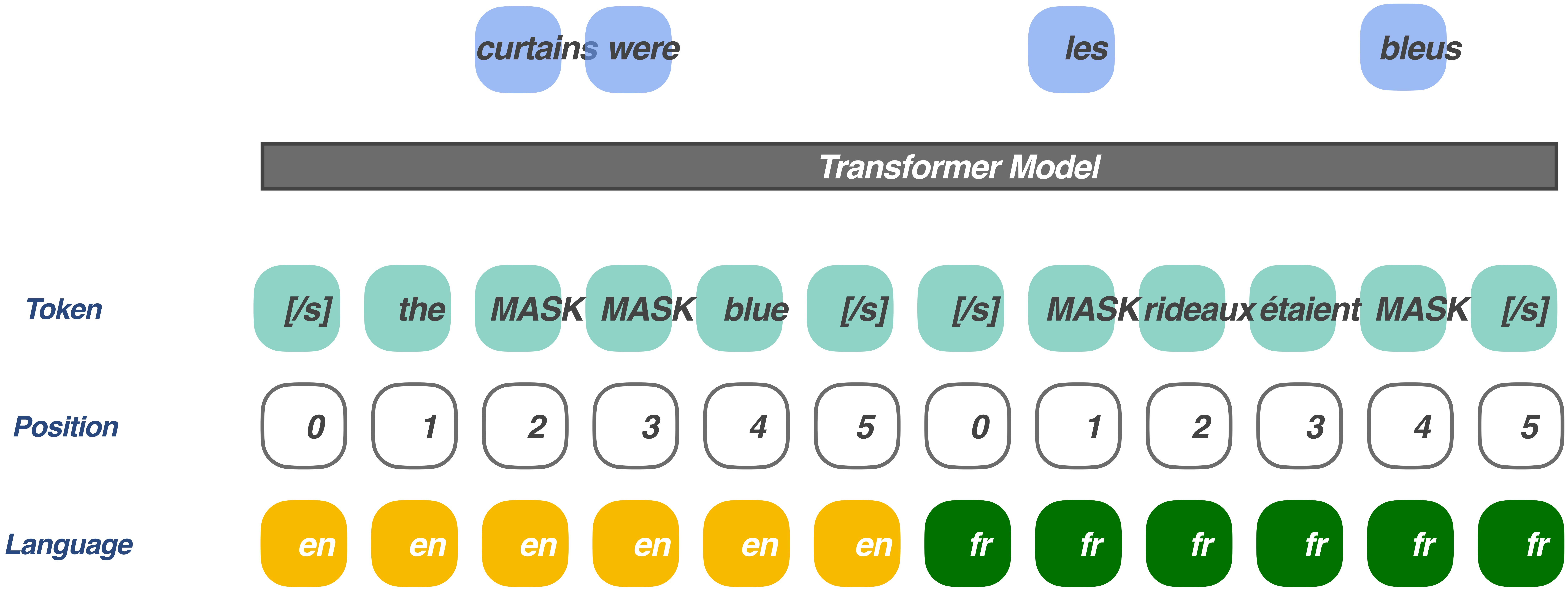
….

Swedish

Amerikanska tjänstemän

…

XLM Multilingual Word Embeddings



Cross-lingual Language Model Pretraining
Guillaume Lample, Alexis Conneau

Multilingual Encoder-decoder

Slovak

sv

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

↓

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

French

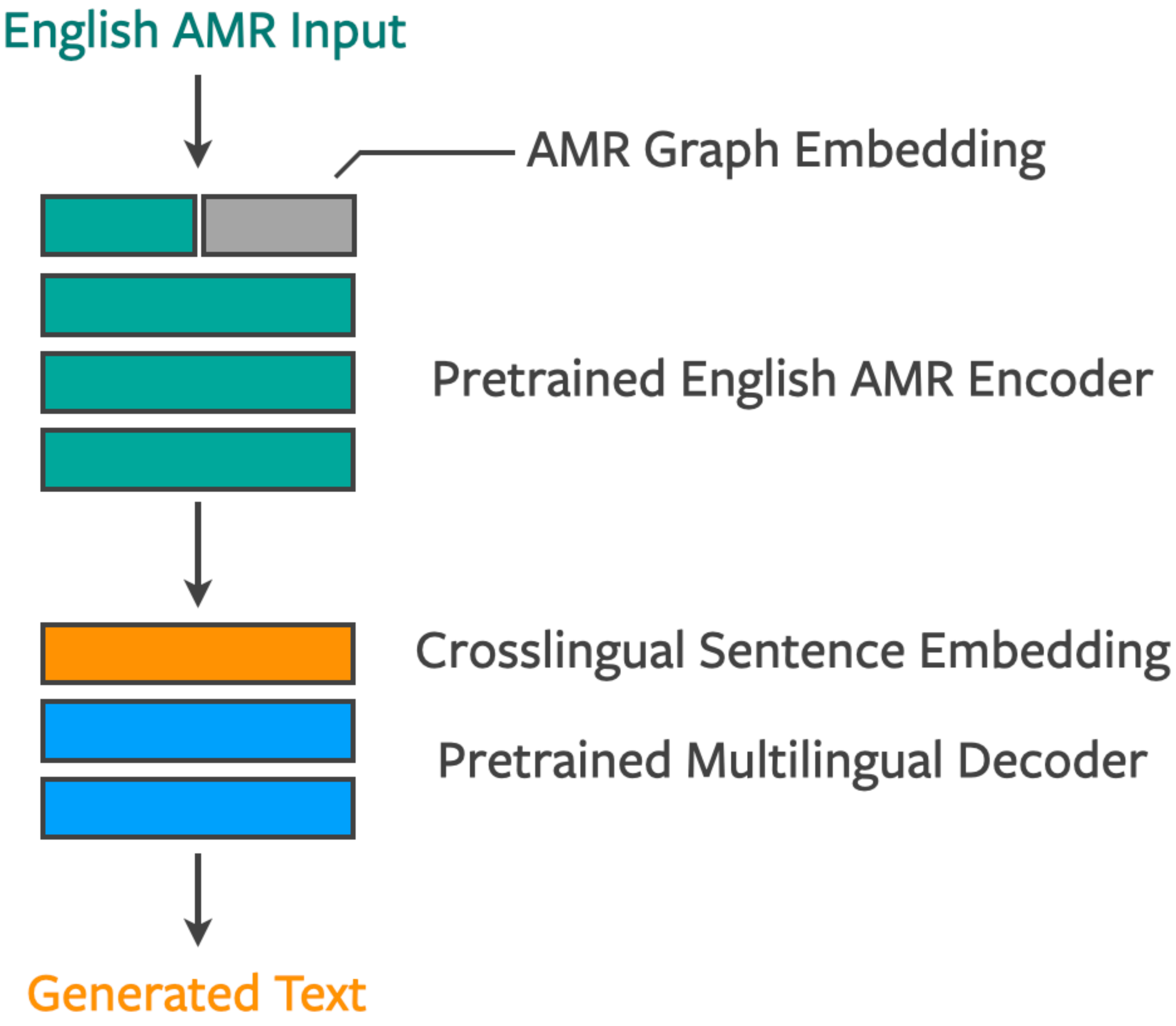
fr

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.

Multilingual AMR-to-Text Model



Encoder
- Pre-trained AMRs
- Structural Embeddings

Decoder
- Pre-trained LMs
- XLM Vocabulary

Multilingual Encoder-Decoder

Training Data (21 languages)

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.

French

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

Spanish

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

Slovak

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

Bulgarian

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

Swedish

Data

Training

- Silver AMR, Europarl Text, 21 languages

Test

- Silver AMR, Europarl text, 21 languages
- Gold AMR, LDC Text, Spanish, German, Italian

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

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: Monolingual vs. Multilingual

Monolingual

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.

Multilingual

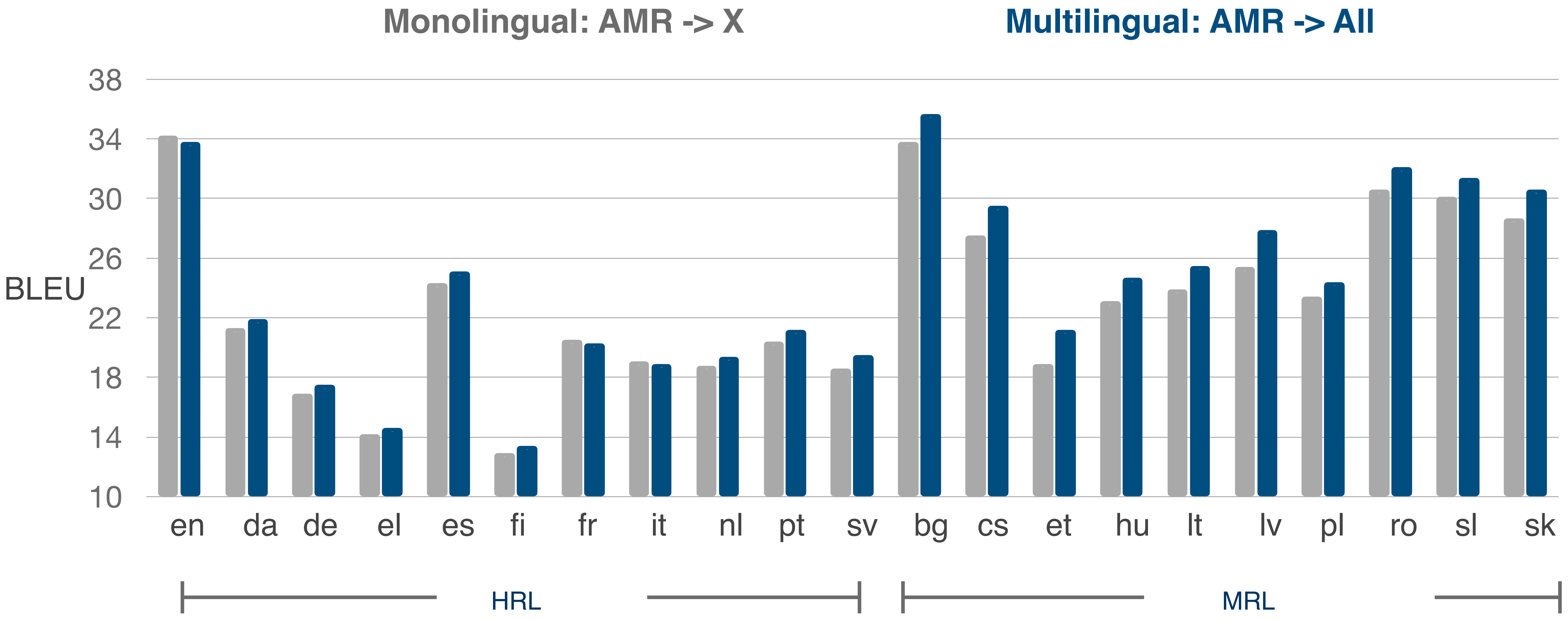
fr

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

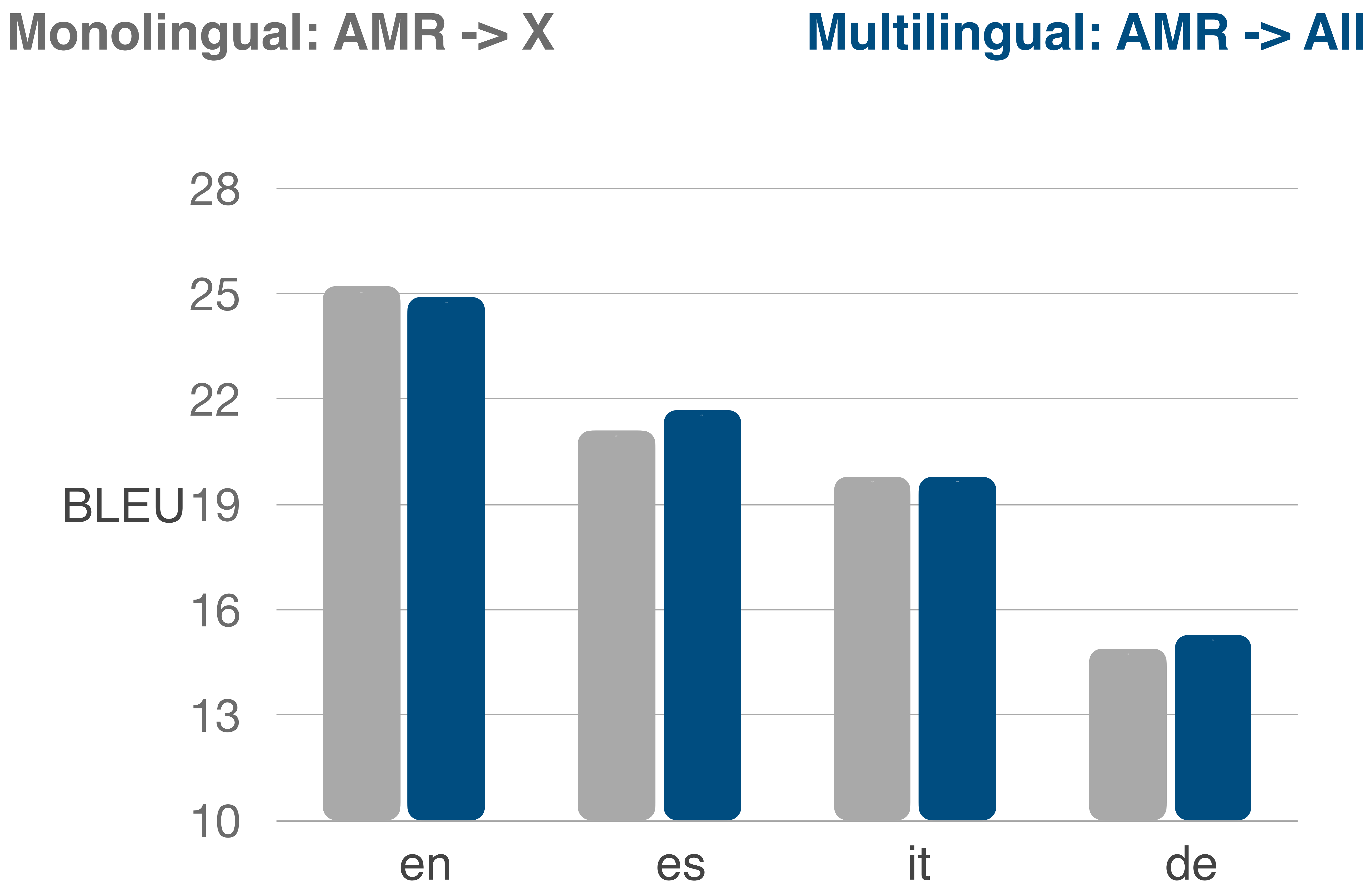
↓

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

Results: Europarl (Silver AMRs)



Results: LTC Data (Gold AMRs) or



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

AMR - English

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

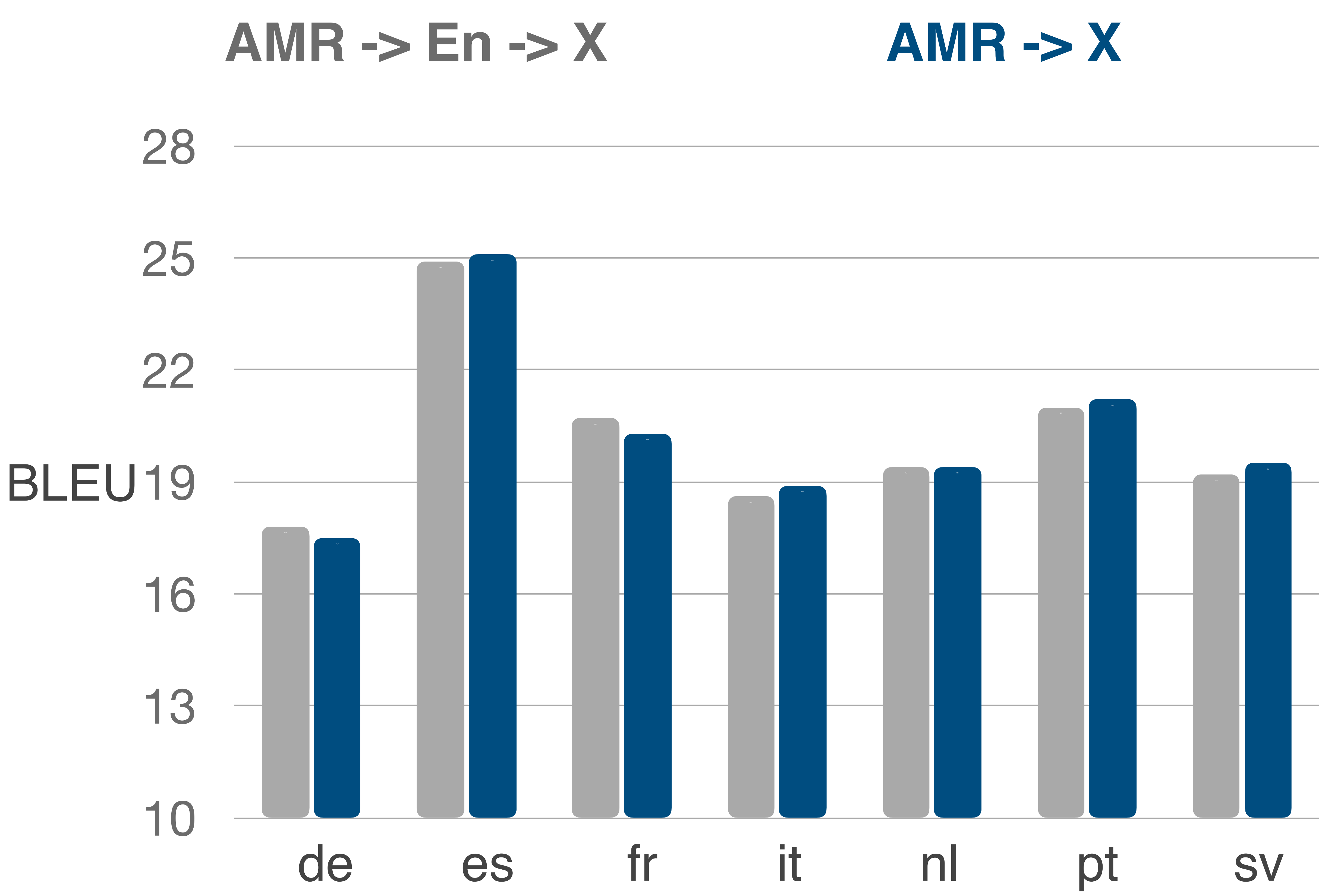
AMR - X

fr

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



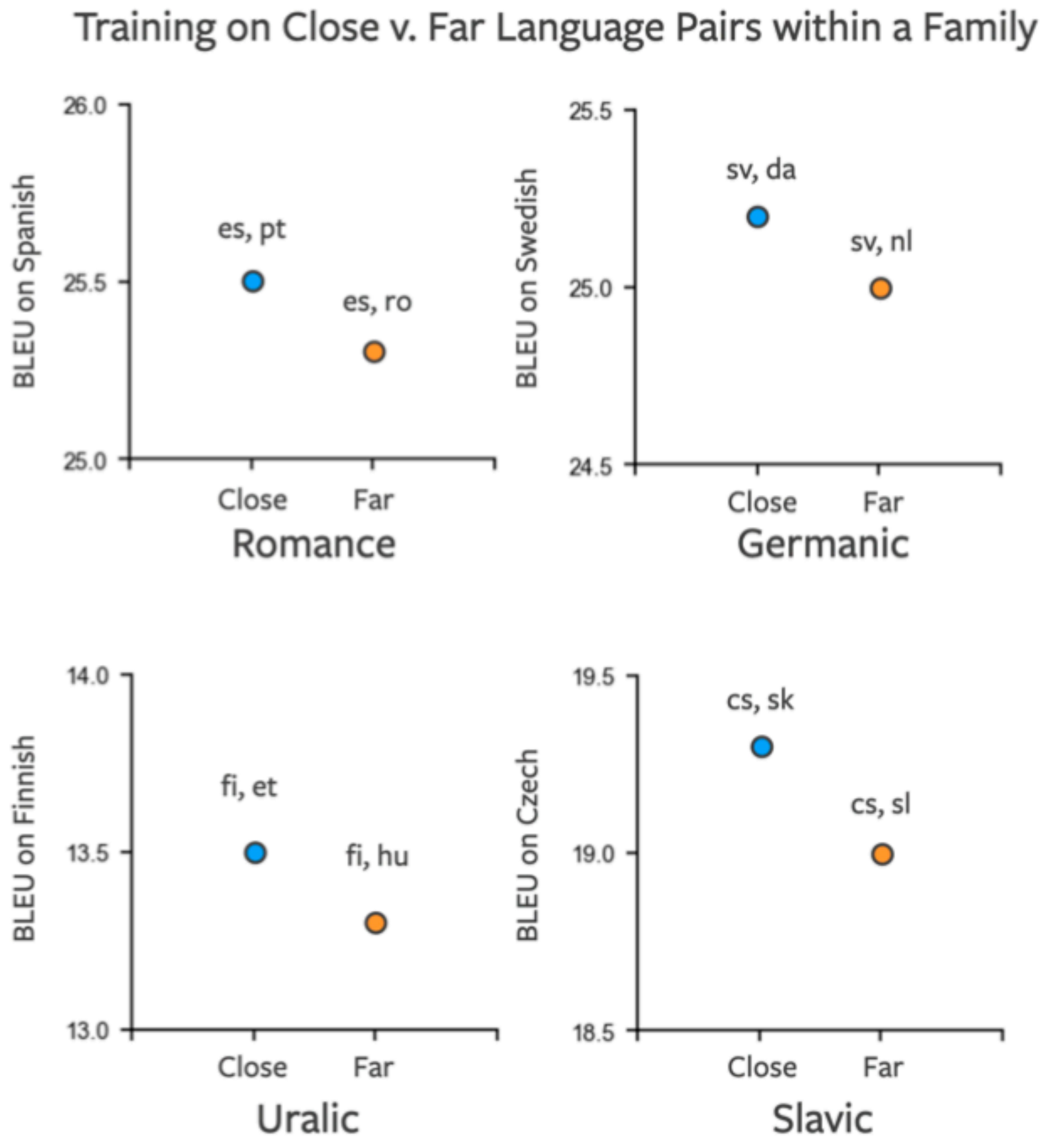
Impact of Training Languages

	Da	De	Nl	Sv
One Language	21.3	17.0	18.5	18.7
ALL Germanic Languages	21.8	21.9	19.6	19.3
21 languages	21.9	17.5	19.4	19.5

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

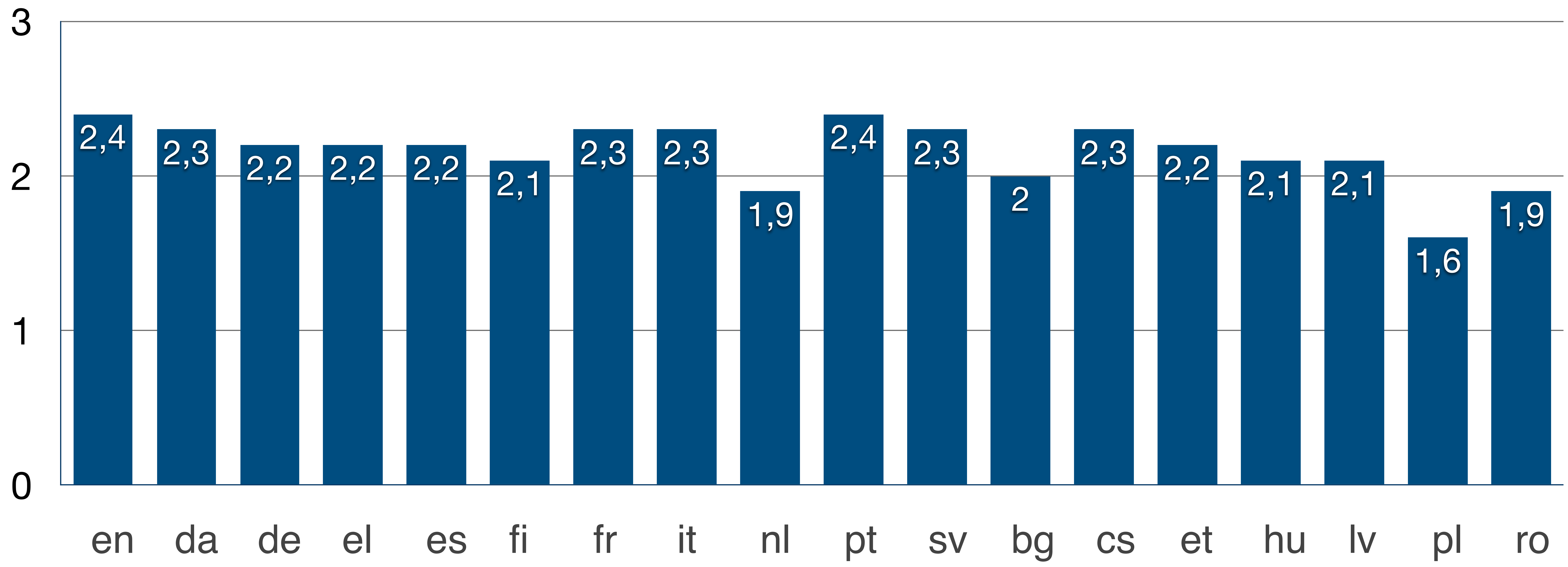


Human-Based Evaluation

- **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



Generated Paraphrases

REF: This point will certainly be the subject of subsequent further debates in the council

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

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

GEN: 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

SY **You**

convert the amr written between "" to French "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"

ChatGPT

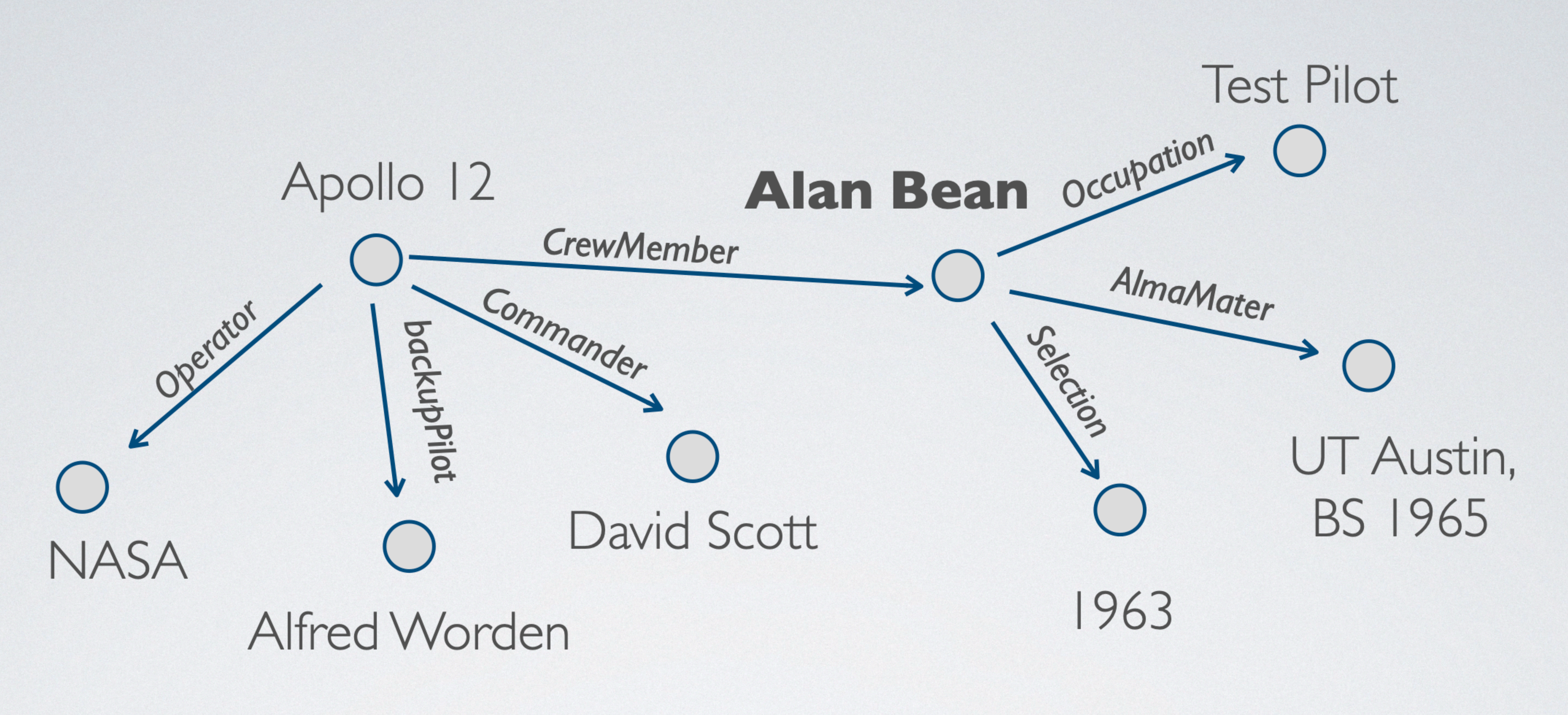
Organiser une réunion aux États-Unis à New York en janvier 2002. La réunion implique une personne occupant un rôle officiel et faisant également partie d'un groupe d'experts.

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

Generating from Knowledge Graphs

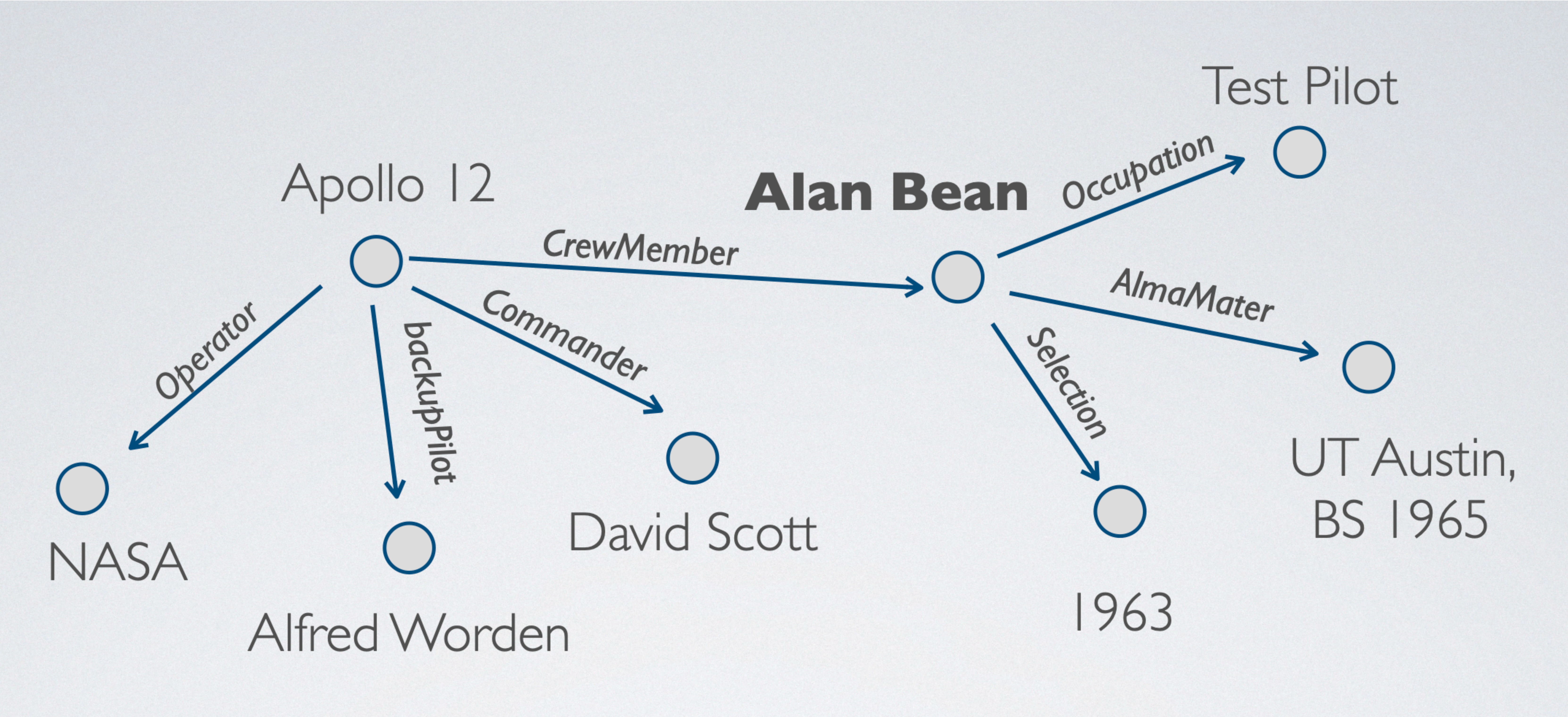
Gardent et al. ACL 2017, Castro-Ferreira et al. 2020, Cripwell et al. 2023
W. Soto-Martinez, Y. Parmentier and C. Gardent AACL 2023

The WebNLG Shared Task



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

	Train+Dev	Test (Seen Category)	Test (Unseen Category)	TOTAL
# (Graph, Text)	20,370	2,495	2,413	**25,298**
# Graphs	7,812	971	891	**9,674**

- DBPedia Graphs with root entities from different DBPedia Categories

- Text is crowd-sourced (human written)

Gardent et al. ACL 2017

WebNLG 2017 : RDF ==> English

	Train+Dev	Test (Seen Category)	Test (Unseen Category)	TOTAL
# (Graph,Text)	20,370	2,495	2,413	**25,298**
# Graphs	7,812	971	891	**9,674**

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 ==> English

6 participants, 10 systems

- 3 symbolic
- 1 statistical
- 5 neural

Gardent et al. ACL 2017

WebNLG 2017 : RDF ==> English

All

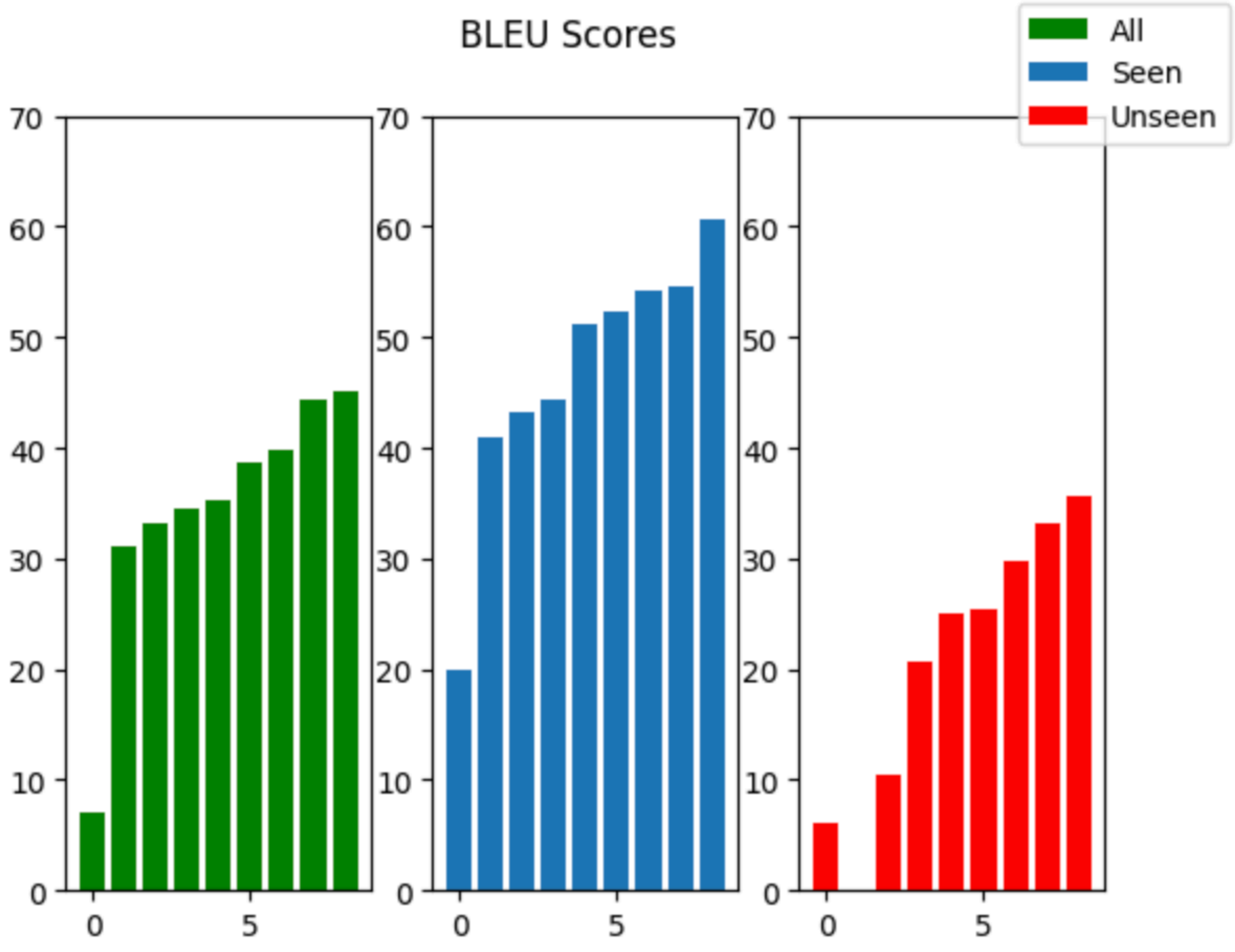
7.07 - 45.13

Seen

19.87 - 60.54

Unseen

5.13 - 35.7



WebNLG 2020

Generation

- RDF ==> English

Castro-Ferreira et al. INLG 2020

WebNLG 2020

Generation

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

Castro-Ferreira et al. INLG 2020

WebNLG 2020

Generation

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

Semantic Parsing

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

Castro-Ferreira et al. INLG 2020

WebNLG 2020

	Train	Dev	Test NLG/SP	TOTAL
# (Graph,Text)	35,426	4,664	5,150	47,395
# Graphs	13,211	1,667	1,779	17,409

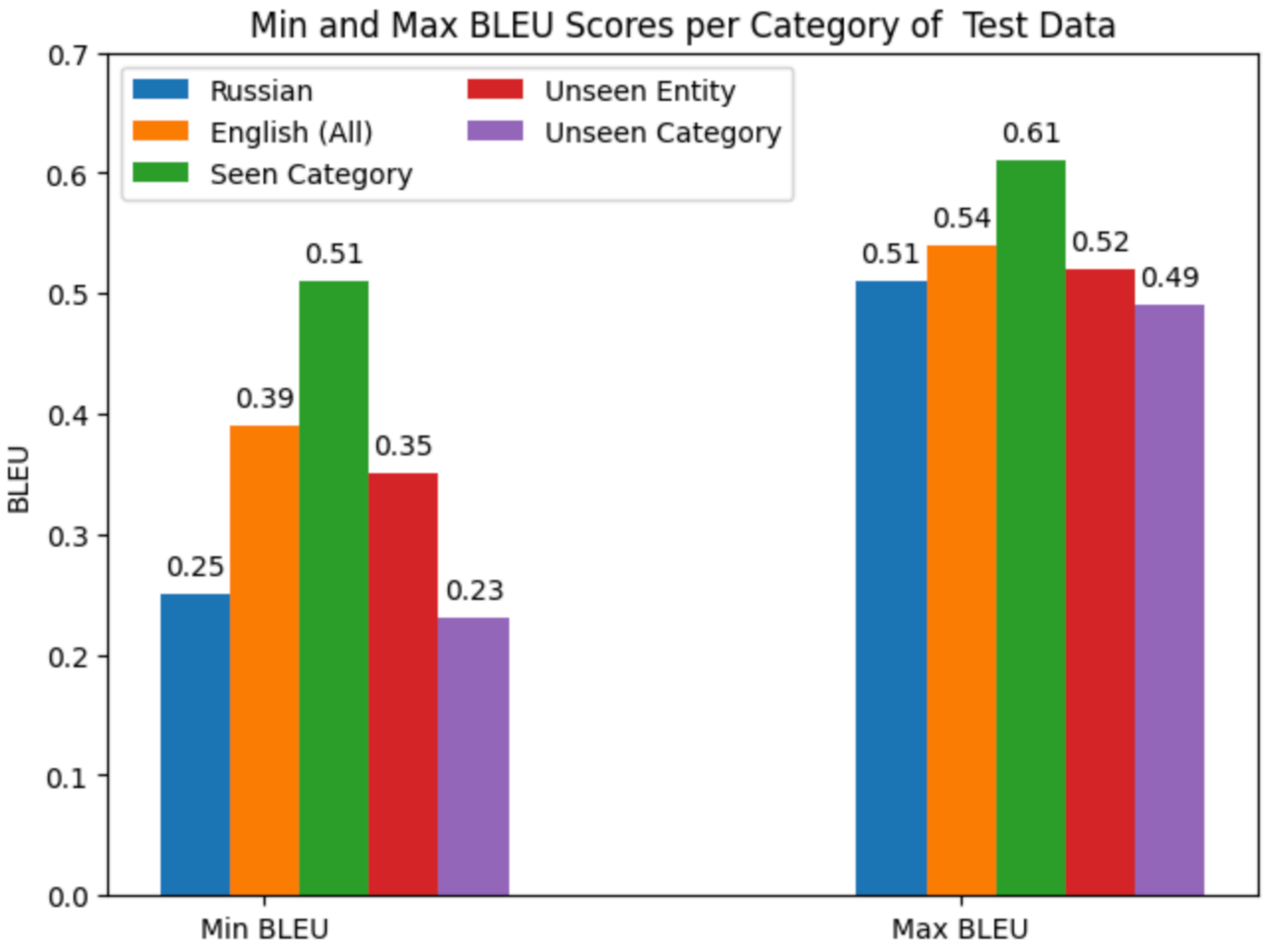
- 16 seen categories
- 3 unseen categories
- Unseen entities

WebNLG 2020 : Participation

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

17 participants

WebNLG 2020 : Results



WebNLG 2023 : Low Resource Languages

	Silver Train	Dev	Test
Breton	13,211	1,399	1,778
Welsh	13,211	1,665	1,778
Irish	13,211	1,665	1,778
Maltese	13,211	1,665	1,778

WebNLG 2023 : Generation + Translation

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

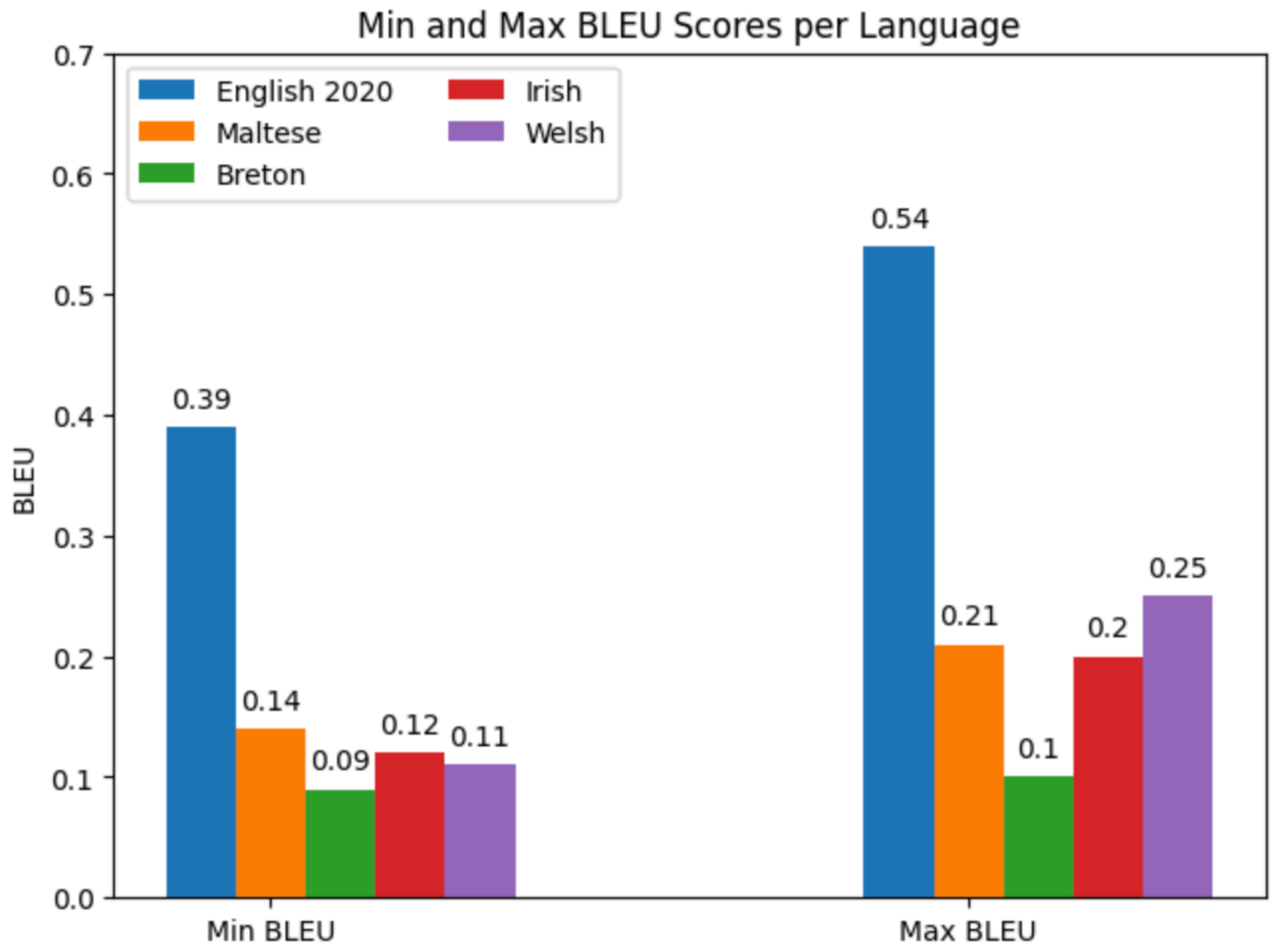
RDF ==> English

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

English ==> LRL

- Machine Translation : NLLB or Google Translate

WebNLG 2023 : Results



NLG+MT Pipeline vs. End-to-End

NLG+MT

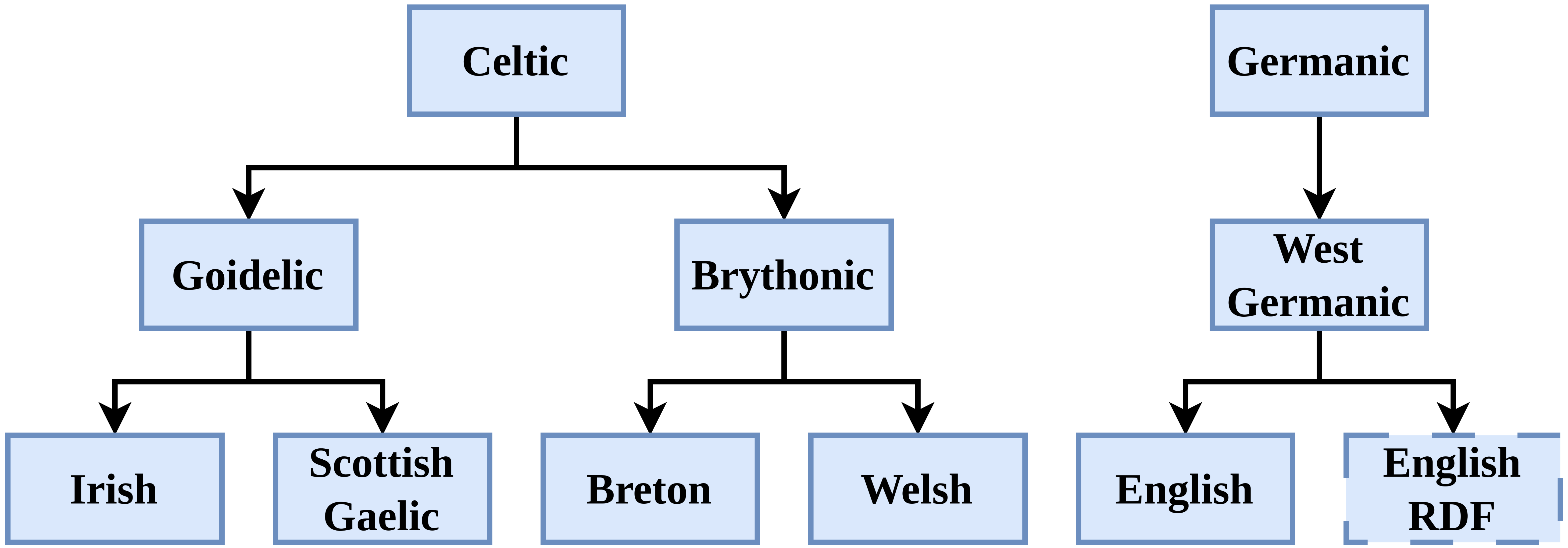
- Only possible for Languages with MT models
- For Breton, Machine Translation is very poor

✗ Generation + Translation

✗ Fine-Tuning (BLEU : 0.10)

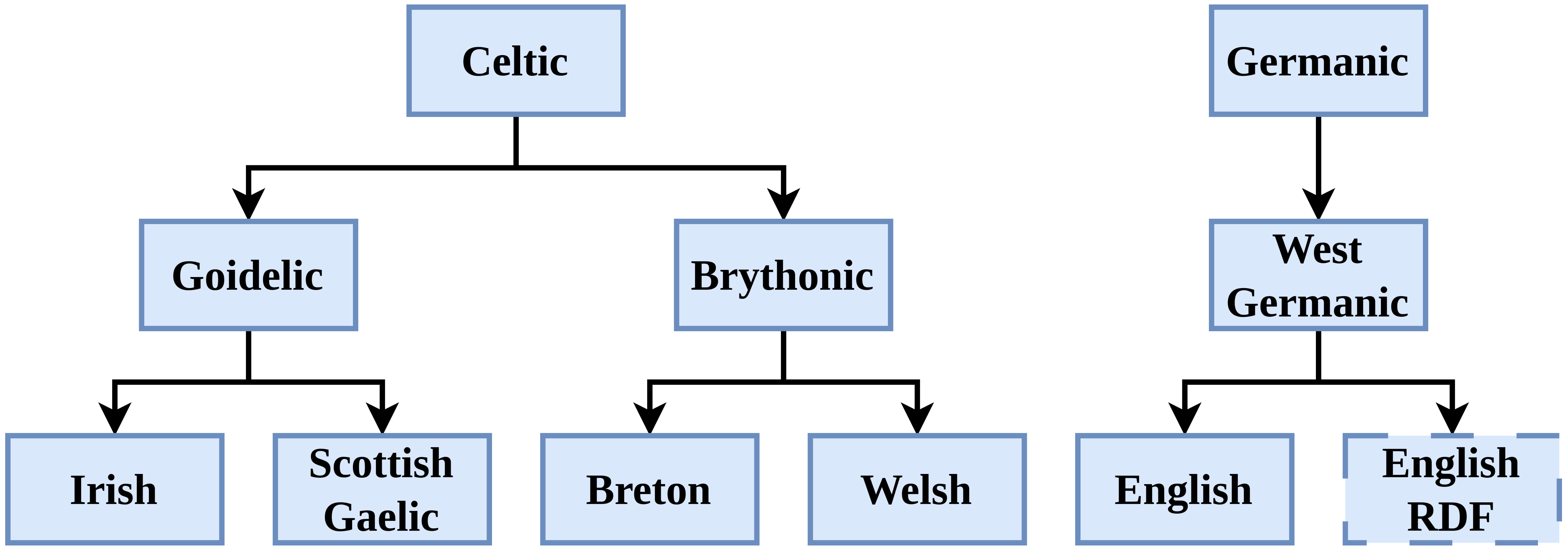
✓ Soft-Prompt Fine-tuning

WebNLG 2023 : Soft Prompt Fine Tuning



- Fine tune mT5 on WebNLG Dev set
- Structured soft-prompt

WebNLG 2023 : Structured Soft Prompt



Soft Prompt

50 Tokens Task	15 Tokens Source Family	15 Tokens Source Genus	15 Tokens Source Language	15 Tokens Target Family	15 Tokens Target Genus	15 Tokens Target Language	n Tokens Input Sequence

Pre-training

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

Monolingual data

	Task	Source			Target			Original Input Sequences					
		Family	Genus	Lang.	Family	Genus	Lang.						
Input Batch	Masked LM	Germanic	West Germanic	RDF	Germanic	West Germanic	RDF	<S>	Einstein	<P>	<mask>	<P>	Poland
	Prefix LM	Germanic	West Germanic	English	Germanic	West Germanic	English	Thank	you	for	<mask>	<pad>	<pad>
	Suffix LM	Celtic	Britonic	Welsh	Celtic	Britonic	Welsh	<mask>	honno	?	<pad>	<pad>	<pad>
	Deshuffling	Celtc	Britonic	Breton	Celtic	Britonic	Breton	skuizh	?	out	Ha	<pad>	<pad>
	Generate	Celtc	Goidelic	Irish	Celtic	Goidelic	Irish	Seo	<mask>	<pad>	<pad>	<pad>	<pad>

Pre-training

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

Monolingual Data

	Task	Source			Target			Original Input Sequences					
		Family	Genus	Lang.	Family	Genus	Lang.						
Input Batch	Masked LM	Germanic	West Germanic	RDF	Germanic	West Germanic	RDF	<S>	Einstein	<P>	<mask>	<P>	Poland
	Prefix LM	Germanic	West Germanic	English	Germanic	West Germanic	English	Thank	you	for	<mask>	<pad>	<pad>
	Suffix LM	Celtic	Britonic	Welsh	Celtic	Britonic	Welsh	<mask>	honno	?	<pad>	<pad>	<pad>
	Deshuffling	Celtc	Britonic	Breton	Celtic	Britonic	Breton	skuizh	?	out	Ha	<pad>	<pad>
	Generate	Celtc	Goidelic	Irish	Celtic	Goidelic	Irish	Seo	<mask>	<pad>	<pad>	<pad>	<pad>

Step 2. Fine-tuning on RDF-Text data

Pre-training

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

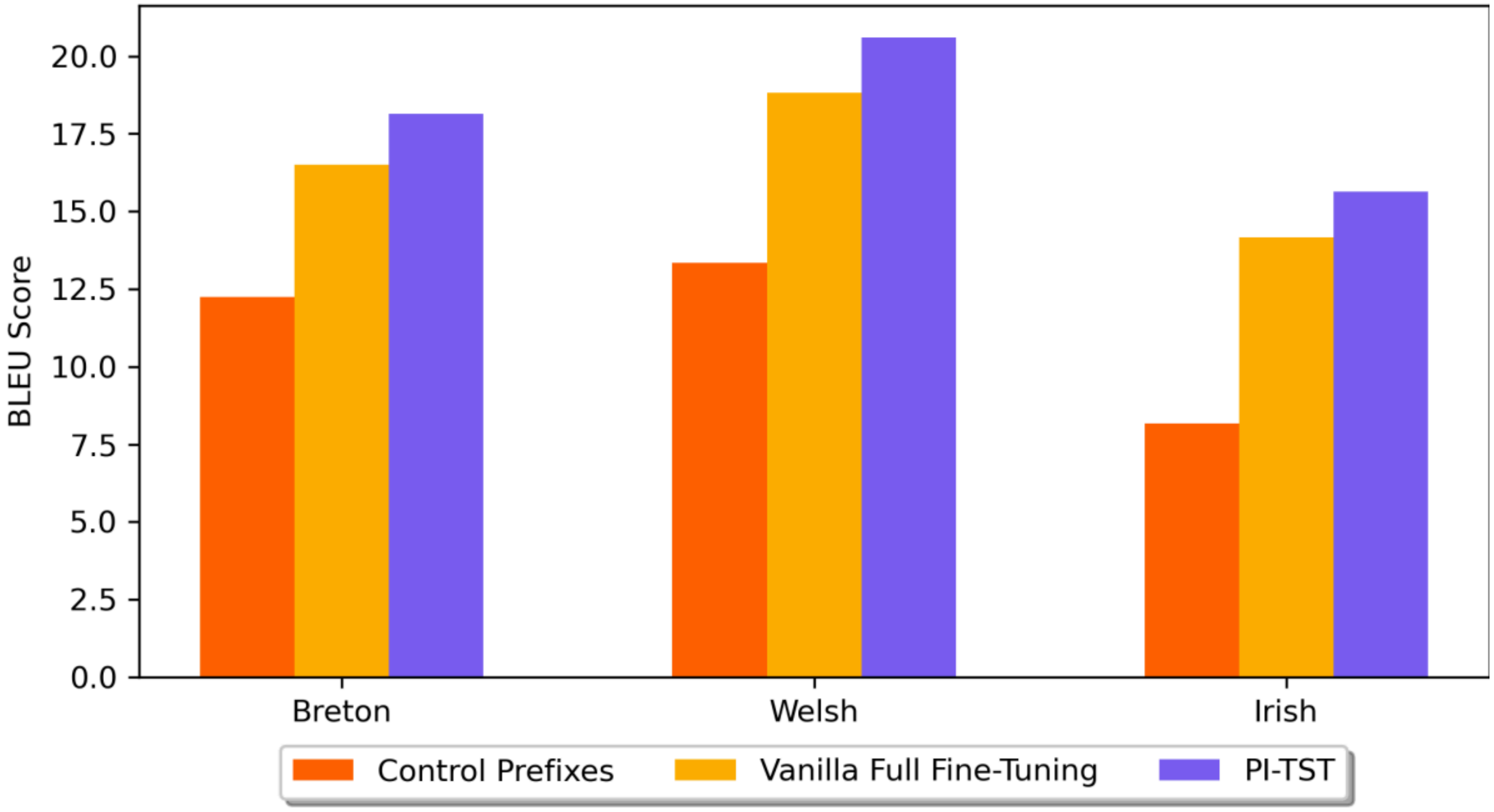
Monolingual Data

	Task	Source			Target			Original Input Sequences					
		Family	Genus	Lang.	Family	Genus	Lang.						
Input Batch	Masked LM	Germanic	West Germanic	RDF	Germanic	West Germanic	RDF	<S>	Einstein	<P>	<mask>	<P>	Poland
	Prefix LM	Germanic	West Germanic	English	Germanic	West Germanic	English	Thank	you	for	<mask>	<pad>	<pad>
	Suffix LM	Celtic	Britonic	Welsh	Celtic	Britonic	Welsh	<mask>	honno	?	<pad>	<pad>	<pad>
	Deshuffling	Celtc	Britonic	Breton	Celtic	Britonic	Breton	skuizh	?	out	Ha	<pad>	<pad>
	Generate	Celtc	Goidelic	Irish	Celtic	Goidelic	Irish	Seo	<mask>	<pad>	<pad>	<pad>	<pad>

Step 2. Fine-tuning on RDF-Text data

Inference. Using the target language soft prompt

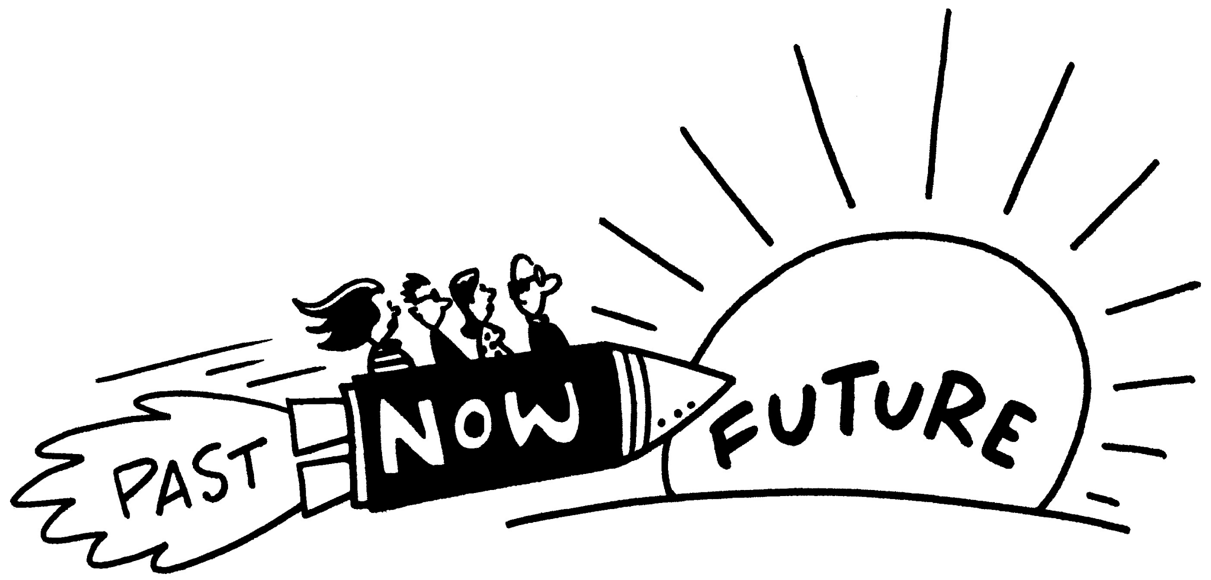
WebNLG 2023 : Results



Key Takeaways

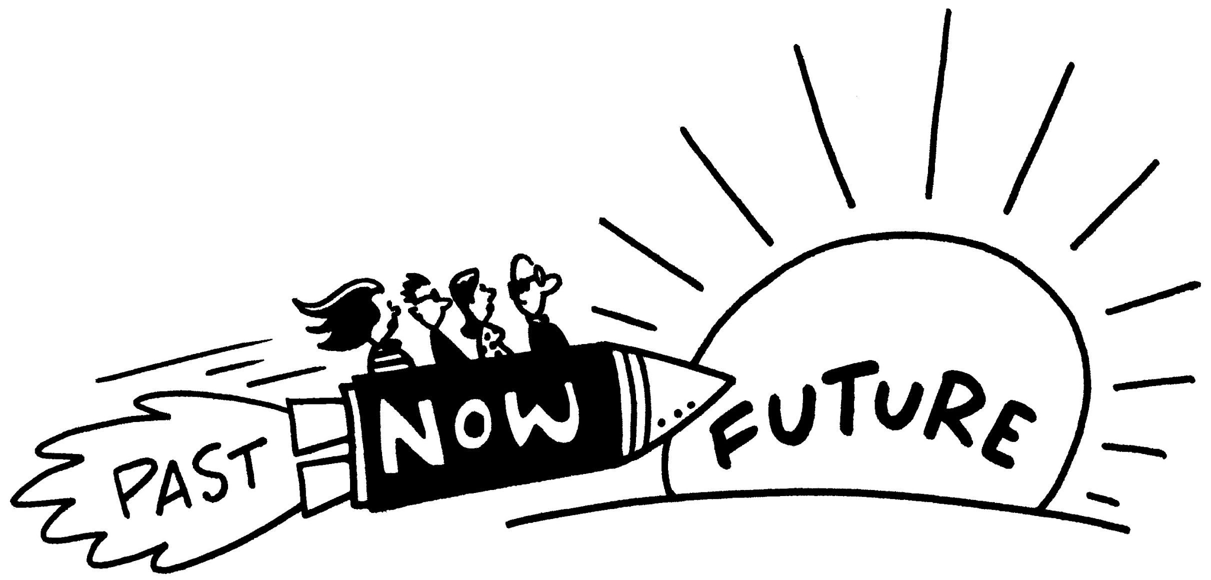
- **Pre-Training:** 2017 vs 2020, Pre-training improve results
- **Target Language**: Results are better for English than for Russian
- **Generalisation:** Models underperform on unseen data
- **LRL**: results are poor for LRL
- **LRL Models**: Pipeline models are used but depends on having MT models 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) —> 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

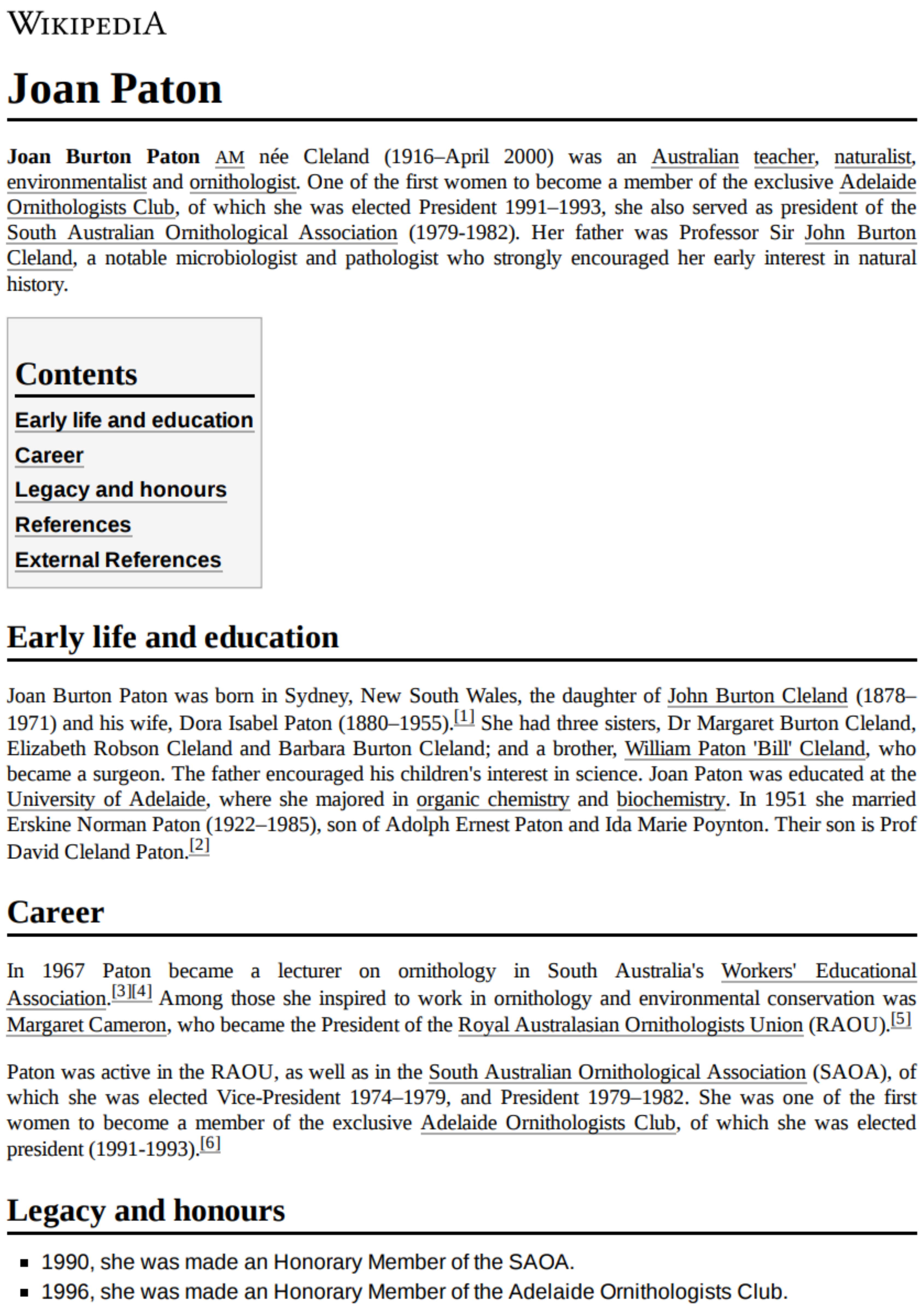
Angela **Fan and Claire Gardent**
"Generating Full Length Wikipedia Biographies. The Impact of Gender Bias on the Retrieval-Based Generation of Women Biographies."
ACL 2022

Generating Wikipedia Biographie

PERSON NAME

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 microbiologist and pathologist who strongly encouraged her early interest in natural history.

Contents

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 (1880–1955).[1] She had three sisters, Dr Margaret Burton Cleland, Elizabeth Robson Cleland and Barbara Burton Cleland; and a brother, William Paton 'Bill' 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 Erskine Norman Paton (1922–1985), son of Adolph Ernest Paton and Ida 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][4] Among those she inspired to work in ornithology and environmental conservation was Margaret Cameron, who became the President of the Royal Australasian Ornithologists Union (RAOU).[5]

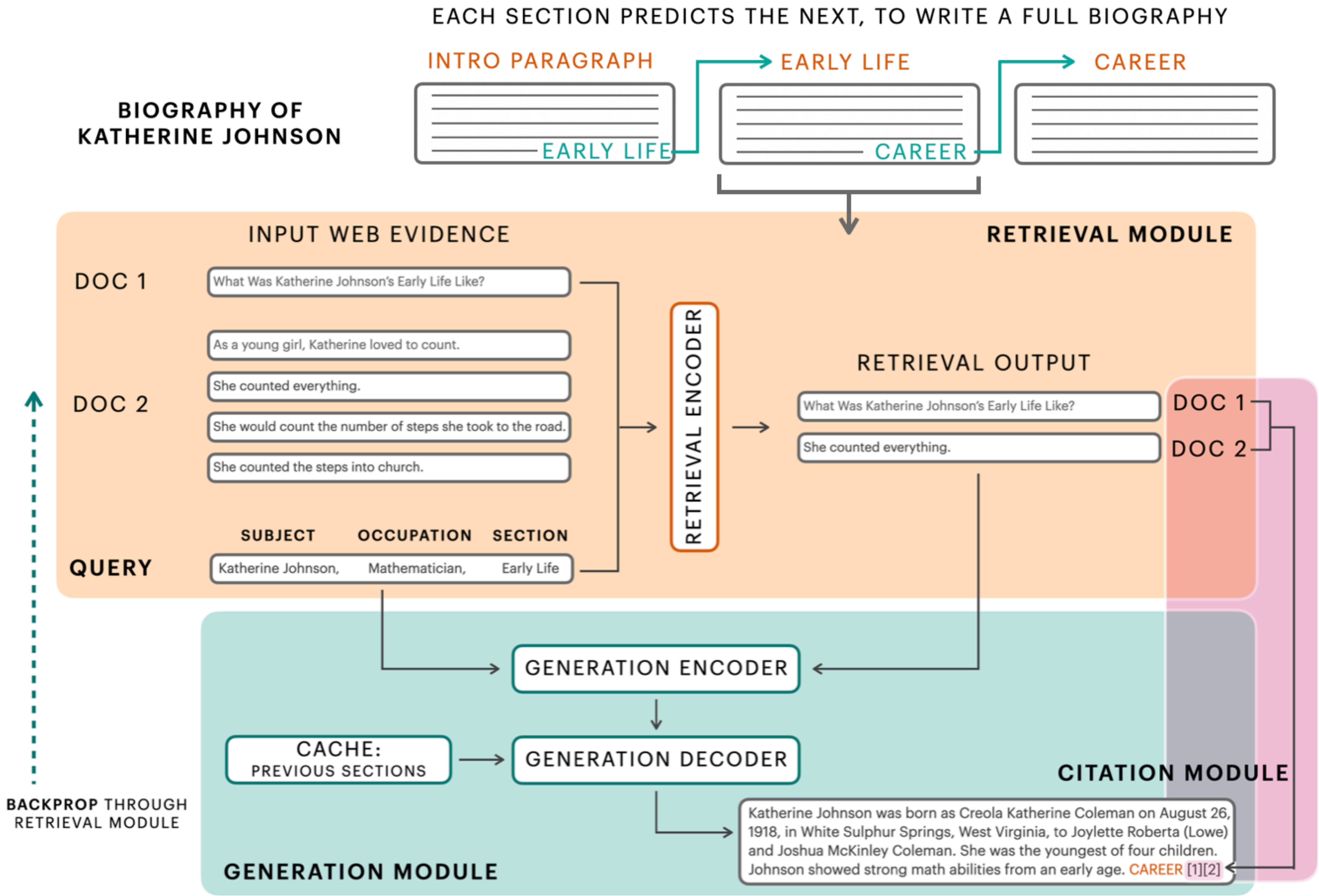
Paton was active in the RAOU, as well as in the South Australian Ornithological Association (SAOA), of which she was elected Vice-President 1974–1979, and President 1979–1982. She was one of the first women to become a member of the exclusive Adelaide Ornithologists Club, of which she was elected president (1991-1993).[6]

Legacy and honours

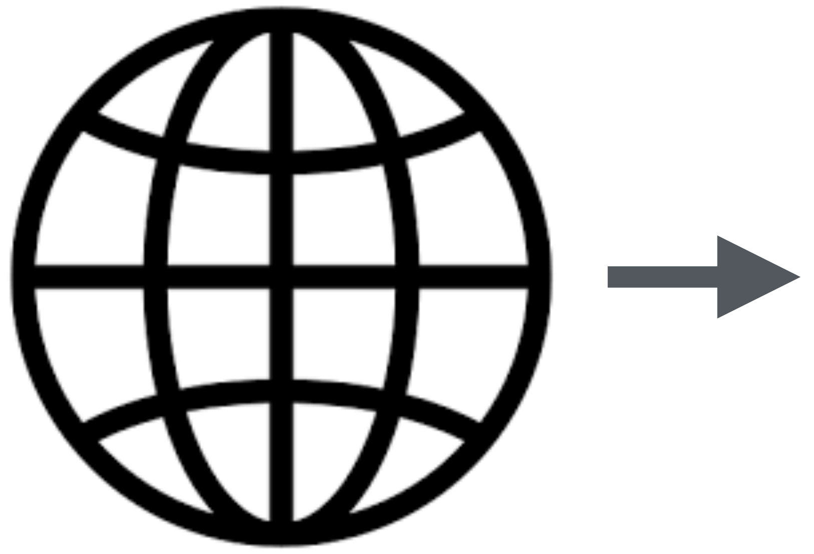
- 1990, she was made an Honorary Member of the SAOA.
- 1996, she was made an Honorary Member of the Adelaide Ornithologists Club.

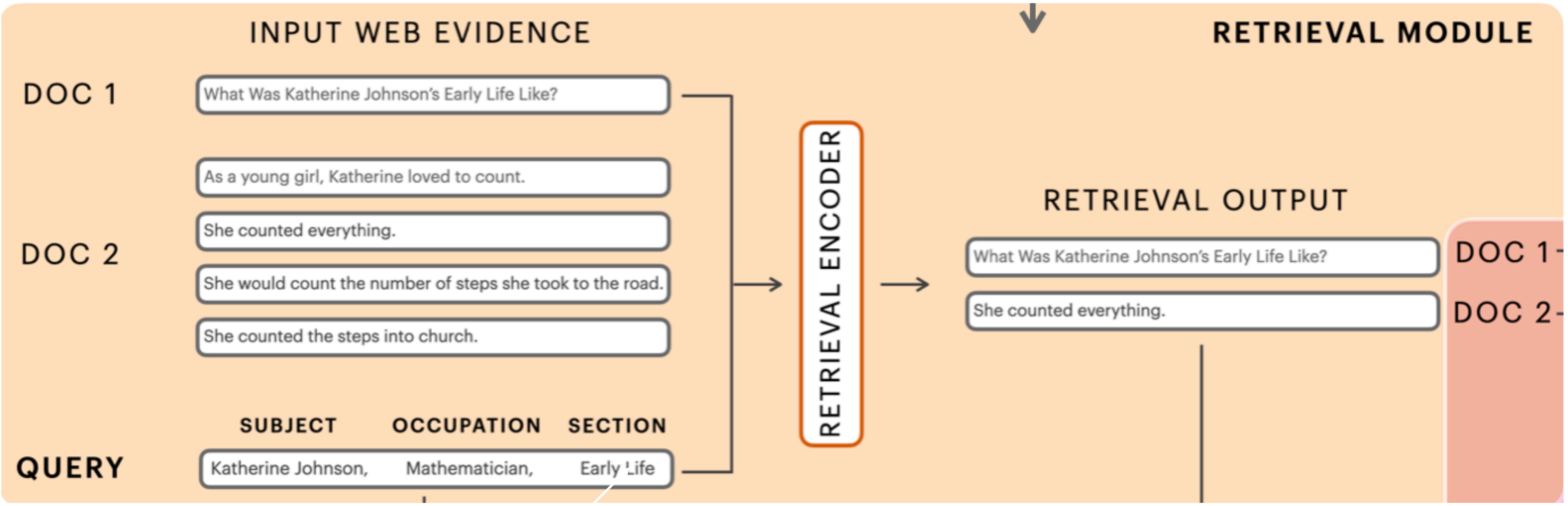
Challenges

- Gather relevant evidence
- Generate structured Text
- Factuality


EACH SECTION PREDICTS THE NEXT, TO WRITE A FULL BIOGRAPHY
BIOGRAPHY OF KATHERINE JOHNSON
INTRO PARAGRAPH
EARLY LIFE
CAREER
EARLY LIFE
CAREER
RETRIEVAL MODULE
INPUT WEB EVIDENCE
DOC 1
What Was Katherine Johnson's Early Life Like?
DOC 2
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.
SUBJECT
OCCUPATION
SECTION
QUERY
Katherine Johnson,
Mathematician,
Early Life
RETRIEVAL ENCODER
RETRIEVAL OUTPUT
What Was Katherine Johnson's Early Life Like?
She counted everything.
DOC 1
DOC 2
GENERATION ENCODER
CACHE: PREVIOUS SECTIONS
GENERATION DECODER
CITATION MODULE
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]
BACKPROP THROUGH RETRIEVAL MODULE
GENERATION MODULE

Extraction





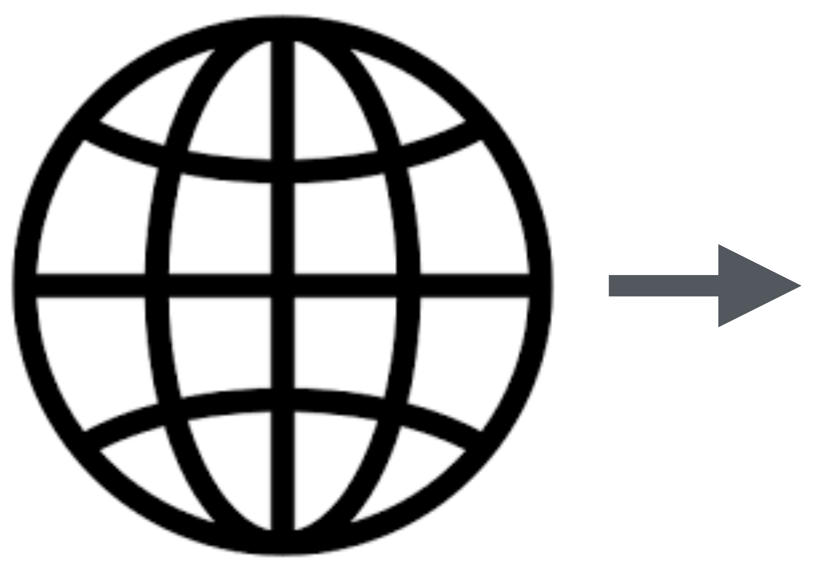
QUERY

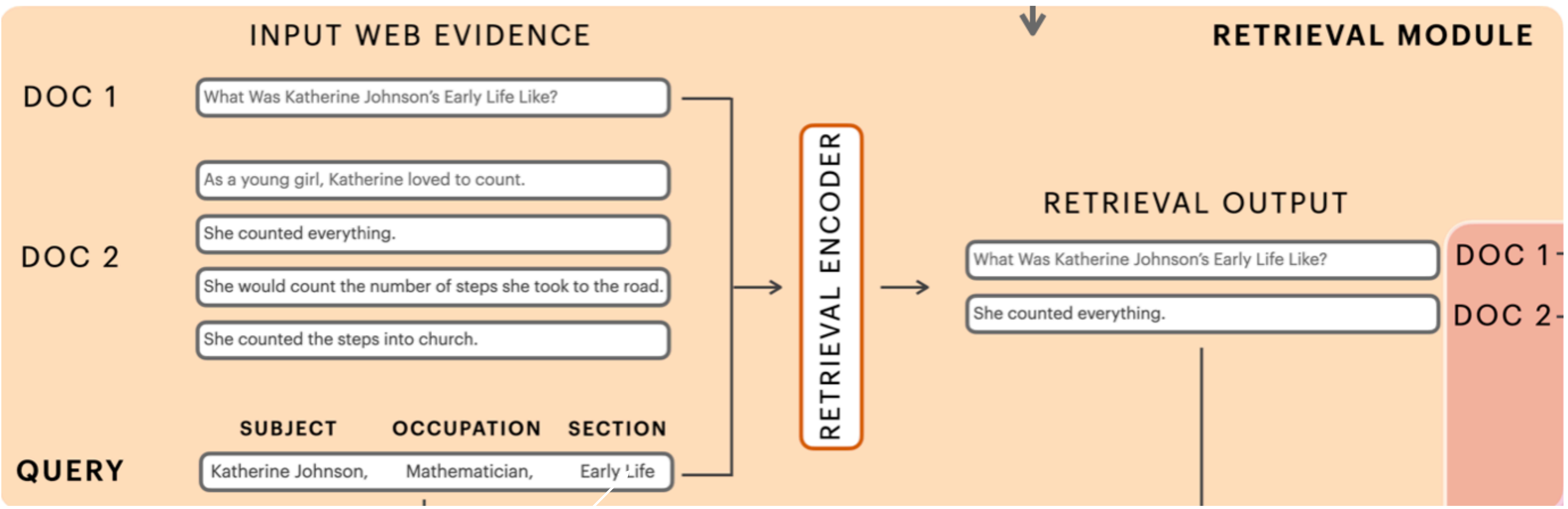
Katherine Johnson

Mathematician

Early Life

Extraction





QUERY

Katherine Johnson

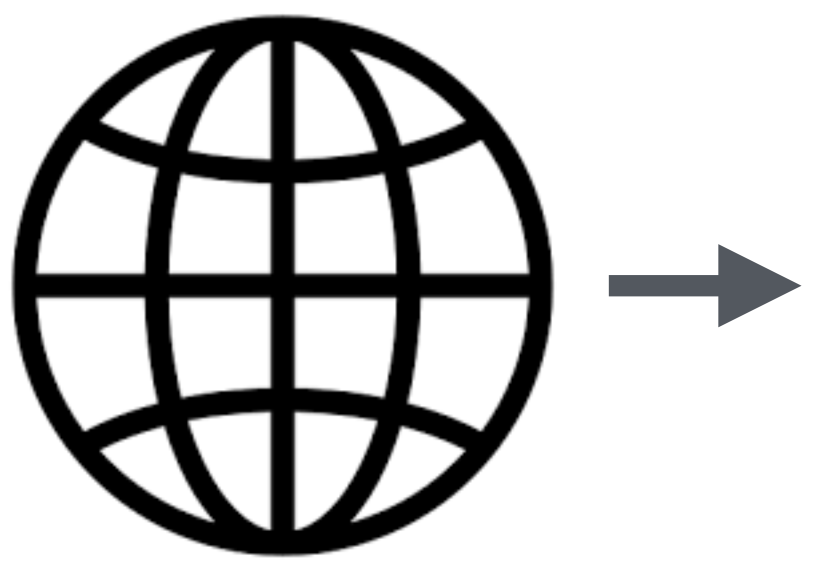
Mathematician

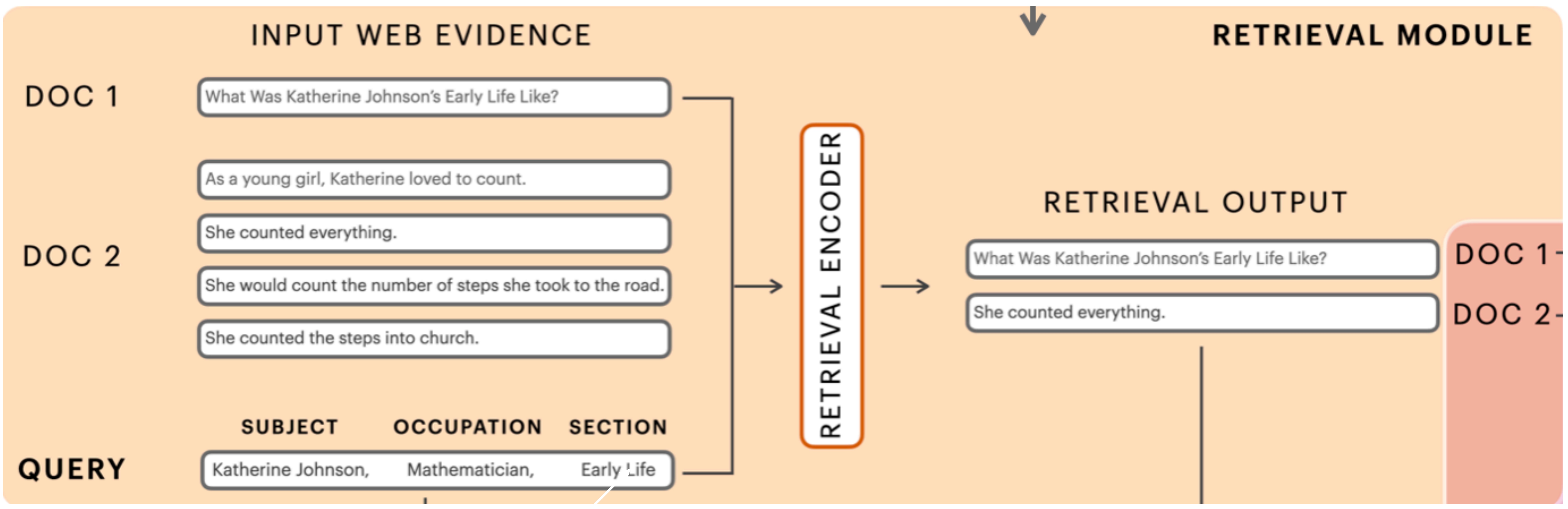
Early Life

RESULT

20 first documents segmented in sentences

Extraction





QUERY

Katherine Johnson

Mathematician

Early Life

Web Search

20 first documents segmented in sentences

Semantic Filtering

40 most similar sentences (1,000 words)

Generation

QUERY

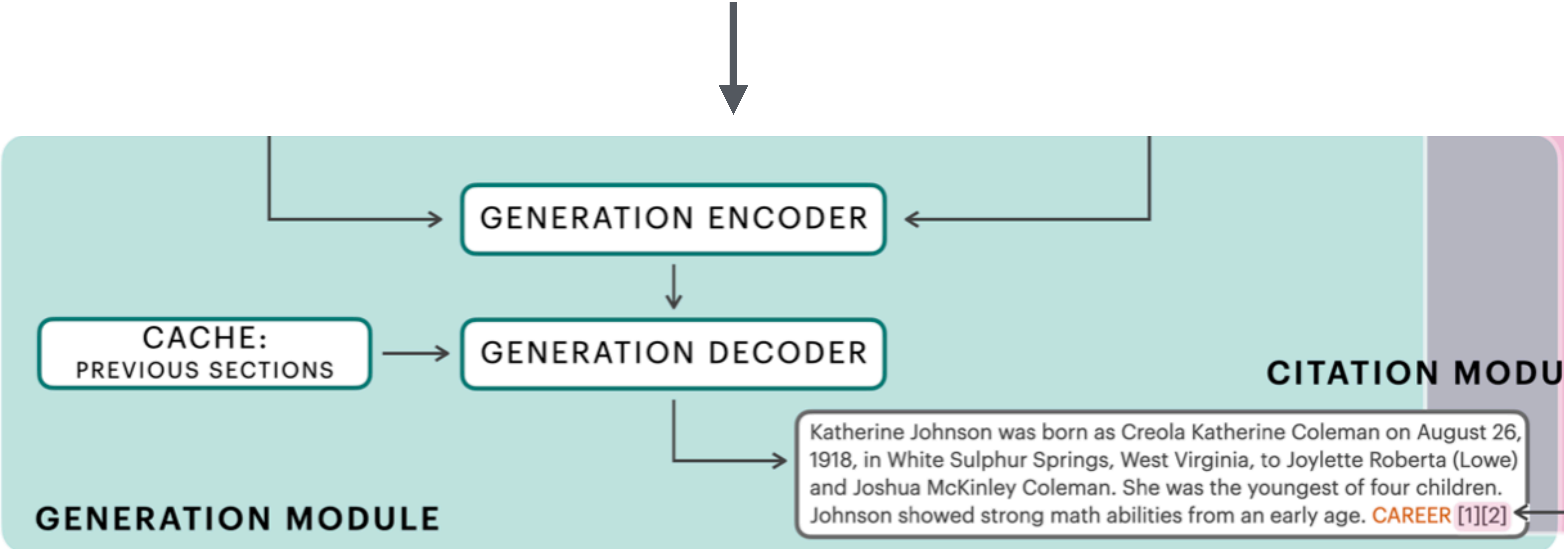
Katherine Johnson

Mathematician

Early Life

EXTRACTION

1,000 words



Cache Transformer-XL

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

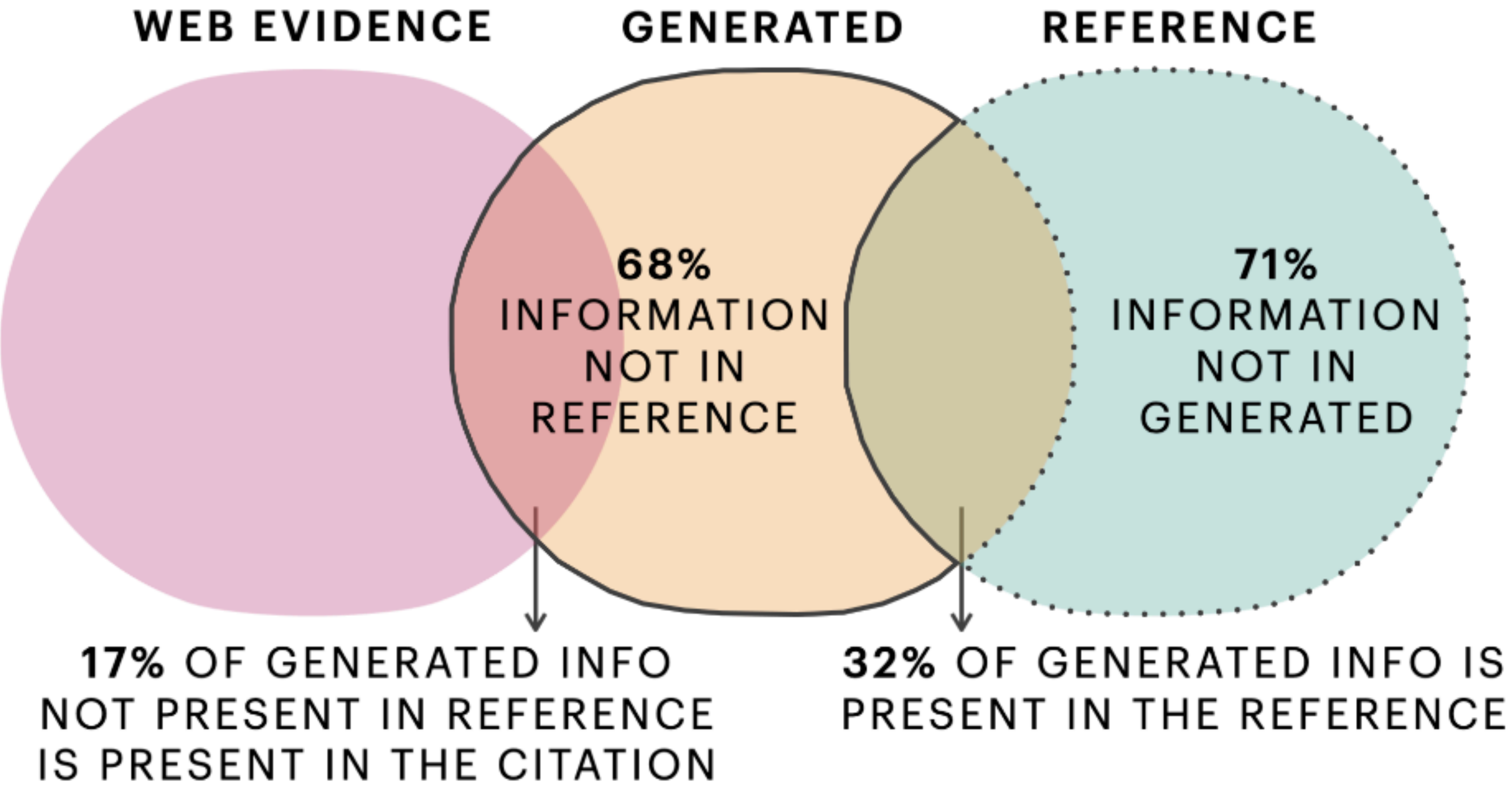
Ablation

Model	ROUGE-L	Entailment	Named Entity Coverage
BART Pretraining + Finetuning	17.4	15.8	21.9
+ Retrieval Module	18.8	17.2	23.1
+ Caching Mechanism	19.3	17.9	23.4

IR and the cache mechanism allow for statistically significant improvement

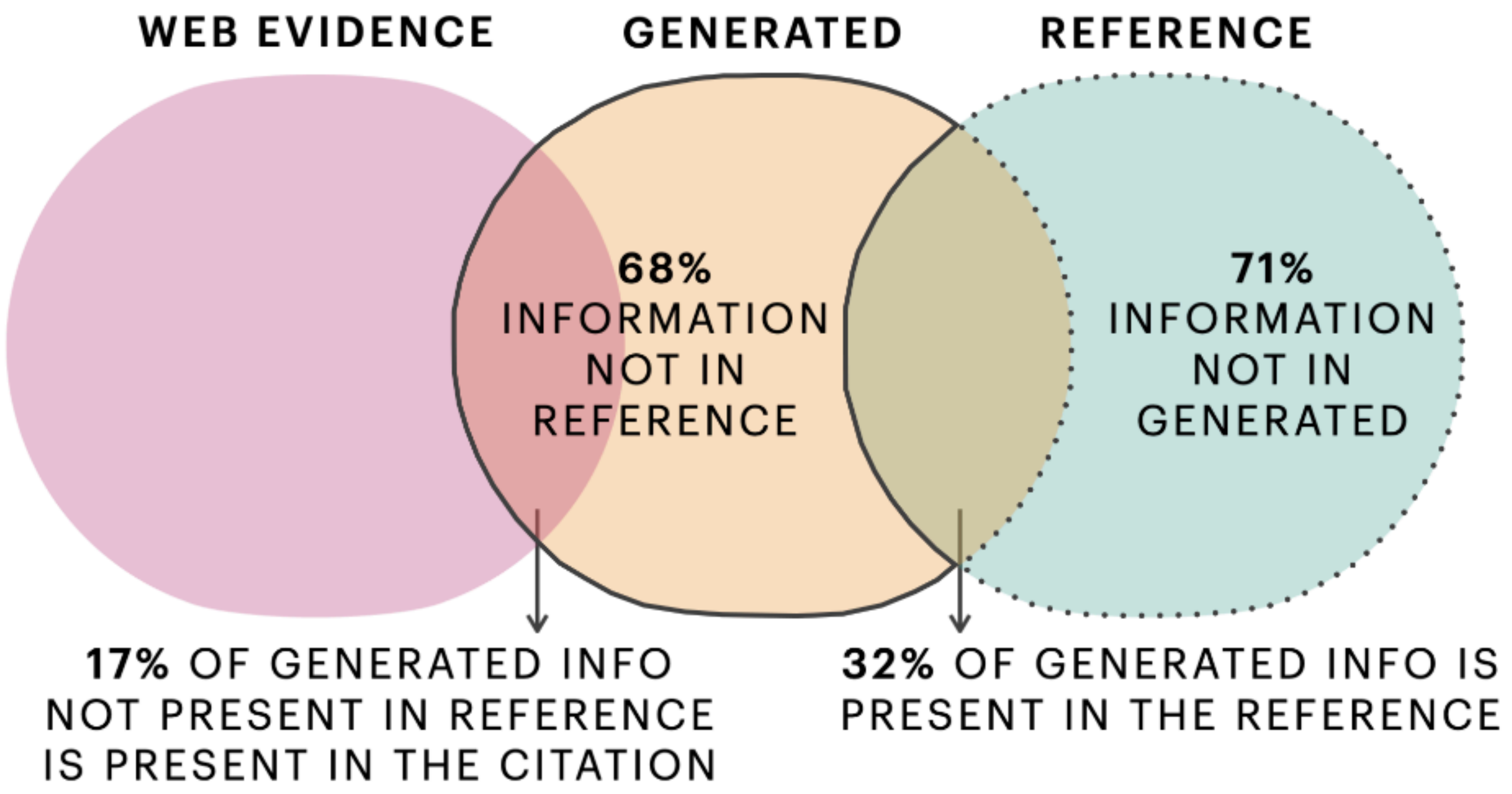
Human Evaluation of factuality

- 71% of information in the reference text is not in the generated text.



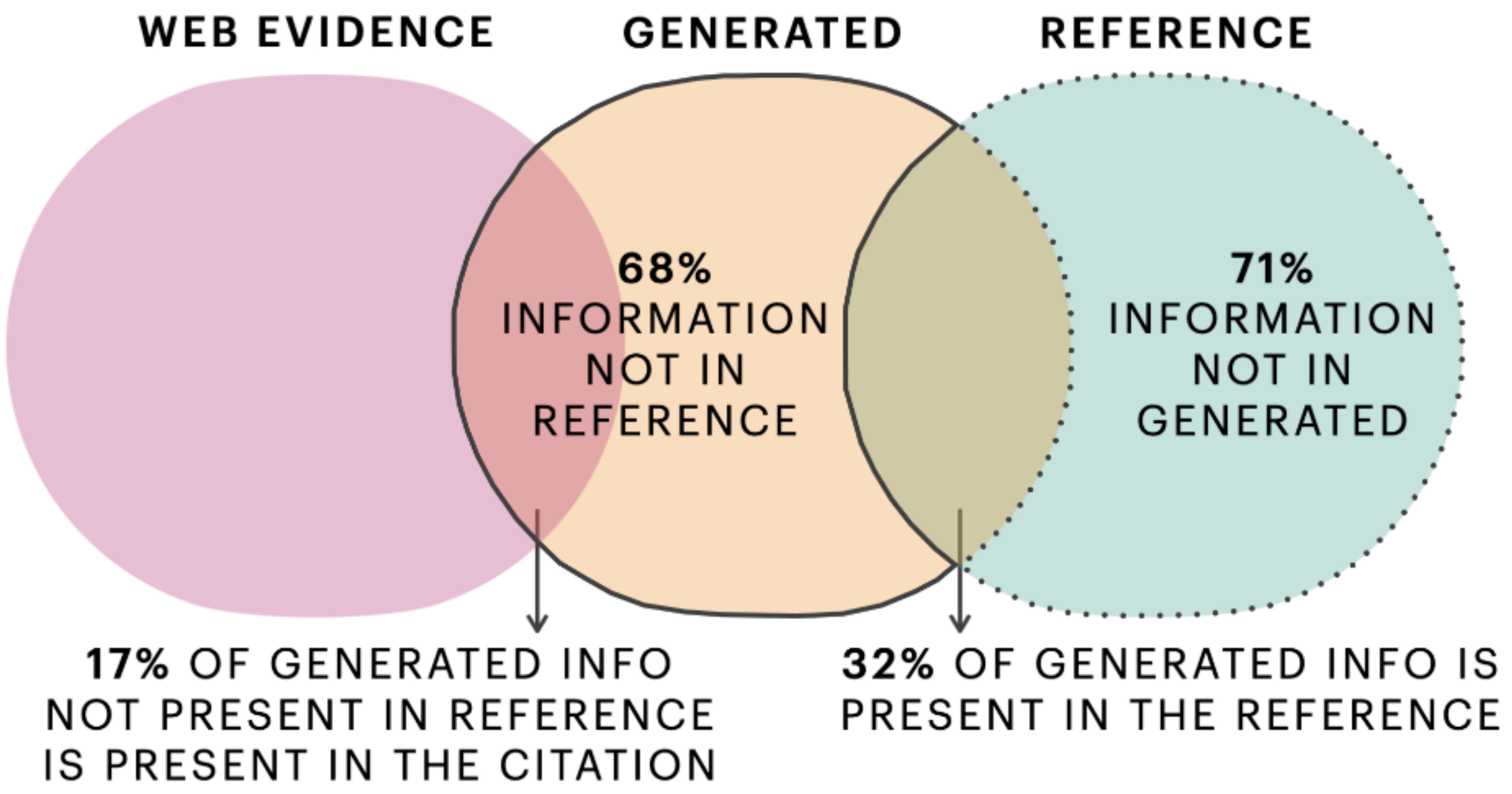
Human Evaluation of factuality

- 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



Human Evaluation of factuality

- 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

WikiSum Evaluation Dataset	
Average Number of Sections	7.2
Average Length of a Section	151.0
Average Length of Total Article	892.3
Avg overlap of Web Hits and Biography	39.8%
Our Evaluation Dataset	
Average Number of Sections	5.8
Average Length of a Section	132.3
Average Length of Total Article	765.9
Avg Number of Web Hits (max 20)	18.1
Avg overlap of Web Hits and Biography	24.9%

Data Bias

Woman Biographies are shorter

- Fewer sections
- Shower sections
- Fewer words

WikiSum Evaluation Dataset	
Average Number of Sections	7.2
Average Length of a Section	151.0
Average Length of Total Article	892.3
Avg overlap of Web Hits and Biography	39.8%
Our Evaluation Dataset	
Average Number of Sections	5.8
Average Length of a Section	132.3
Average Length of Total Article	765.9
Avg Number of Web Hits (max 20)	18.1
Avg overlap of Web Hits and Biography	24.9%

Data Bias

Less Web evidence for women

- Fewer documents: 18 in average

WikiSum Evaluation Dataset	
Average Number of Sections	7.2
Average Length of a Section	151.0
Average Length of Total Article	892.3
Avg overlap of Web Hits and Biography	39.8%
Our Evaluation Dataset	
Average Number of Sections	5.8
Average Length of a Section	132.3
Average Length of Total Article	765.9
Avg Number of Web Hits (max 20)	18.1
Avg overlap of Web Hits and Biography	24.9%

Data Bias

Less web evidence for woman

Smaller overlap between web evidence and Wikipedia biographies

WikiSum Evaluation Dataset	
Average Number of Sections	7.2
Average Length of a Section	151.0
Average Length of Total Article	892.3
Avg overlap of Web Hits and Biography	39.8%
Our Evaluation Dataset	
Average Number of Sections	5.8
Average Length of a Section	132.3
Average Length of Total Article	765.9
Avg Number of Web Hits (max 20)	18.1
Avg overlap of Web Hits and Biography	24.9%

Fewer Evidence, Lower Quality Output

Model	WikiSum Test	Women	Scientists	Women in Asia	Women in Africa
BART Pretraining	19.0	17.4	18.2	16.7	16.4
+ Retrieval	21.4	18.8	19.3	17.9	17.1
+ Caching	21.8	19.3	19.7	18.4	17.3

- 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 3.5
29/12/2023

ChatGPT does not know Ashley McKenzie

Winner of the 2016 Toronto international film festival

SY **You**

write a wikipedia biographie for ashley mckenzie

ChatGPT

I'm sorry, but as of my last knowledge update in January 2022, I don't have information on a specific individual named Ashley McKenzie who would warrant a Wikipedia-style biography. If Ashley McKenzie is a private individual or someone who has not gained significant public recognition up to that point, it would be inappropriate for me to generate false or misleading information.

If Ashley McKenzie is a public figure, artist, scientist, or anyone who has made contributions to society and gained notable recognition after my last update, I won't have that information. To create a fictional biography, I'd need some details or characteristics you'd like included. Alternatively, if Ashley McKenzie is a real person who gained recognition after 2022, I recommend checking more recent and reliable sources for accurate information to include in a biography.

ChatGPT 3.5
29/12/2023

ChatGPT knows
Henri Poincaré

SY **You**
write a wikipedia biographie for henri poincaré

ChatGPT
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.

Thanks !