Verbalising Graphs into High-, Medium- and Low-Resource Languages

Chaire IA xNLG: Generating from Multiple Sources into Multiple Languages

Claire Gardent

CNRS / LORIA

Anja Belz, Thiago Castro-Ferreira, Liam Cripwell, Angela Fan, ALbert Gatt, Nikolai Ilinyskh, Chris van der Lee, Simon Mille, Diego Moussalem, Yannick Parmentier, Laura Perez-Beltrachini, Anastasia Shimorina, William Soto-Martinez

Graphs are frequent Data Structures

- Knowledge Graphs

Graphs are frequent Data Structures

- Knowledge Graphs
- RDF Stores

Graphs are frequent Data Structures

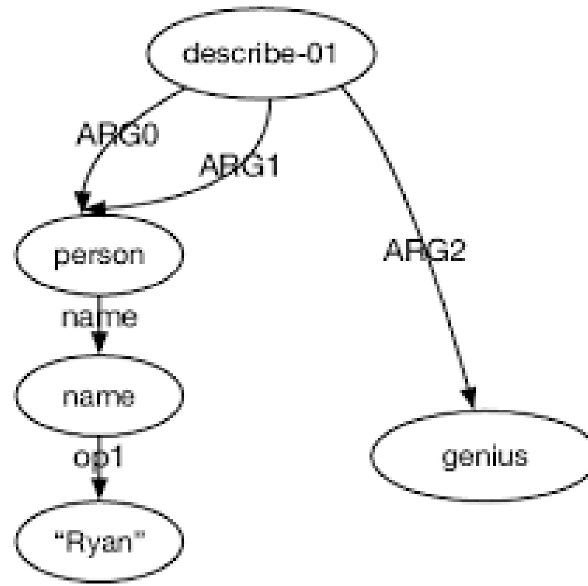
- Knowledge Graphs
- RDF Stores
- Meaning Representations

Graphs are frequent Data Structures

- Knowledge Graphs
- RDF Stores
- Meaning Representations
- Tabular Data

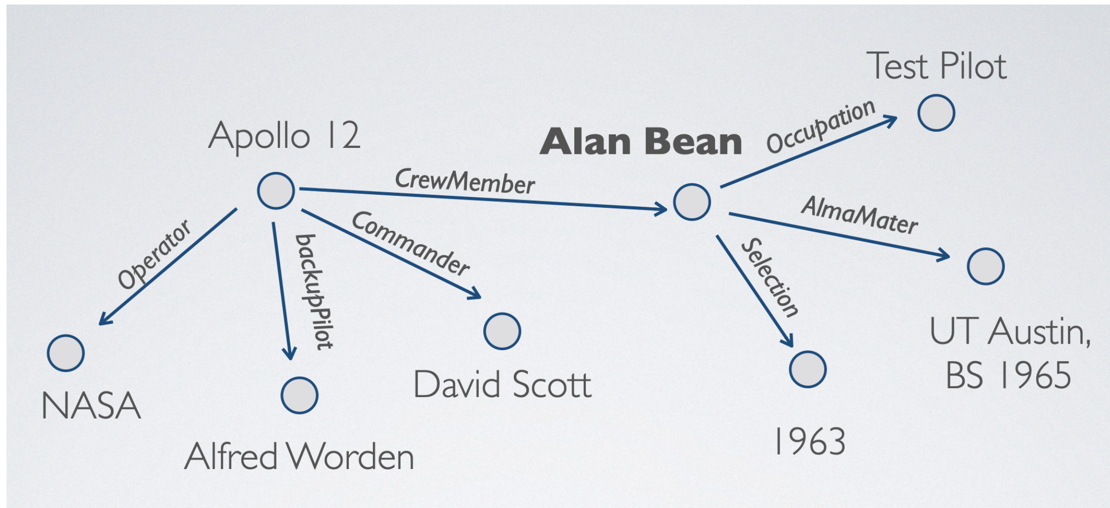
Abstract Meaning Representation (AMR)

Ryan describes himself as a genius



Knowledge Graphs

RDF Graph



English Text

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

Challenges

- Structured input has a different surface form

Challenges

- Structured input has a different surface form
- Structured Input is underspecified

Challenges

- Structured input has a different surface form
- Structured Input is underspecified
- Lack of parallel Graph/text data

Challenges

- Structured input has a different surface form
- Structured Input is underspecified
- Lack of parallel Graph/text data
- Decoding into languages with varied morphology and word order

Outline

Multilingual Models

- AMR $\Rightarrow$ 21 EU Languages

Outline

Multilingual Models

- AMR $\Rightarrow$ 21 EU Languages

Pre-trained Multilingual Models

- RDF $\Rightarrow$ English, Russian

Outline

Multilingual Models

- AMR $\Rightarrow$ 21 EU Languages

Pre-trained Multilingual Models

- RDF $\Rightarrow$ English, Russian

Parameter Efficient Fine Tuning

- RDF $\Rightarrow$ Breton, Welsh, Irish, Maltese

Outline

Multilingual Models

- AMR $\Rightarrow$ 21 EU Languages

Pre-trained Multilingual Models

- RDF $\Rightarrow$ English, Russian

Parameter Efficient Fine Tuning

- RDF $\Rightarrow$ Breton, Welsh, Irish, Maltese
- AMR $\Rightarrow$ 6 High- and 6 Low-Resource Languages

Outline

Multilingual Models

- AMR $\Rightarrow$ 21 EU Languages

Pre-trained Multilingual Models

- RDF $\Rightarrow$ English, Russian

Parameter Efficient Fine Tuning

- RDF $\Rightarrow$ Breton, Welsh, Irish, Maltese
- AMR $\Rightarrow$ 6 High- and 6 Low-Resource Languages

Prompting LLMs

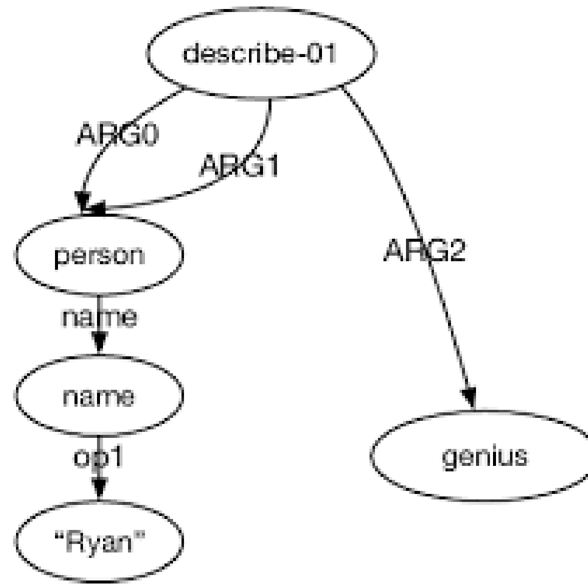
- RDF $\Rightarrow$ English, Russian

AMR $\Rightarrow$ 21 EU Languages

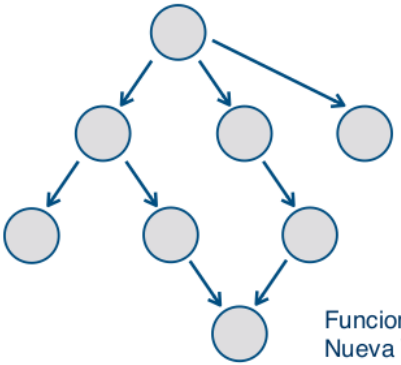
Fan and Gardent EMNLP 2020

Abstract Meaning Representation (AMR)

Ryan describes himself as a genius



AMR → 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.

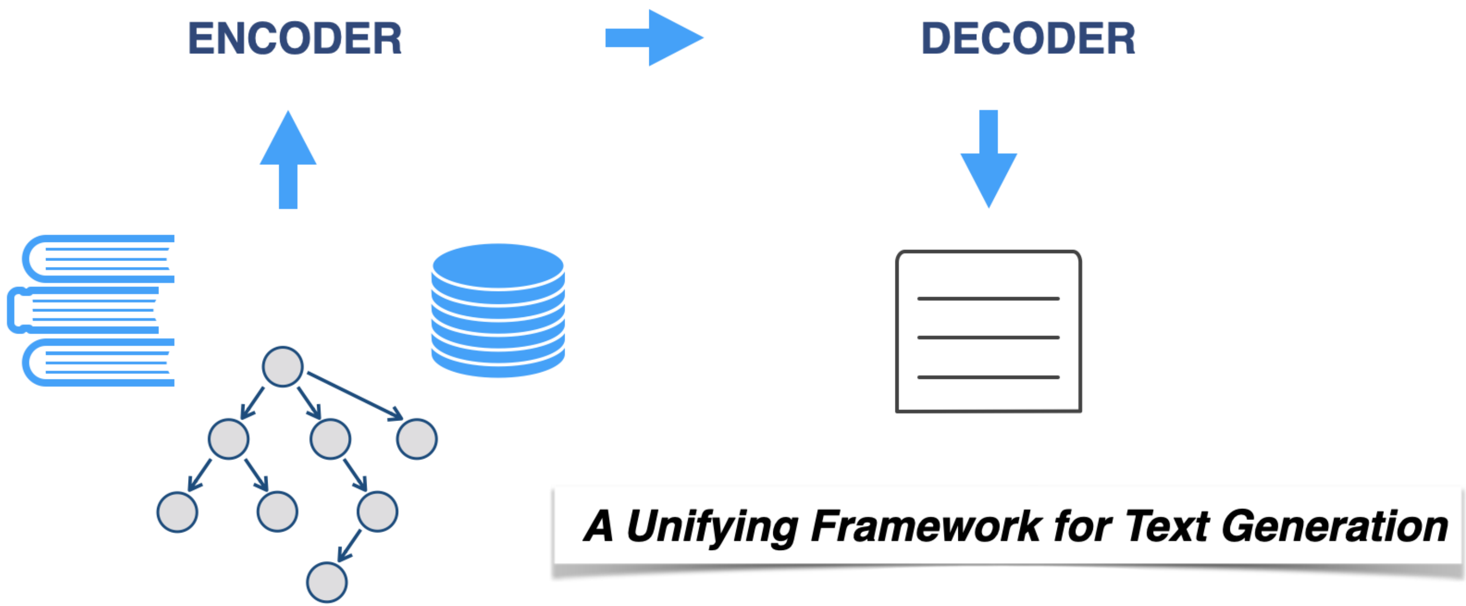
US officials held an expert group meeting in January 2002 in New York.

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

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

Romance, Germanic, Slavic, Uralic

The Encoder-Decoder Framewok



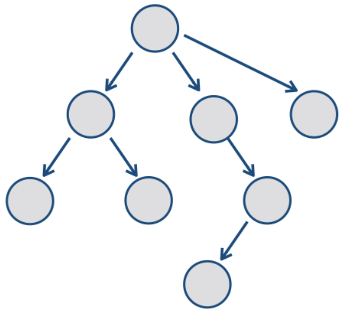
Two networks

- The encoder produces a continuous representation of the input
- The decoder generates a text from this representation

AMR Encoding

- Transformer encoder
- Linearise (and simplify) AMRs
- Graph structure
  - Node: token + distance from root + branch
- Pretraining (Masked Language Modelling objective)
  - on 30M silver AMRs

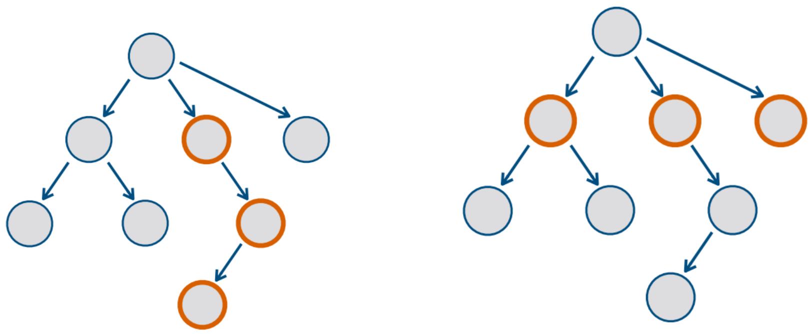
Linearising



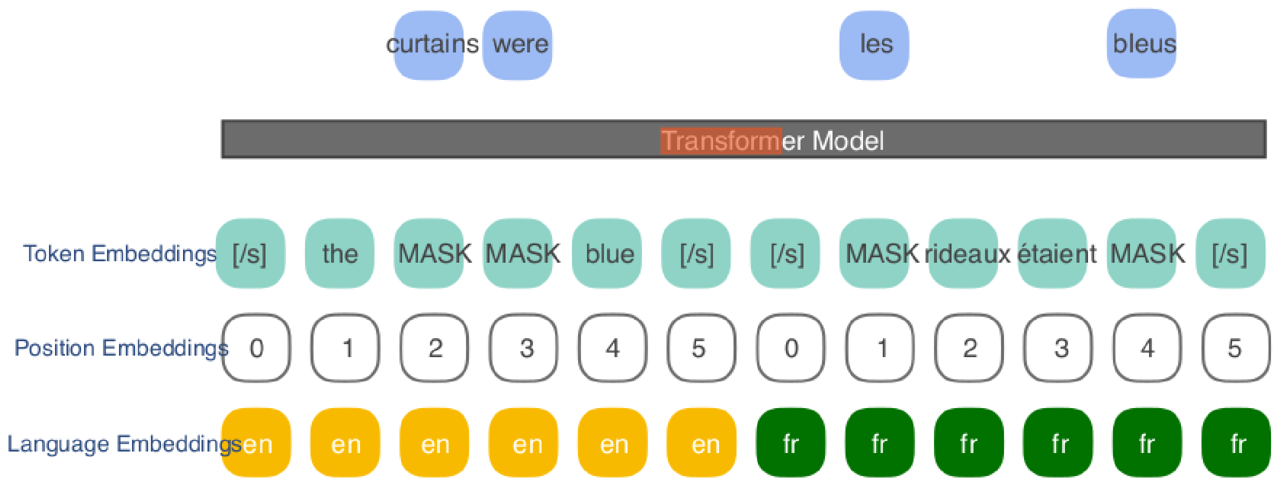
hold
:ARG0 person : ARG0-of have-org-role :ARG1 :op1 United :op2 States :ARG2 official
:ARG1 meet :ARG0 person :ARG1-of expert :ARG2-of group
:time date-entity :year 2002 :month 1
:location city :op1 New :op2 York

Encoding Structure

Add branch and sibling embeddings (aka positional embedding for graphs)

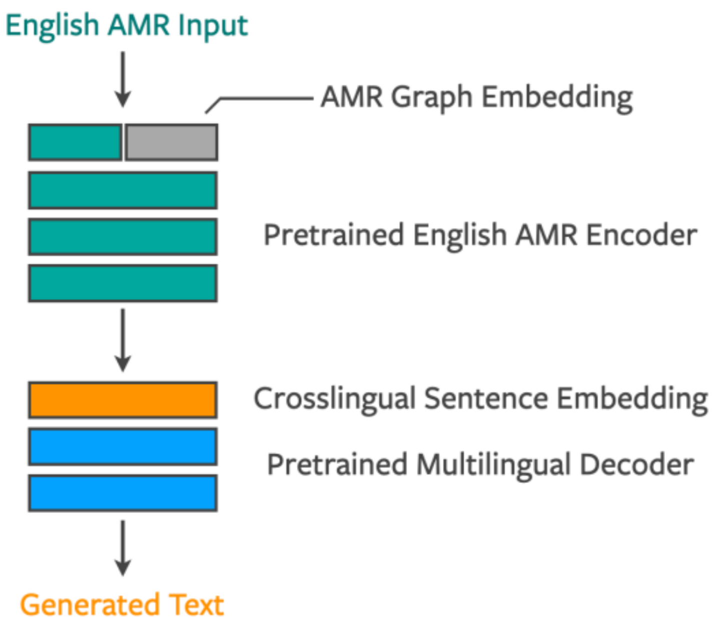


Multilingual Decoding



- Crosslingual embeddings (XLM Sentence Piece Model and Vocabulary)
- Language Models pretrained on 30M sentences (for each language)

Leveraging Pretraining



- Encoder: pretraining on Silver AMRs
- Decoder: language model pretraining

Multilingual decoding

- Prefix each training instance with a control token
- Trained on multilingual Europarl data

Training Data

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

French	Spanish	Slovak	Bulgarian	Swedish
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 januári 2002 v New Yorku.	Американските служители проведоха среща на експертна група през януари 2002 г. в Ню Йорк.	Amerikanska tjänstemän höll ett expertgruppsmöte i januari 2002 i New York.

- Europarl: 21 Languages
- Input AMR: create AMR structure with JAMR parser

Test Data

- Silver AMR: 21 languages, Europarl
- Gold AMR: 4 languages (Damonte and Cohen, NAACL 2018)

Comparison: Bilingual vs Multilingual

Bilingual Baseline

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: Bilingual vs Multilingual

Bilingual Baseline

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 Model

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: Bilingual vs Multilingual

Bilingual Baseline

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 Model

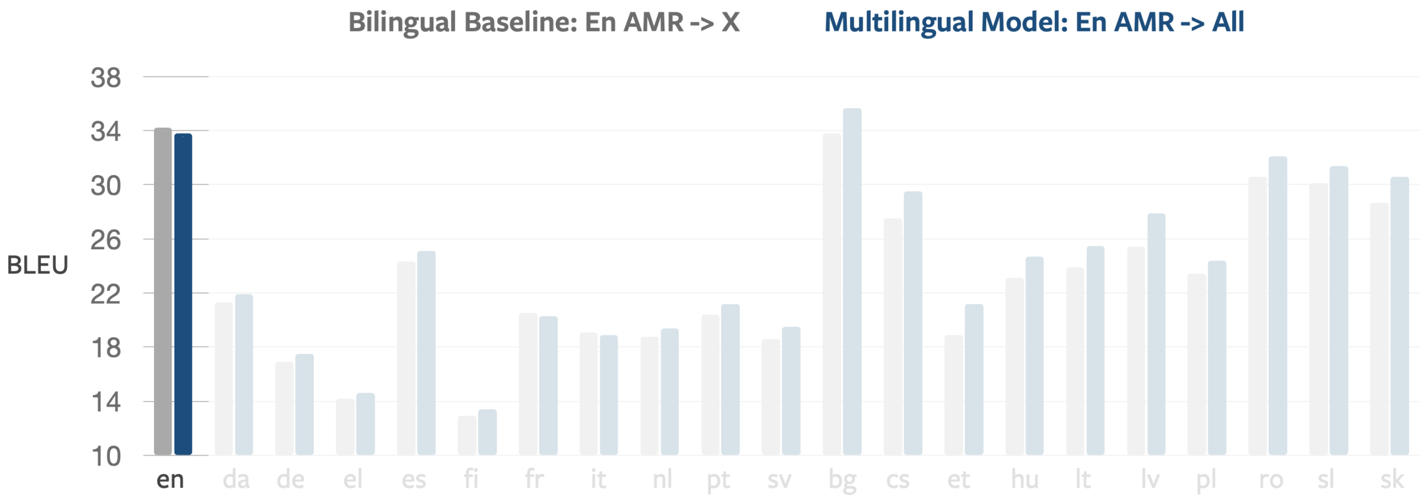
SV

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

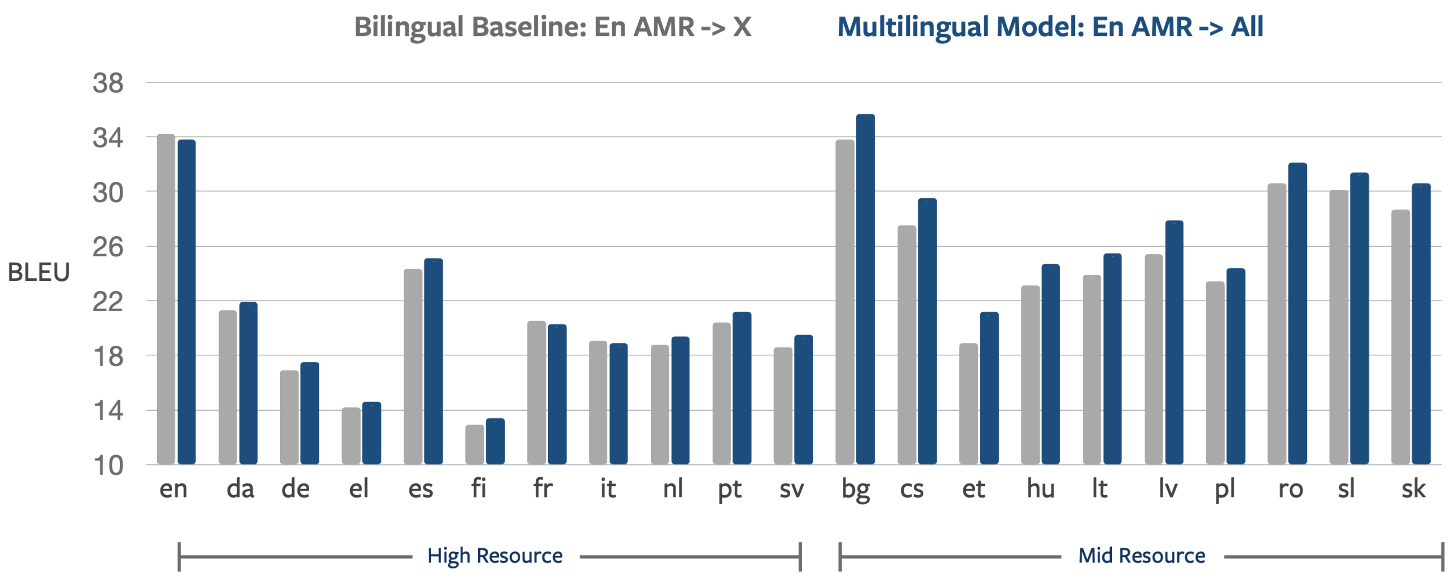
↓

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

Results: Europarl

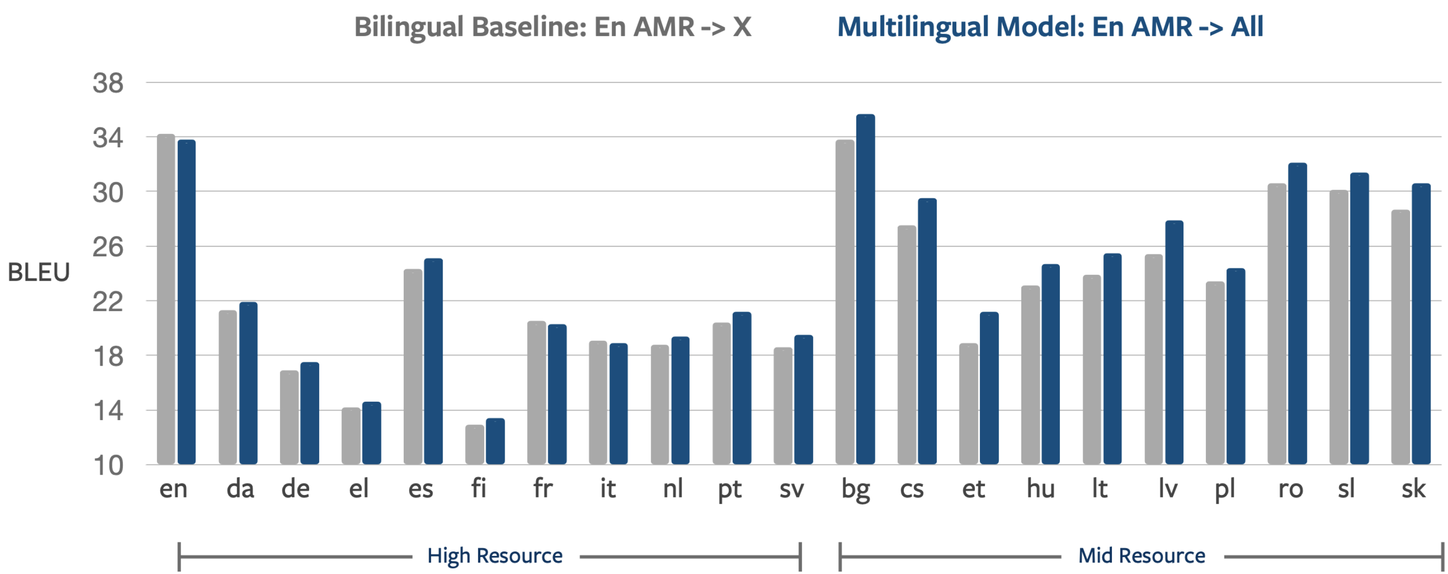


Results: Europarl



The multilingual model generally outperforms monolingual models

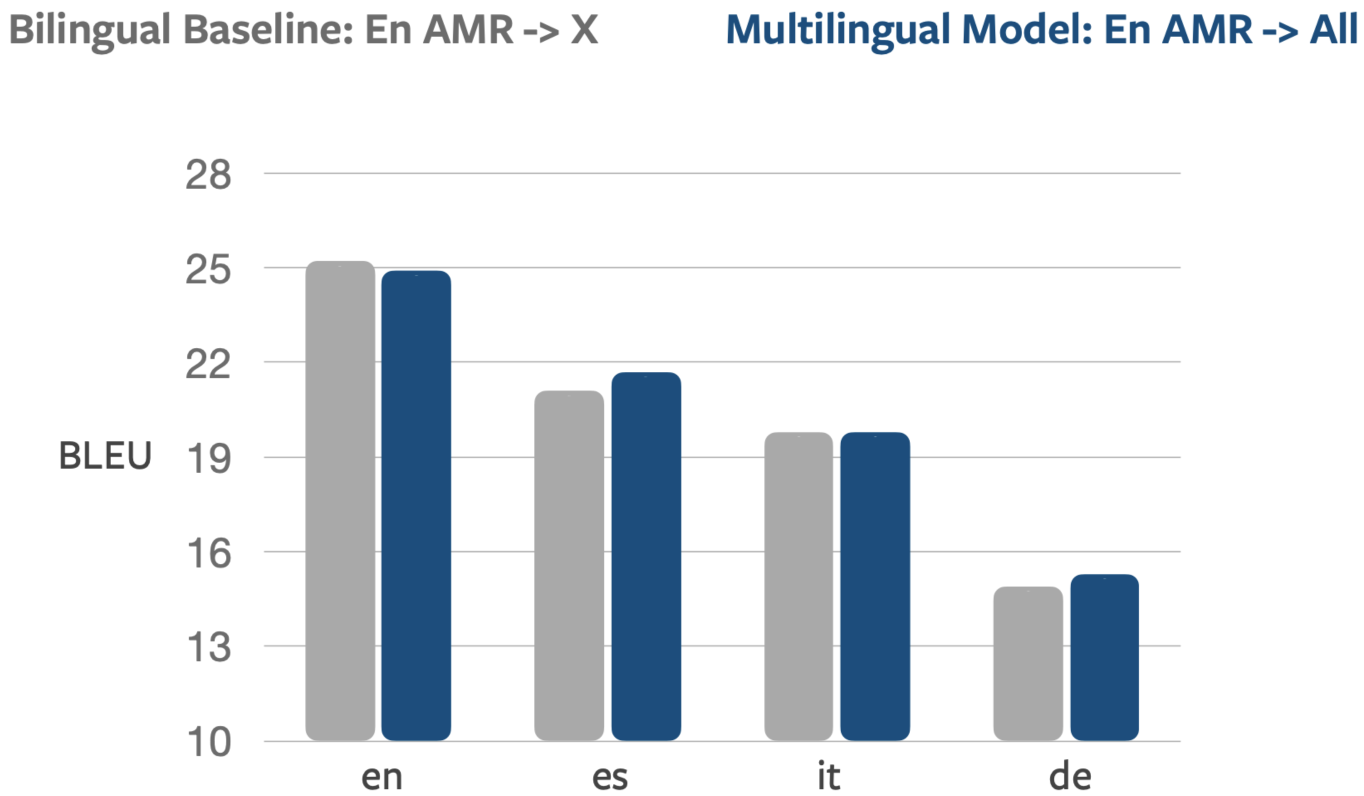
Results: Europarl



The multilingual model generally outperforms monolingual models

The difference is stronger on Mid-Resource Languages

Results: Gold AMR



The difference also holds when generating from gold AMRs

Comparison: Hybrid vs Multilingual

Hybrid Translation Model

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 to English

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

Multilingual Model

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: Hybrid vs Multilingual

Hybrid Translation Model

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 to English

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

↓ Translation Model

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

Multilingual Model

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.

Generate: AMR → English

Comparison: Hybrid vs Multilingual

Hybrid Translation Model

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 to English

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

↓ Translation Model

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

Generate: AMR $\rightarrow$ English

Translate: English $\rightarrow$ X

Multilingual Model

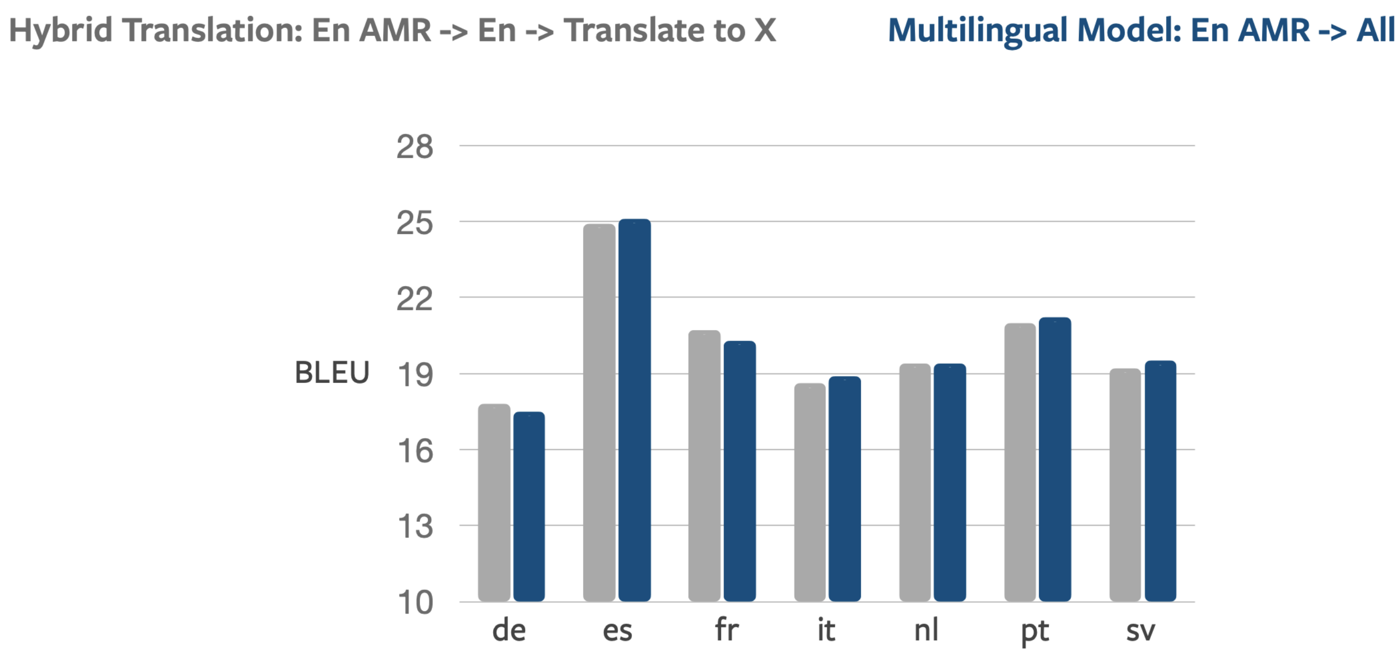
fr

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

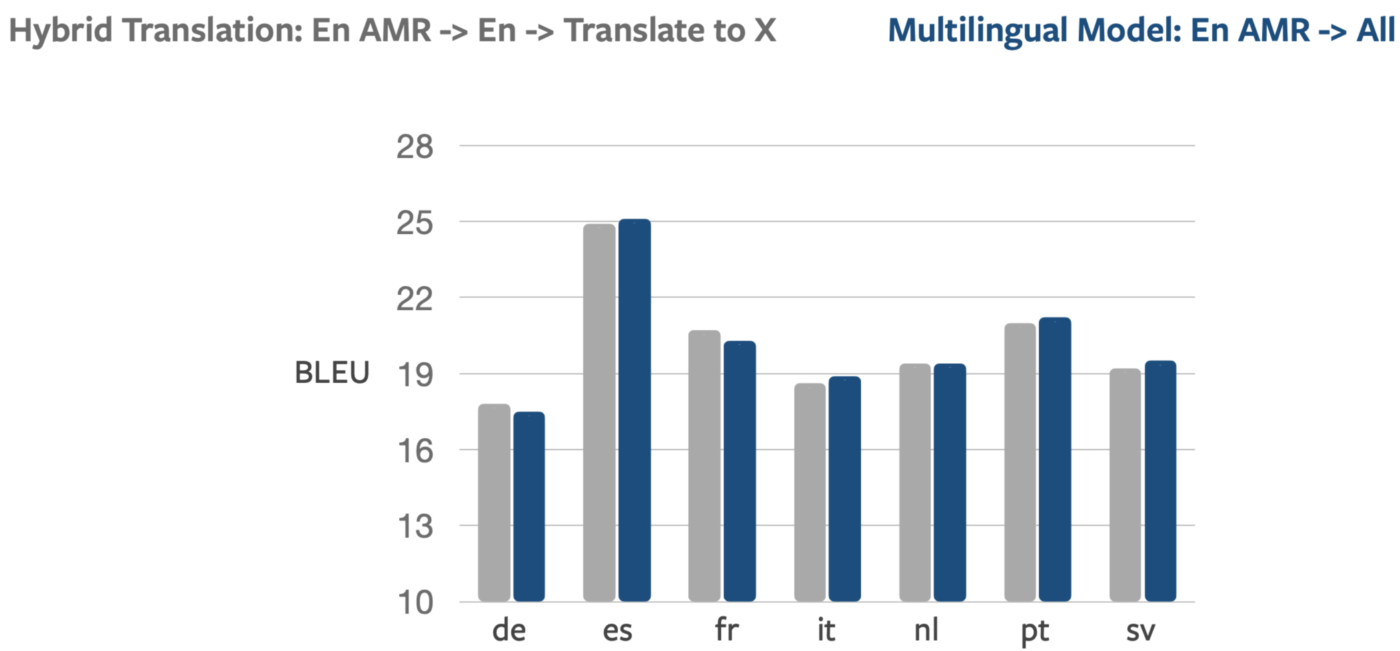
↓

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

Comparison: Hybrid vs Multilingual



Comparison: Hybrid vs Multilingual



The multilingual model performs similarly to the Gen&Translate pipeline while trained on much fewer data

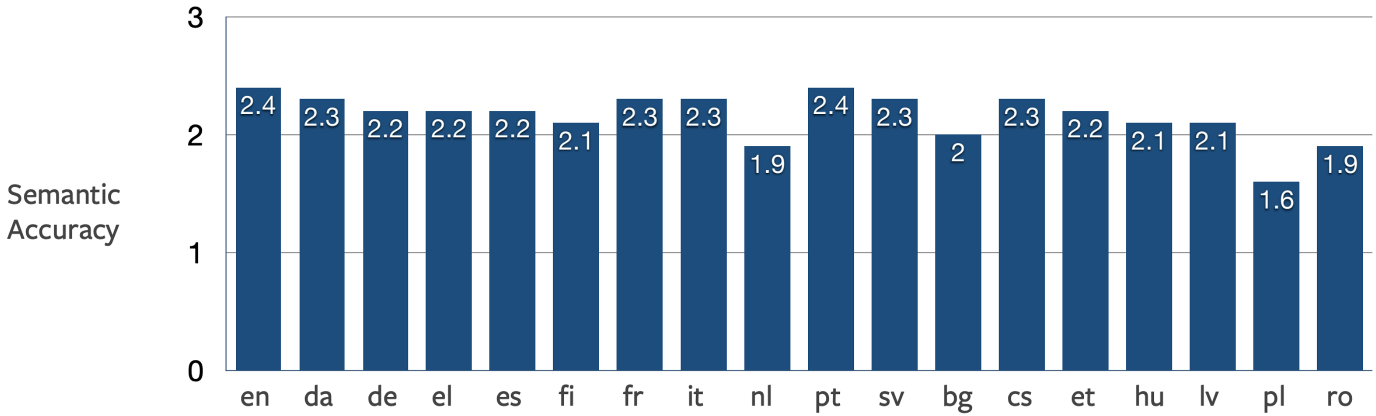
Human Evaluation

- Semantic Accuracy:
  Does the generated text correctly paraphrase the reference?
- Morphology:
  Is the morphology correct? Are agreement constraints e.g., verb/subject, noun/adjective respected?
- Word Order:
  Is the word order natural sounding?

Human Evaluation

- Evaluators: colleagues from NLP mailing lists
- 50 sentences per language
    - Half low BLEU
    - Half high BLEU

Human Evaluation: Semantic Accuracy



- 3 = correct
- 2 = minor differences
- 1 = incorrect

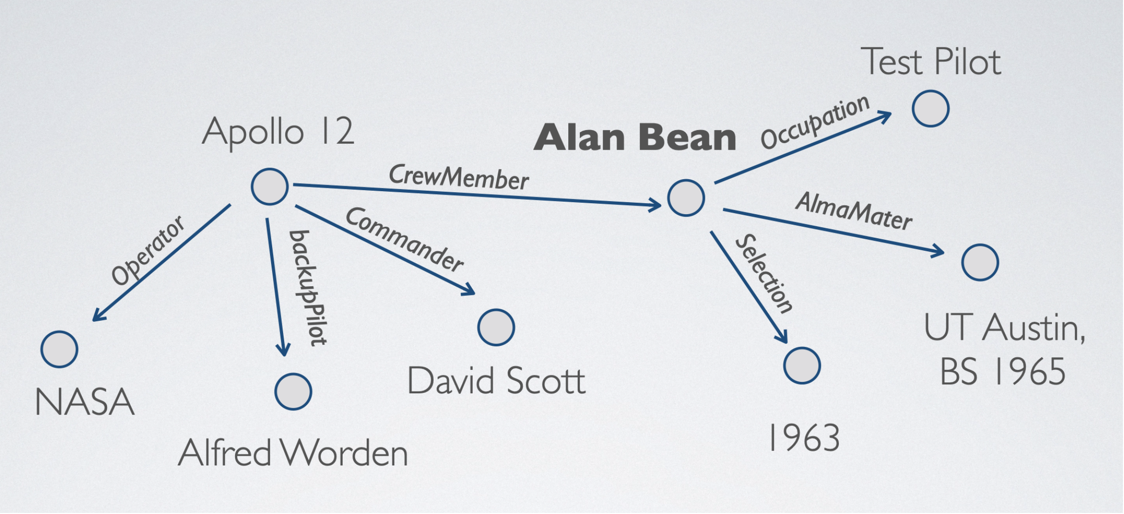
Pre-training and Multilingual techniques permits bridging the gap between English-Centric AMR graphs and target languages with varied syntax and morphology

Knowledge Graphs $\Rightarrow$ English, Russian

Gardent et al. ACL 2017, Castro-Ferreira et al. INLG 2020

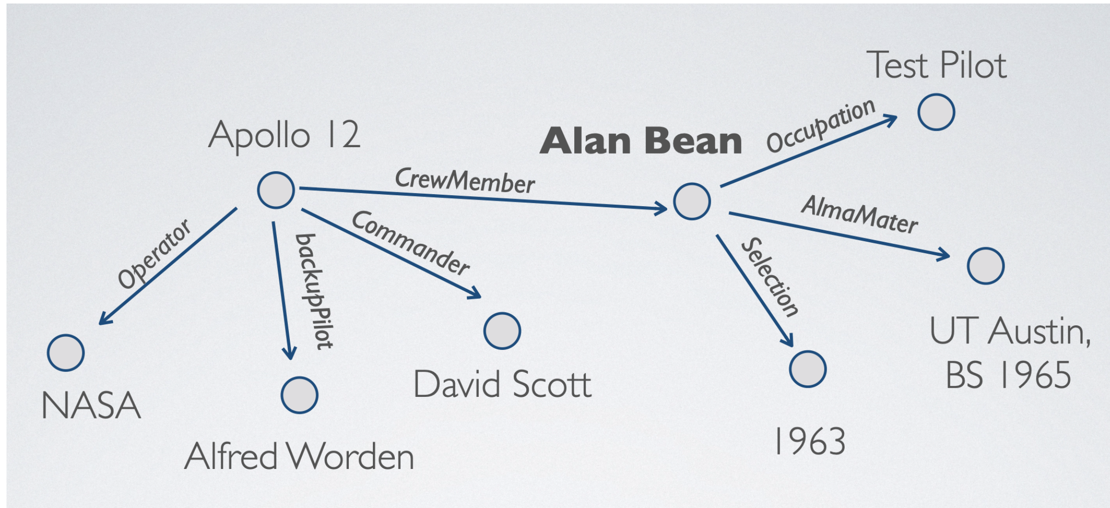
The WebNLG Challenge

RDF Graph



The WebNLG Challenge

RDF Graph



English Text

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

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 entity of various categories.
- English texts are crowdsourced

WebNLG 2017: RDF $\Rightarrow$ 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:

- Astronaut, University, Monument, Building, Comics Character, Food, Airport, SportsTeam, City and WrittenWork

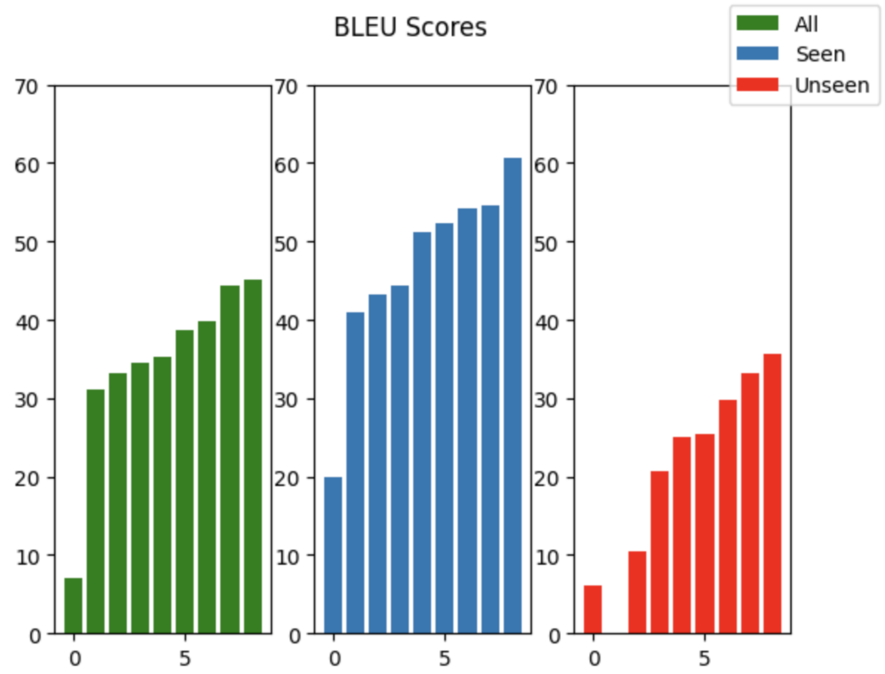
5 **unseen** categories:

- Athlete, Artist, MeanOfTransportation, CelestialBody, Politician

WebNLG 2017: RDF ⇒ English

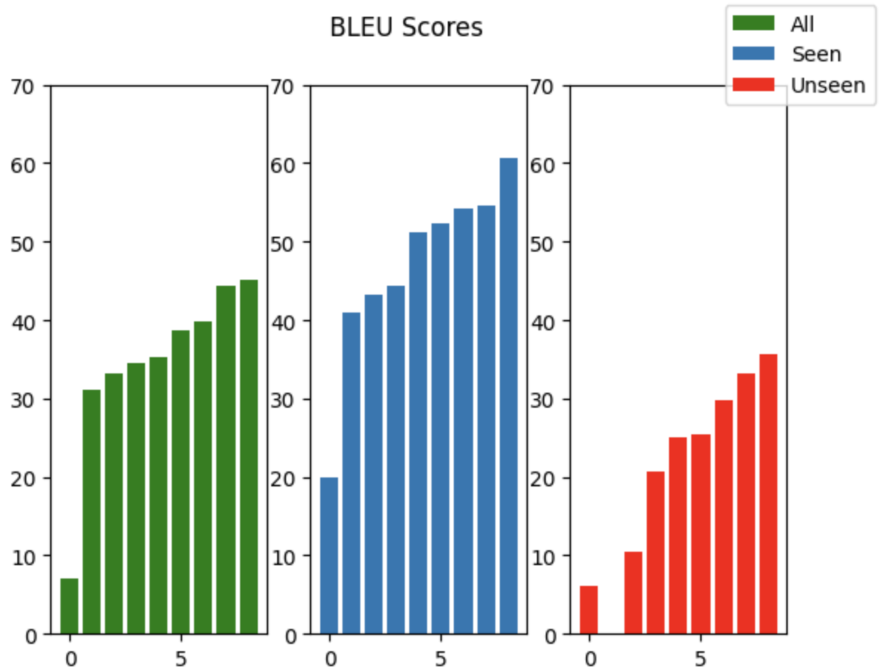
- 6 participants, 10 systems
- Models: 3 rule-based, 1 SMT, 5 neural

WebNLG 2017: RDF ⇒ English



ALL: 7.07 - 45.13, **Seen**: 19.87 - 60.54, **Unseen**: 5.13 - 35.7

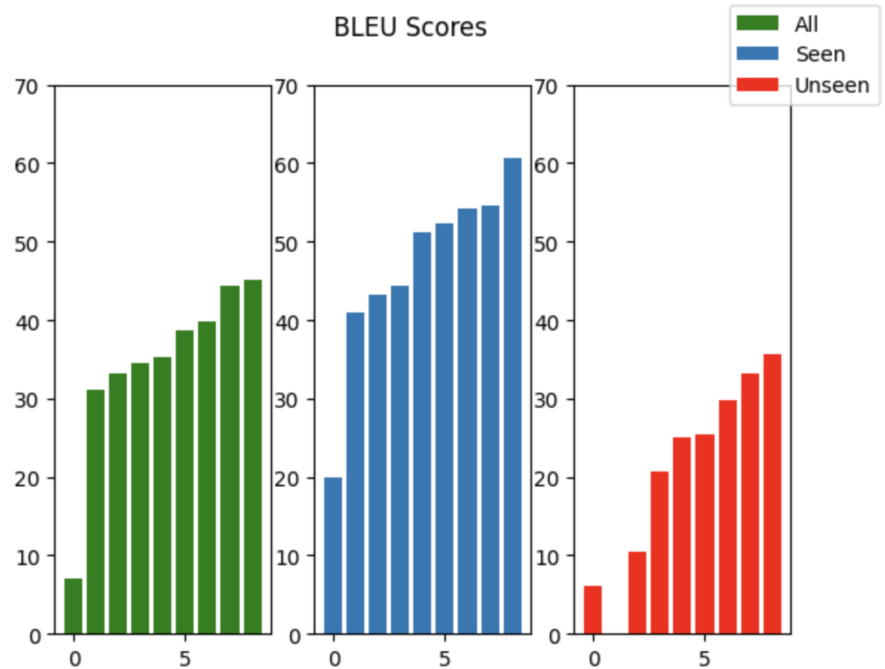
WebNLG 2017: RDF ⇒ English



ALL: 7.07 - 45.13, **Seen**: 19.87 - 60.54, **Unseen**: 5.13 - 35.7

Strong differences between models

WebNLG 2017: RDF ⇒ English



ALL: 7.07 - 45.13, **Seen**: 19.87 - 60.54, **Unseen**: 5.13 - 35.7

Strong differences between models

All models degrades on Unseen Data

WebNLG 2020

Natural Language Generation

- RDF $\Rightarrow$ English
- RDF $\Rightarrow$ Russian

WebNLG 2020

Natural Language Generation

- RDF $\Rightarrow$ English
- RDF $\Rightarrow$ Russian

Semantic Parsing

- English $\Rightarrow$ RDF
- Russian $\Rightarrow$ RDF

WebNLG 2020: RDF ⇒ English

	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

Astronaut, University, Monument, Building, Comics Character, Food, Airport, SportsTeam, City, WrittenWork, Athlete, Artist, CelestialBody, MeanOfTransportation, Politician, Company

3 **unseen** categories:

Film, Scientist, and MusicalWork

Unseen entities: graphs from seen categories, but unseen root entity

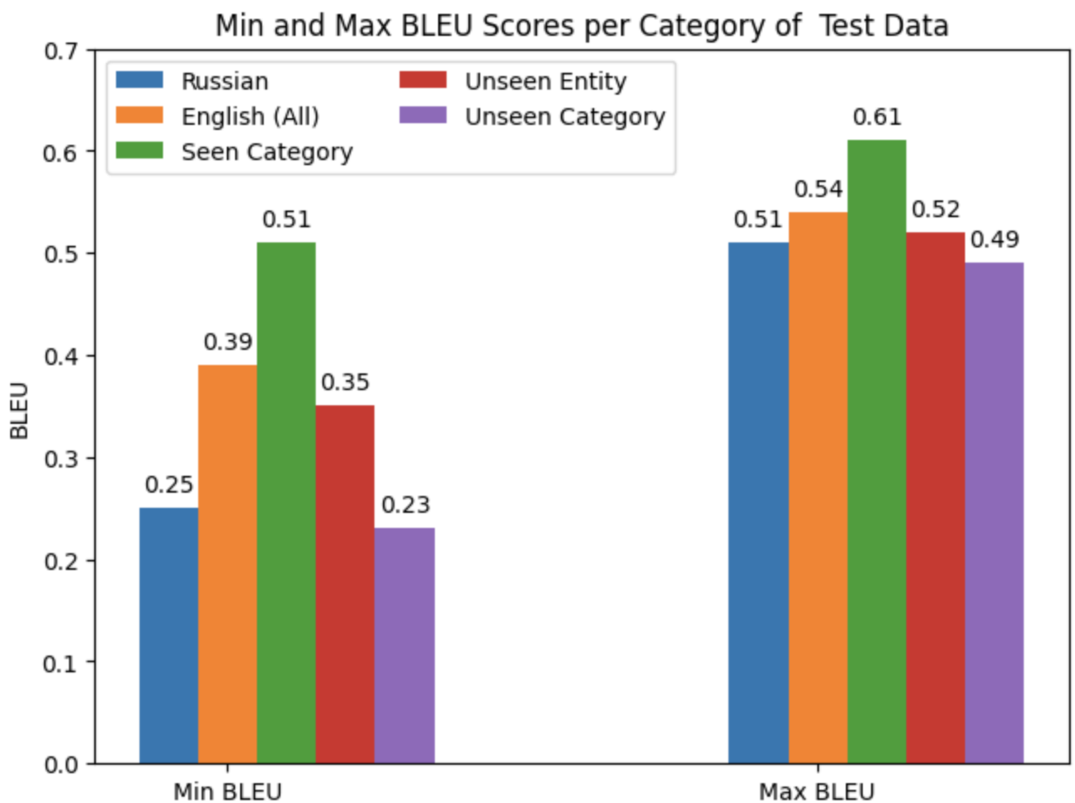
E.g., *Nie Haisheng* in category *Astronaut*

WebNLG 2020: Participation

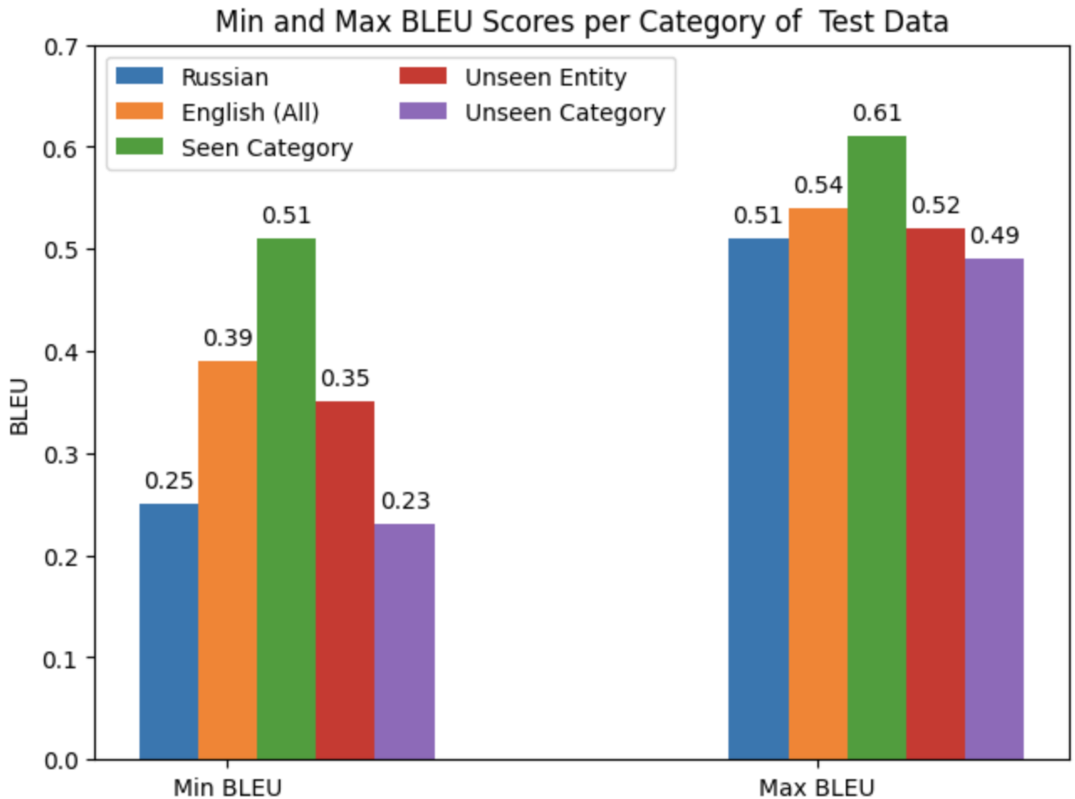
System	Affiliation	Country
MED	Sber AI Lab	Russia
RALI-UMONTRÉAL	Université de Montréal	Canada
ORANGE-NLG	Orange Labs	France
CUNI-UFAL	Charles University	Czechia
TGEN	AIST	Japan
BT5	Google	US
UPC-POE	Universitat Politècnica de Catalunya	Spain
DANGNT-SGU	Saigon University	Vietnam
HUAWEI	Huawei Noah's 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 teams submitted 48 system runs

WebNLG 2020: Results

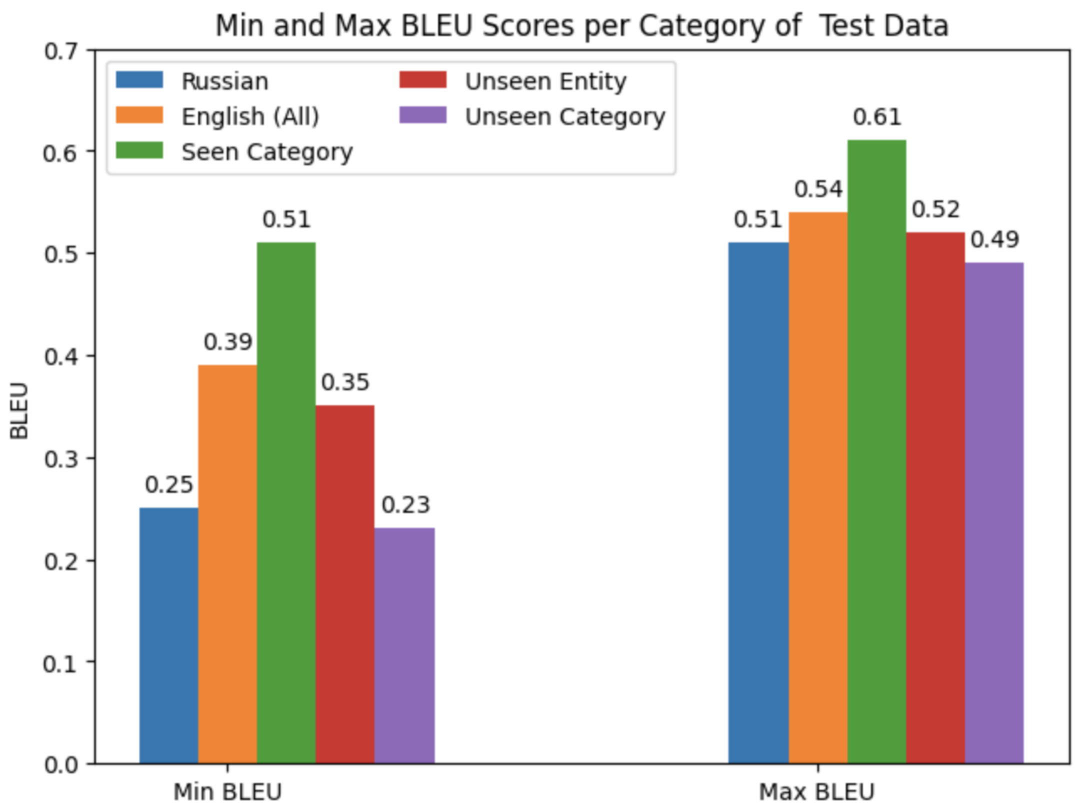


WebNLG 2020: Results



Results are better for English than for Russian

WebNLG 2020: Results



Results are better for English than for Russian

Pre-training improves results: +16 BLEU points for English w.r.t. 2017

WebNLG 2023: Low Resource Languages

Data

	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: Low Resource Languages

Data

	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

Participants

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	-	✓	✓	✓	-

WebNLG 2023: Low Resource Languages

Data

	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

Participants

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	-	✓	✓	✓	-

No training Data

WebNLG 2023: Pipeline NLG+MT Models

Participants

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	-	✓	✓	✓	-

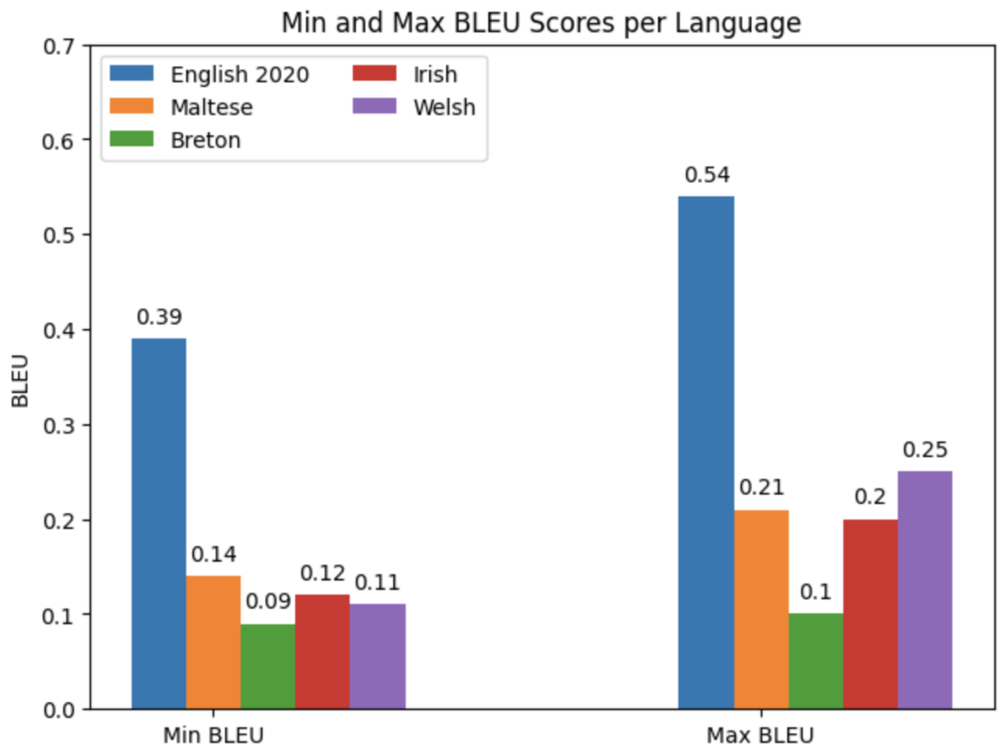
RDF $\Rightarrow$ English

- T5 or mT5 fine-tuned on English WebNLG data
- GPT3-5 in context learning, no fine-tuning

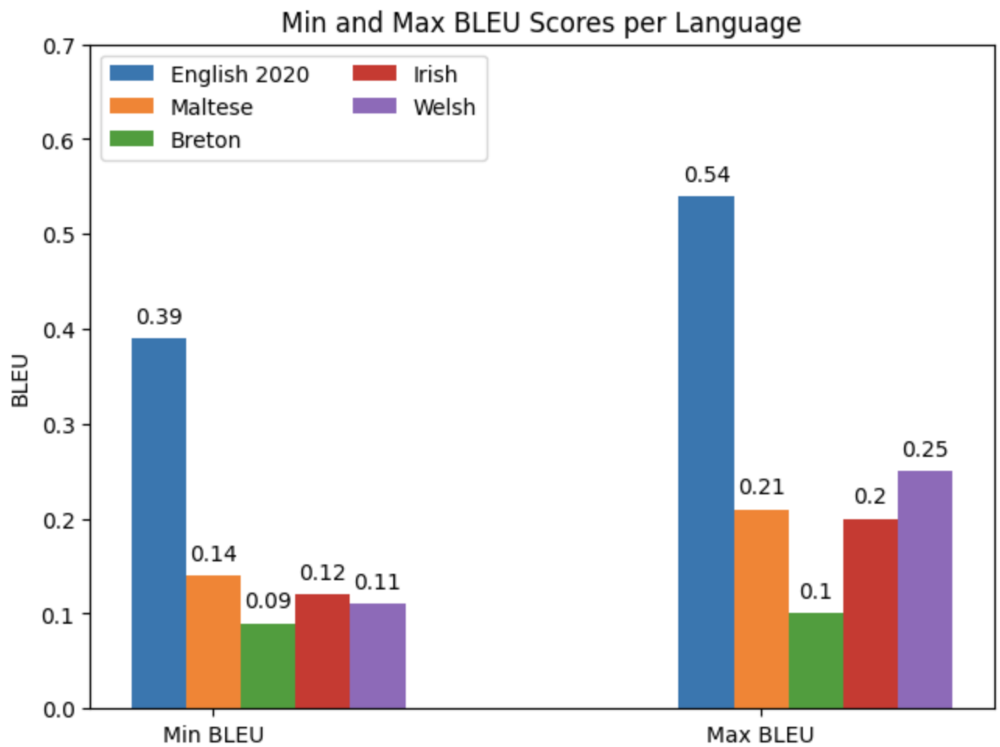
English $\Rightarrow$ LR Language

- Machine Translation: NLLB or Google Translate

WebNLG 2023: Results

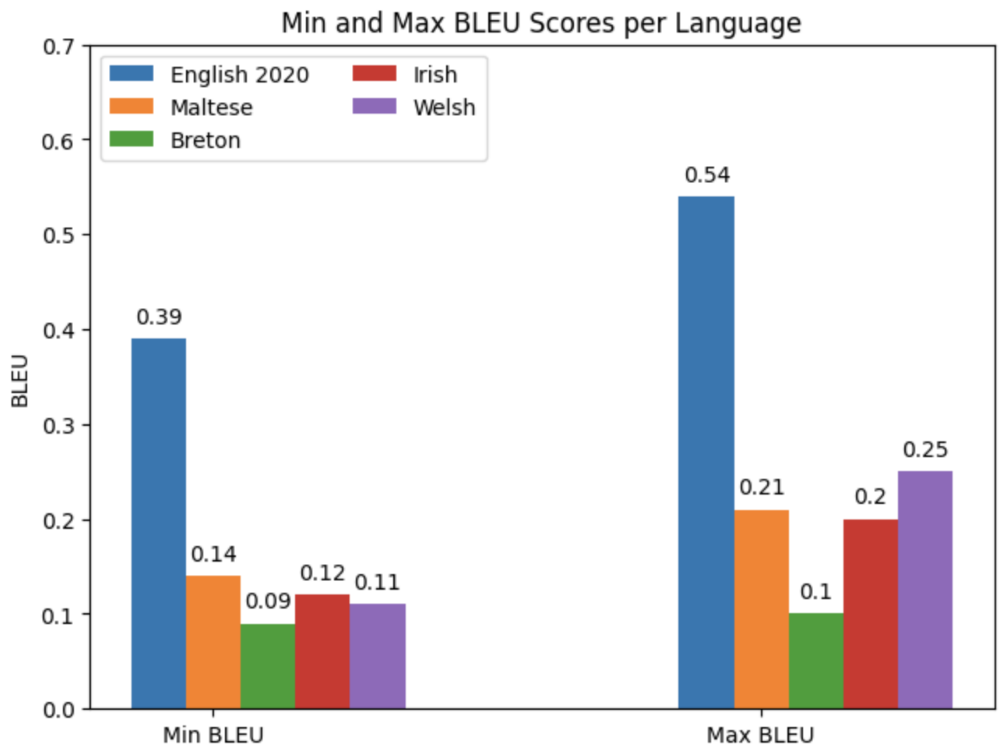


WebNLG 2023: Results



Strong degradation overall compared to results on English

WebNLG 2023: Results



Strong degradation overall compared to results on English

Very poor output for Breton

End-to-End RDF $\Rightarrow$ Celtic Language

Soto-Martinez et al. AACL-IJCNLP 2023

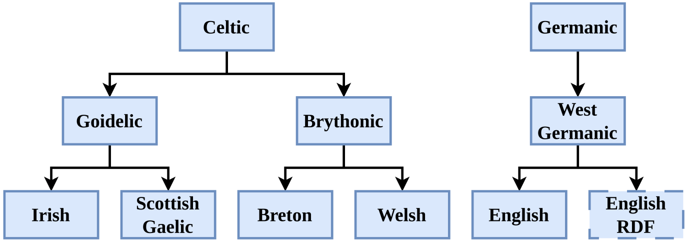
Pipeline vs. End-to-End

For Breton, there is no (good) MT system

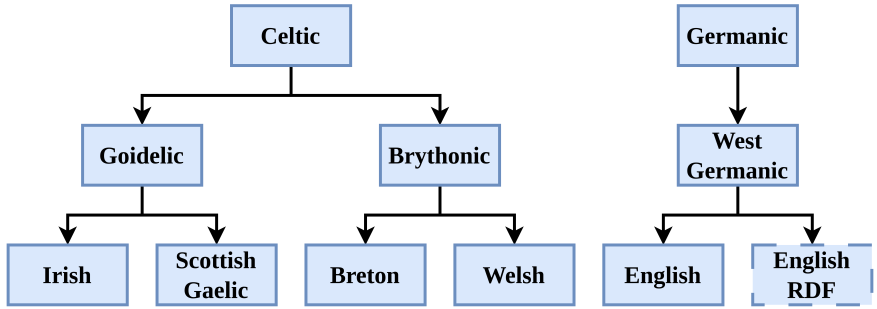
✗ NLG+MT pipeline

✓ Parameter Efficient Fine Tuning (PEFT)

- Soft-Prompt (Prefix Tuning)
- Structured to capture language relatedness



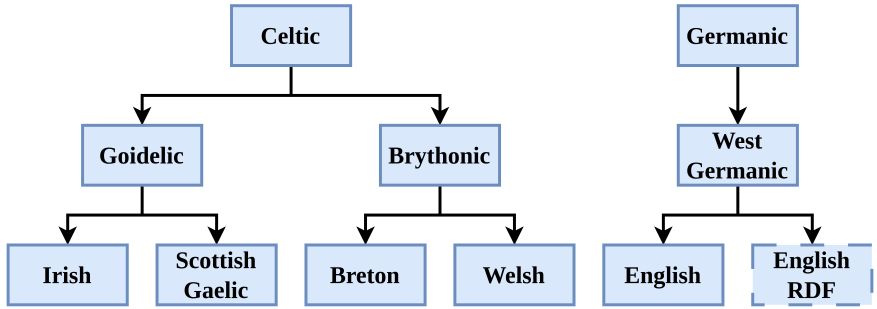
Phylogenetic Tree



Soft Prompt

The soft-prompt is decomposed into Family, Genus, and Language sub-prompts.

Phylogenetic Tree



Soft Prompt

The soft-prompt is decomposed into Family, Genus, and Language sub-prompts.

- ***Better Transfer:***
  - The sub-prompts for Family and Genus are updated each time a training instance from the corresponding Family/Genus is processed
  - Allows LR languages to benefit from the training data of their related languages
- ***Less Noise:***
  - Prevents the mixture of training data to introduce too much noise to the model.

Training and Testing

Step 1: Self-supervised Training (Language Models)

Trains the Soft Prompt on unsupervised, monolingual tasks

	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>

Training and Testing

Step 1: Self-supervised Training (Language Models)

Trains the Soft Prompt on unsupervised, monolingual tasks

	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 Dev RDF-to-Text data (RDF-to-Text Models)

Trains the RDF-to-Text Task sub-prompt for each target language

Training and Testing

Step 1: Self-supervised Training (Language Models)

Trains the Soft Prompt on unsupervised, monolingual tasks

	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>

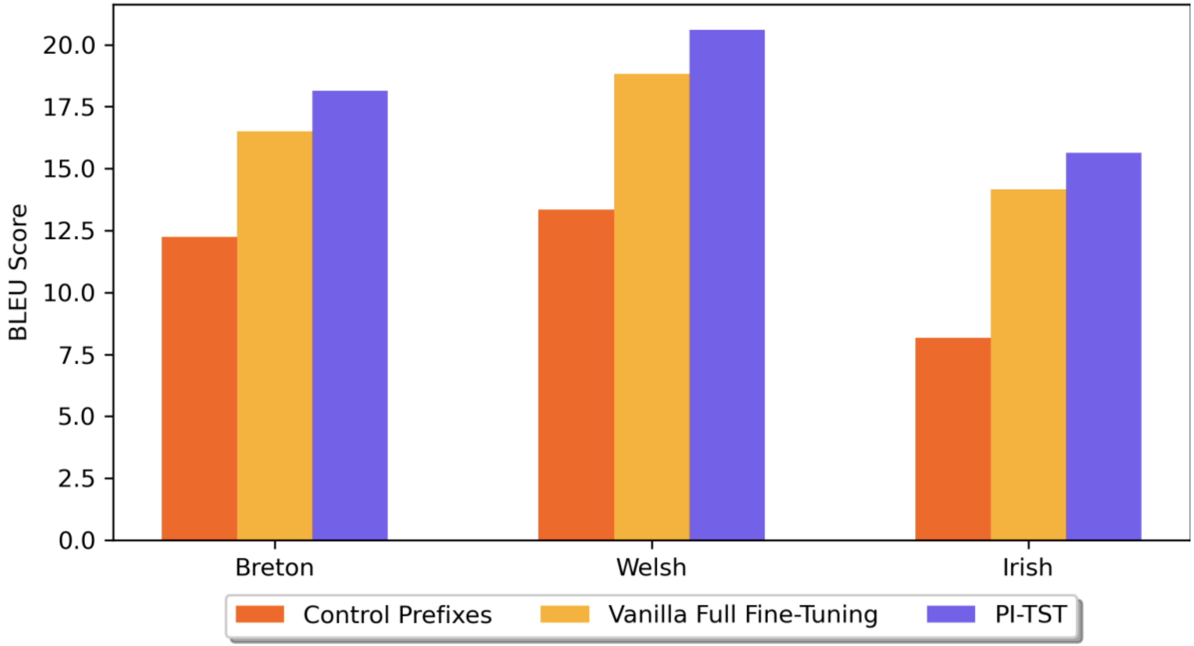
Step 2: Fine-Tuning on Dev RDF-to-Text data (RDF-to-Text Models)

Trains the RDF-to-Text Task sub-prompt for each target language

Inference

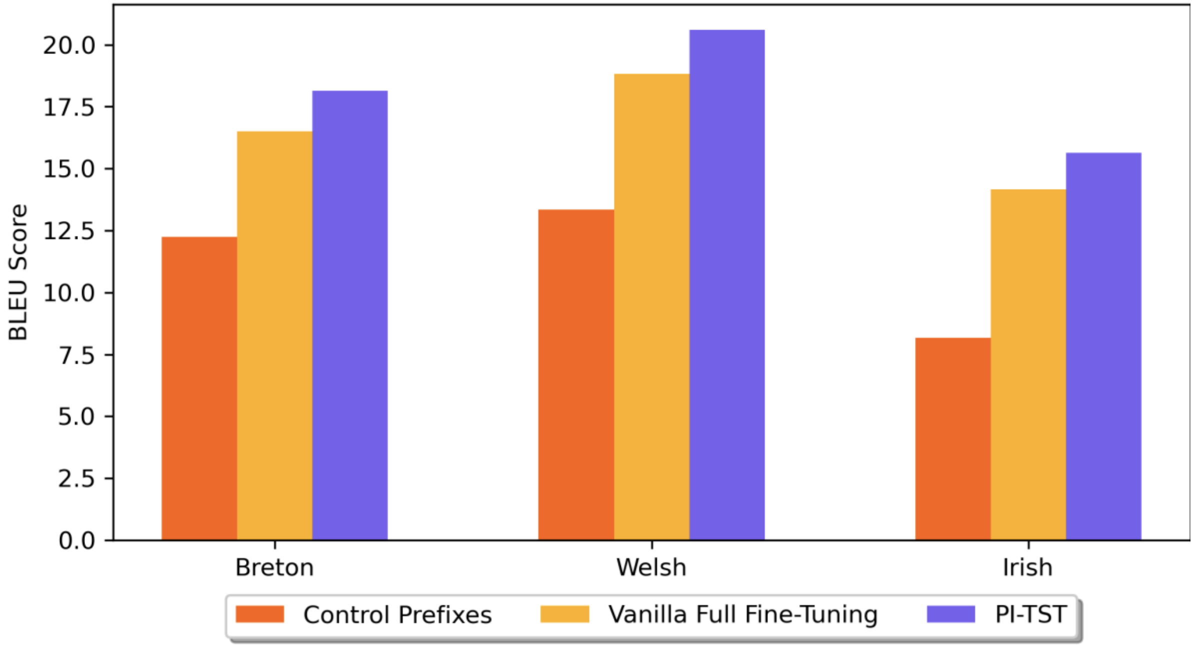
The Language sub-prompt is set to the target language.

Results



BLEU Score
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Breton
Welsh
Irish
Control Prefixes
Vanilla Full Fine-Tuning
PI-TST

Results



Phylogenetic prefix-tuning outperforms full fine-tuning and a SoTA approach for KG-to-Text generation

Key Takeaways

- Pretraining (2017 vs 2020) improves performance

Key Takeaways

- Pretraining (2017 vs 2020) improves performance
- Performance degrades on out of domain data (unseen)

Key Takeaways

- Pretraining (2017 vs 2020) improves performance
- Performance degrades on out of domain data (unseen)
- Performance is very poor for Low Resource Languages (2023)

Key Takeaways

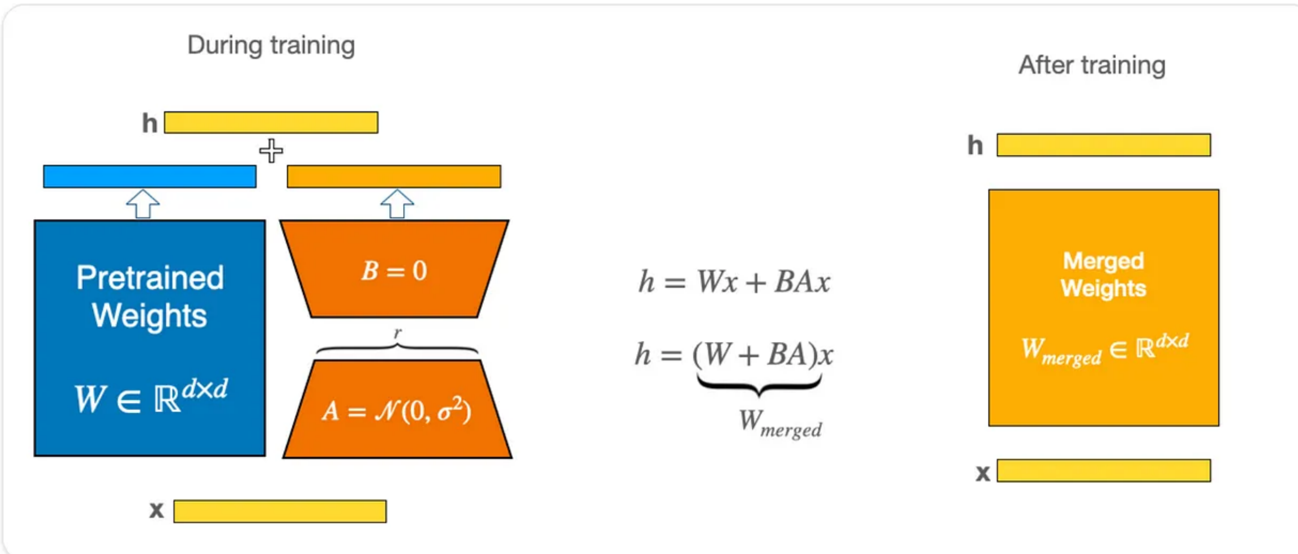
- Pretraining (2017 vs 2020) improves performance
- Performance degrades on out of domain data (unseen)
- Performance is very poor for Low Resource Languages (2023)
- PEFT techniques help improve performance for these languages

  BLEU for Breton: 10 (NLG+MT) $\rightarrow$ 18.15 (PEFT E2E Model)

AMR Graph --> High- and Low-Resource Languages

Soto-Martinez et al., INLG 2024

LoRA (Low Rank Matrices) Adaptation



A new model is created. No overhead during inference

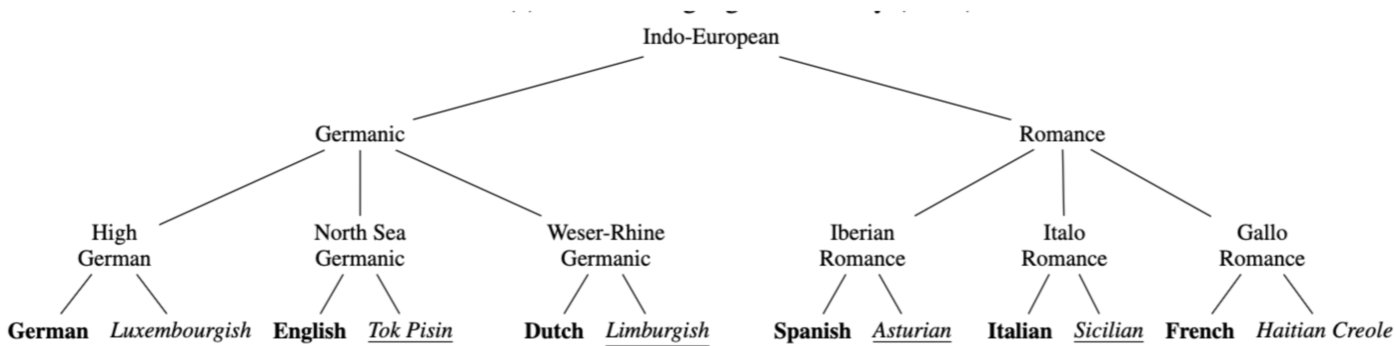
Hierarchical Fine-Tuning

Hierarchical Fine-Tuning

- Iterative fine-tuning of a multilingual model (12 languages) into 12 monolingual models

Phylogenetic Knowledge

- At each iteration, the training languages are chosen using phylogenetic knowledge



(b) Phylogenetic Tree Hierarchy (PTL)

Training and Test Data

Training Data

Dataset	Quality		Instances per Language		
	AMR	Text	Train	Test	Valid
FLORES-200	Silver	Gold	997	506	506
AMR 3.0	Gold	Silver	30 000	1 000	1 000
AMR3.0	Gold	Gold	N/A	1 371	N/A

AMR3.0 55.6K (gold AMR, English Text)

- English text translated using NLLB3.3B
- Filtering using GlotLID

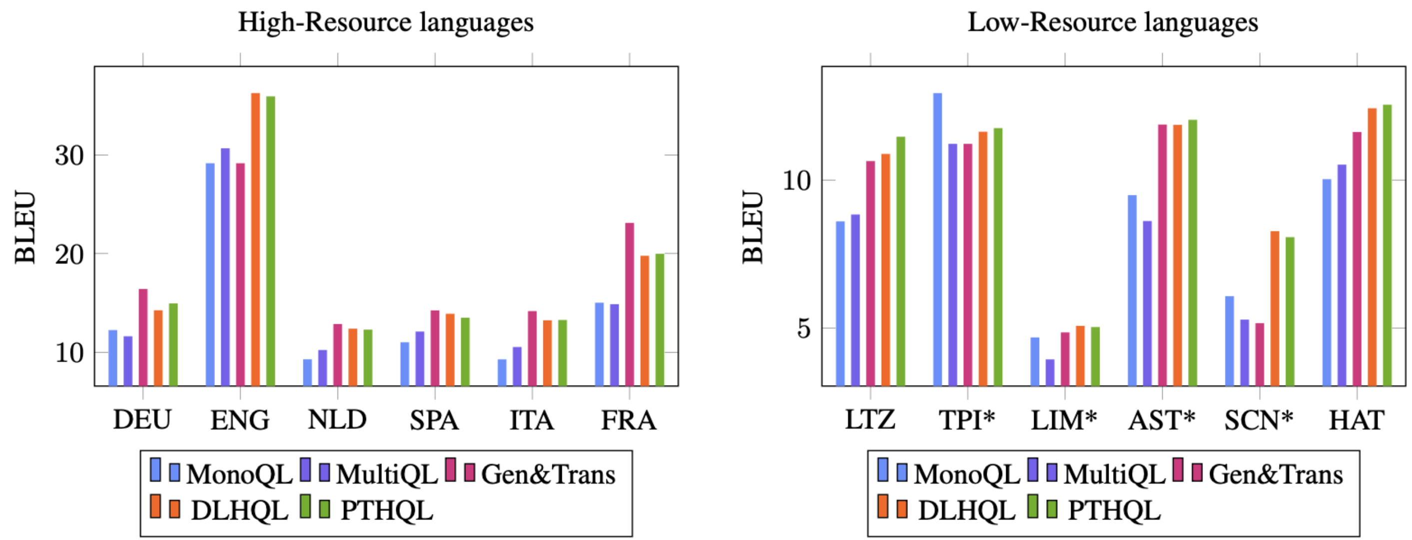
FLORES 200 dev data

- (silver AMR, Text)
- AMR obtained from English text using AMR3-structbart-L

Test Data

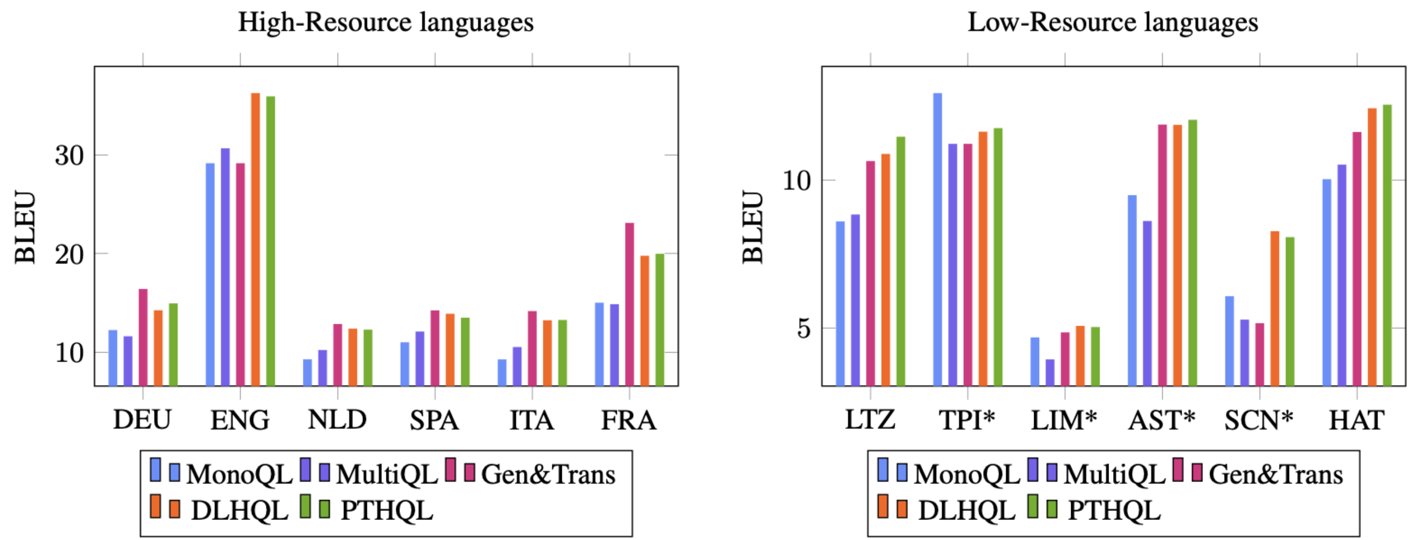
- (gold AMR, human written text) for English, German, Spanish and Italian
- (silver AMR, Text) from FLORES test data

Comparison with Baselines



HQL outperforms or is on par with multi- and monolingual approaches fine tuned using standard LoRA adaptation.

Comparison with Baselines



HQL outperforms the Gen&Trans approach for LR Languages

Tok Pisisn and Asturian (languages unseen by the base model) show a transfer effect as they perform on par with LR languages present in the base model's training data.

Comparison with Baselines

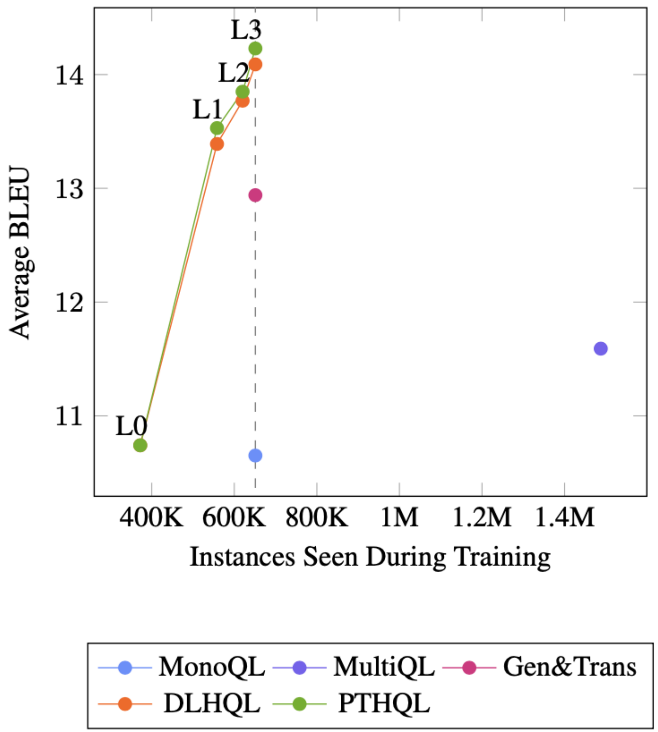


Figure 4: Average BLEU score of all languages against total number of seen training instances. HQL models include results on the intermediary levels of the hierarchy.

Comparison with Baselines

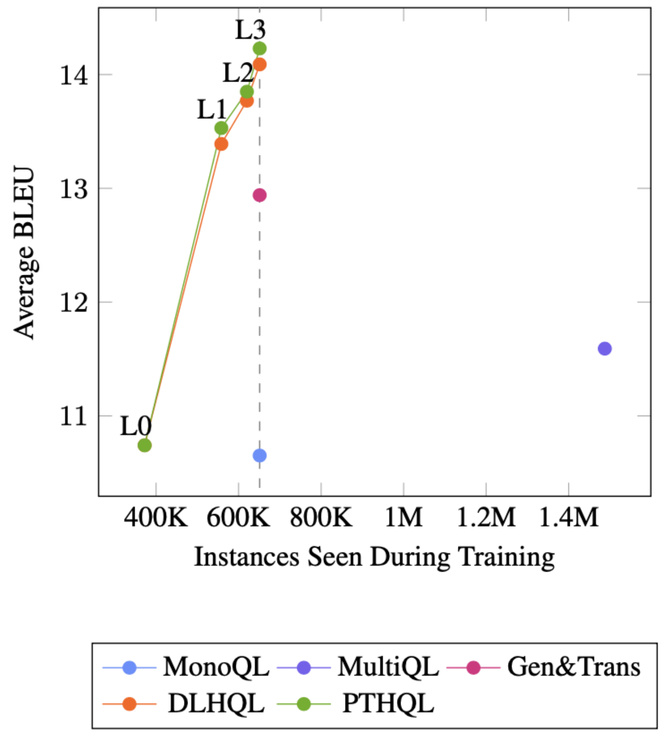


Figure 4: Average BLEU score of all languages against total number of seen training instances. HQL models include results on the intermediary levels of the hierarchy.

HQL optimises faster than the 3 baselines.

Comparison with Baselines

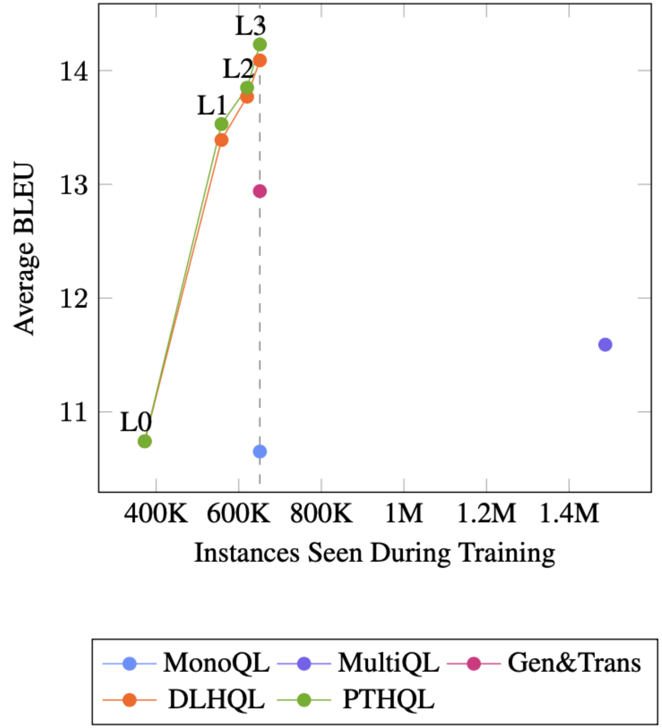


Figure 4: Average BLEU score of all languages against total number of seen training instances. HQL models include results on the intermediary levels of the hierarchy.

On average, HQL outperforms all 3 baselines.

Comparison with Previous Work

Model	DEU	ENG	SPA	ITA
F&G	15.3	24.9	21.7	19.8
Ribeiro	20.6	—	30.7	26.4
Xu	**25.7**	—	31.4	28.4
Martinez	23.2	44.8	34.6	29.0
MonoQL	18.2	**49.2**	38.6	22.7
MultiQL	19.8	42.9	34.1	27.2
Gen&Trans*	**28.0**	**49.2**	**39.6**	**33.8**
DLHQL	21.2	44.2	37.4	29.2
PTHQL	22.8	43.4	37.2	29.7

Table 4: BLEU score on AMR3.0 test data.
English Gen&Trans is simply the result of MonoQL.

HQL performs better or close to previous work on HRL while using fewer data.

Prompting LLMs

Knowledge Graphs $\Rightarrow$ English, Russian

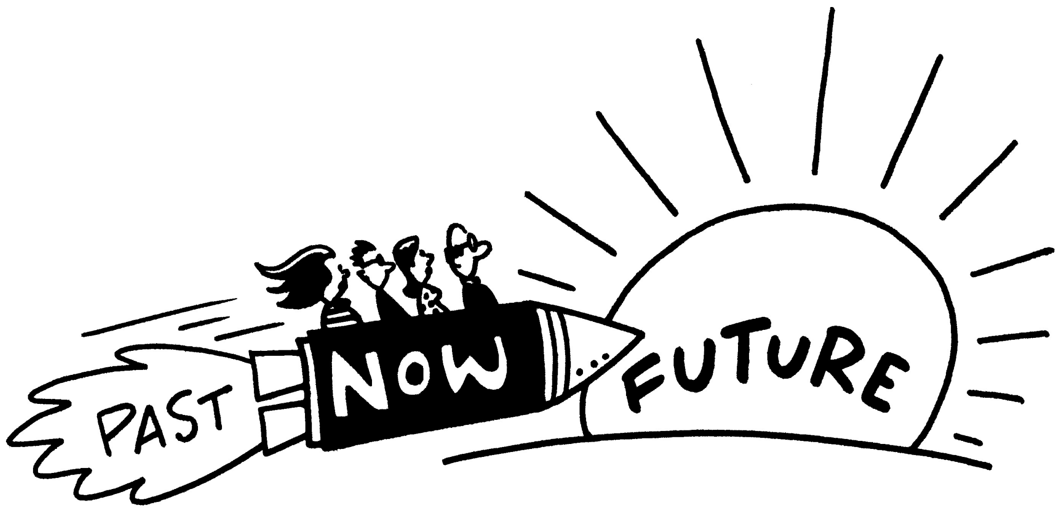
Preliminary Results

Results vary depending on

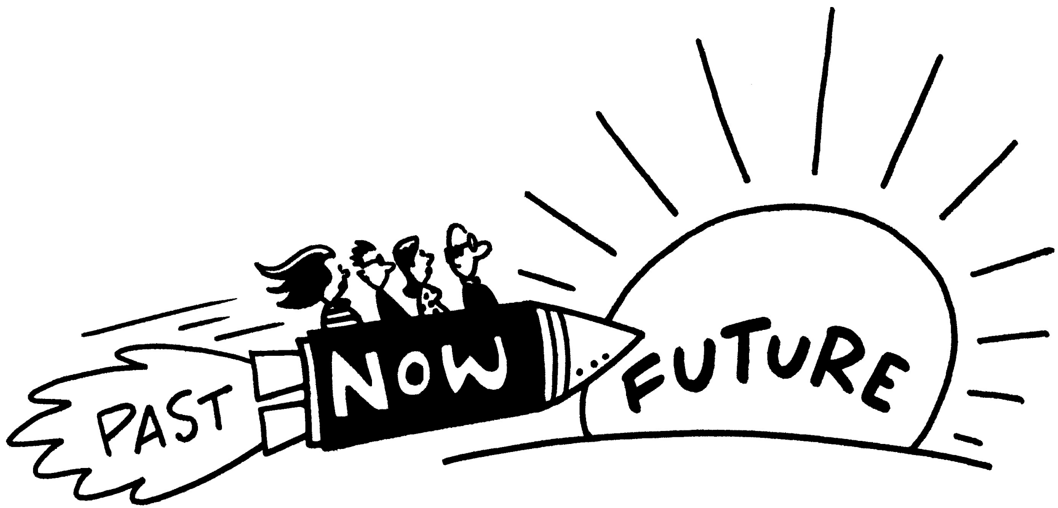
- the target language
  - BLEU English: 39, Russian: 18
- how few shots are selected
  - 2 BLEU points difference between best and worse strategy
- the prompting strategy
  - inputting the whole graph is better than CoT or an incremental strategy
- the input content
  - adding labels help

Preliminary results

Prompt Engineering outperforms the SoTA for both English and Russian

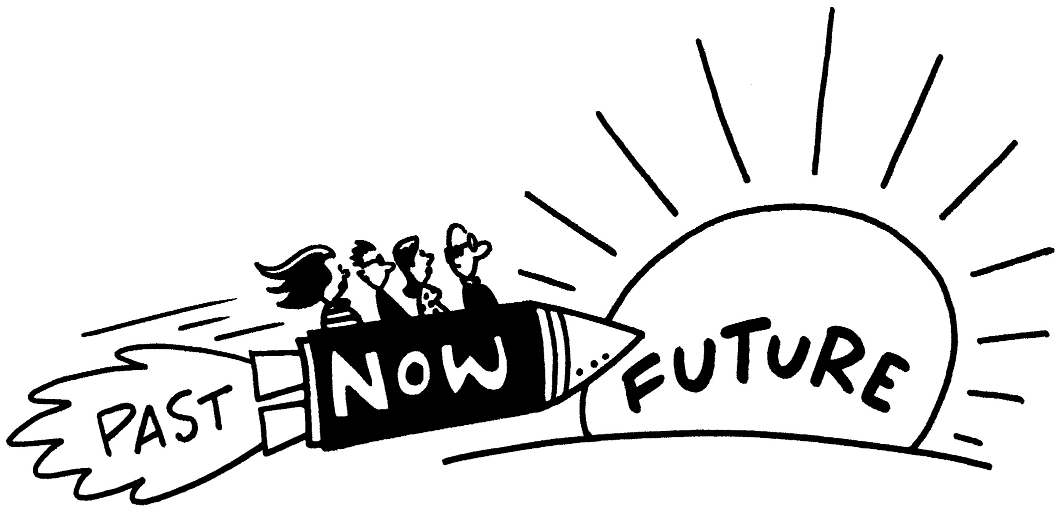


Cross-Modal, Multilingual Graph/Text similarity metrics are needed to



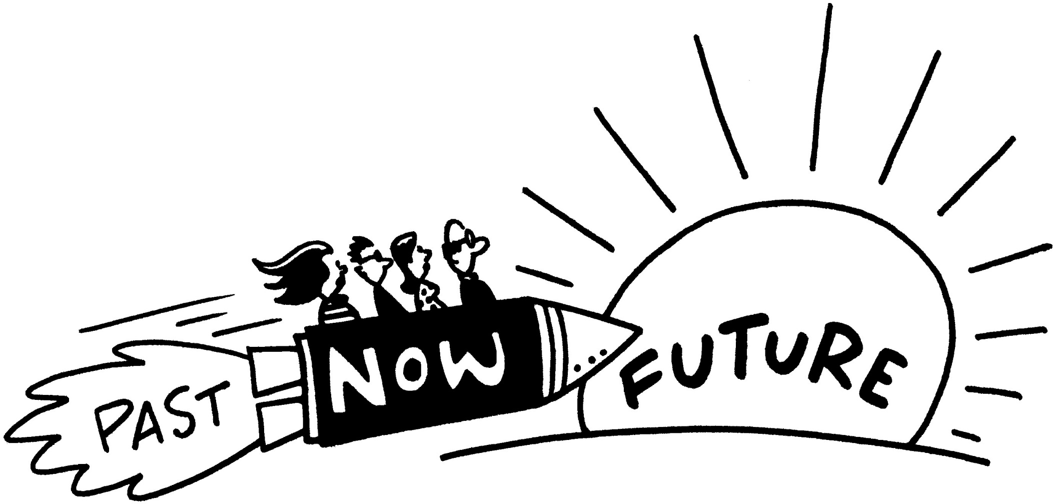
Cross-Modal, Multilingual Graph/Text similarity metrics are needed to

- filter noisy training data



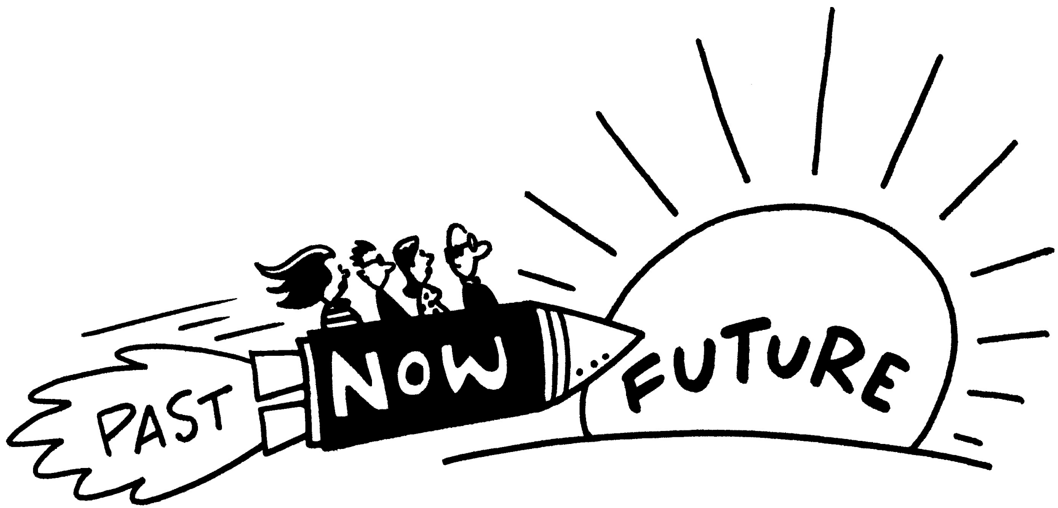
Cross-Modal, Multilingual Graph/Text similarity metrics are needed to

- filter noisy training data
- guide generation



Cross-Modal, Multilingual Graph/Text similarity metrics are needed to

- filter noisy training data
- guide generation
- support a reference less evaluation of Graph-to-Text generation



Cross-Modal, Multilingual Graph/Text similarity metrics are needed to

- filter noisy training data
- guide generation
- support a reference less evaluation of Graph-to-Text generation
- generalise Graph-to-Text Models to other languages and other domains

Questions ?