Génération de textes

A partir de textes et de graphes

Chaire IA xNLG: Generating from Multiple Sources into Multiple Languages

CNRS / LORIA

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR

Outline

Text $\rightarrow$ Text

- Document level Simplification

Outline

Text $\rightarrow$ Text

- Document level Simplification

Graph $\rightarrow$ Text

Outline

Text $\rightarrow$ Text

- Document level Simplification

Graph $\rightarrow$ Text

- Abstract Meaning Representation (AMR) $\rightarrow$ 21 EU languages
- RDF graph $\rightarrow$ English
- RDF graph $\rightarrow$ Low Resource Languages (Breton, Irish, Welsh)

Document Simplification

Cripwell et al. EMNLP Findings 2021, NAACL Findings 2022, EACL 2023, ACL Findings 2023

Example

Complex Input Document

Owls are birds from the order of Strigiformes, comprising over 200 species of mostly solitary and nocturnal birds of prey typified by an upright stance, binocular vision, binaural hearing, and sharp talons. Owls hunt mostly small mammals, insects, and other birds, although a few species specialize in hunting fish.

Simplified Output Document

Owls are birds. There are over 200 species and are all animals of prey. Most of them are solitary and nocturnal. Owls' prey may be birds, large insects (such as crickets), small reptiles (such as lizards) or small mammals (such as mice, rats, and rabbits).

Avg nb of sentences in Input Document: 39

Why Simplify ?

To aid reader comprehension (Mason, 1978; Williams et al., 2003; Kajiwara et al., 2013)

- Adult vs children
- Native vs non Native
- Reading disability
- Expert vs non-Expert

Simplification Operations

Owls are birds from the order of Strigiformes, comprising over 200 species of mostly solitary and nocturnal birds of prey typified by an upright stance, binocular vision, binaural hearing, and sharp talons. Owls hunt mostly small mammals, insects, and other birds, although a few species specialize in hunting fish.

Owls are birds. There are over 200 species ...

Sentence Splitting

Simplification Operations

Owls are birds from the order of Strigiformes, comprising over 200 species of ***mostly solitary and nocturnal birds of prey*** typified by an upright stance, binocular vision, binaural hearing, and sharp talons. Owls hunt mostly small mammals, insects, and other birds, although a few species specialize in hunting fish.

Owls are birds. There are over 200 species ***and are all animals of prey. Most of them are solitary and nocturnal*** .

Rephrasing

Simplification Operations

Owls are birds from the order of Strigiformes, comprising over 200 species of mostly solitary and nocturnal birds of prey ***typified by an upright stance, binocular vision, binaural hearing, and sharp talons*** . Owls hunt mostly small mammals, insects, and other birds, although a few species specialize in hunting fish.

Owls are birds. There are over 200 species and are all animals of prey. Most of them are solitary and nocturnal. Owls’ prey may be birds, large insects (such as crickets), small reptiles (such as lizards) or small mammals (such as mice, rats, and rabbits).

Deletion

Previous Work

Mostly on ***Sentence Simplification***

Previous Work

Mostly on ***Sentence Simplification***

Some work on Sentence-level techniques iteratively applied over a document (Woodsend and Lapata, 2011a; Alva-Manchego et al., 2019b)

Previous Work

Mostly on ***Sentence Simplification***

Some work on Sentence-level techniques iteratively applied over a document (Woodsend and Lapata, 2011a; Alva-Manchego et al., 2019b)

> *Insufficient for maintaining the discourse coherence of the document (Siddharthan, 2003; Alva-Manchego et al., 2019b)*.

Previous Work

Mostly on ***Sentence Simplification***

Some work on Sentence-level techniques iteratively applied over a document (Woodsend and Lapata, 2011a; Alva-Manchego et al., 2019b)

> *Insufficient for maintaining the discourse coherence of the document (Siddharthan, 2003; Alva-Manchego et al., 2019b)*.

A sentence-level model that uses context information to influence document simplification (Sun et al. 2020)

Previous Work

Mostly on ***Sentence Simplification***

Some work on Sentence-level techniques iteratively applied over a document (Woodsend and Lapata, 2011a; Alva-Manchego et al., 2019b)

> *Insufficient for maintaining the discourse coherence of the document (Siddharthan, 2003; Alva-Manchego et al., 2019b).*

A sentence-level model that uses context information to influence document simplification (Sun et al. 2020)

> *Unable to outperform the baseline* (Sun et al. 2021)

Our Proposal

Plan-Guided, Context-Aware Document Simplification

PLAN + SIMPLIFY

Our Proposal

Plan-Guided, Context-Aware Document Simplification

PLAN + SIMPLIFY

PLAN: predict a simplification operation for each sentence in the input document

Our Proposal

Plan-Guided, Context-Aware Document Simplification

PLAN + SIMPLIFY

PLAN: predict a simplification operation for each sentence in the input document

$$c_1, \ldots, c_n \Rightarrow \hat{o}_1, \ldots, \hat{o}_n$$

Our Proposal

Plan-Guided, Context-Aware Document Simplification

PLAN + SIMPLIFY

PLAN: predict a simplification operation for each sentence in the input document

$c_1, \ldots, c_n \Rightarrow \hat{o}_1, \ldots, \hat{o}_n$

$c_1, \ldots, c_n \Rightarrow split, delete, split, rephrase, rephrase, copy$

Our Proposal

Plan-Guided, Context-Aware Document Simplification

PLAN + SIMPLIFY

PLAN: predict a simplification operation for each sentence in the input document

$c_1, \ldots, c_n \Rightarrow \hat{o_1}, \ldots, \hat{o}_n$

$c_1, \ldots, c_n \Rightarrow split, delete, split, rephrase, rephrase, copy$

SIMPLIFY: Controlled Sentence Simplification

$c_i, \hat{o}_i \Rightarrow s_i$

Our Proposal

Plan-Guided, Context-Aware Document Simplification

PLAN + SIMPLIFY

PLAN: predict a simplification operation for each sentence in the input document

$c1, \ldots, c_n \Rightarrow \hat{o1}, \ldots, \hat{o}_n$

$c1, \ldots, c_n \Rightarrow split, delete, split, rephrase, rephrase, copy$

SIMPLIFY: Context-Aware, Controlled Sentence Simplification

$Context, c_i, \hat{o}_i \Rightarrow s_i$

Planning Simplification Operations

$c1, \ldots, c_n \Rightarrow \hat{o}, \ldots, \hat{o}_n$

Challenges

Simplification Operations have different requirements

Splitting

- mainly depends on the ***input sentence's internal structure***

  *The man **who** sleeps snores → The man sleeps. He snores.*

  *John went shopping **after** he left work → John left work. Afterwards he went shopping.*

Challenges

Simplification Operations have different requirements

Splitting

- mainly depends on the ***input sentence's internal structure***

  *The man **who** sleeps snores → The man sleeps. He snores.*

  *John went shopping **after** he left work → John left work. Afterwards he went shopping.*

Deletion, copy and rephrase

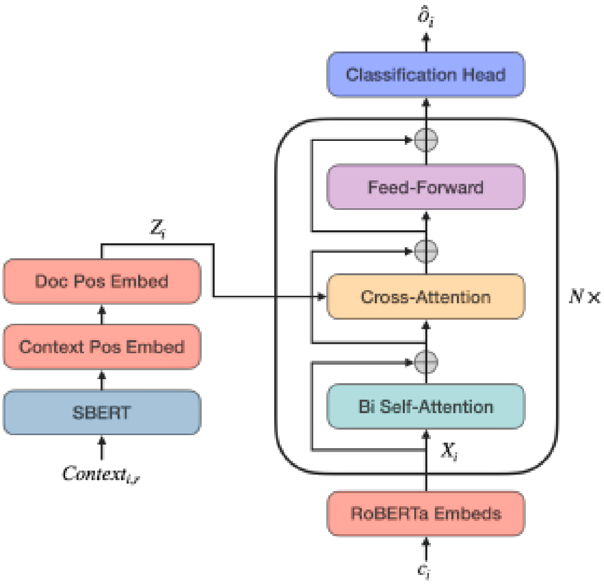
- are mostly ***context dependent*** .

  A sentence can only be omitted if it is either **redundant** with, or of **minor semantic import** relative to, other sentences in the document

Planning Model

RoBERTa classifier with cross-attention over the context

- layers initialised with weights from a context-independent classifier



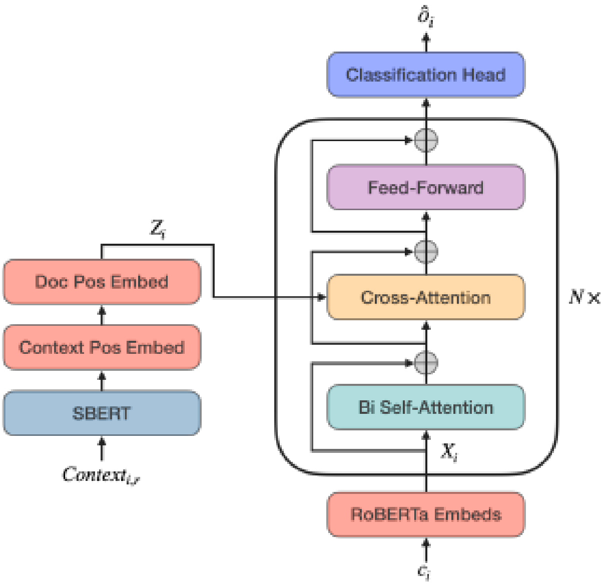
Planning Model

RoBERTa classifier with cross-attention over the context

- layers initialised with weights from a context-independent classifier

Internal structure

- ***Token level*** encoder for c_i



Planning Model

RoBERTa classifier with cross-attention over the context

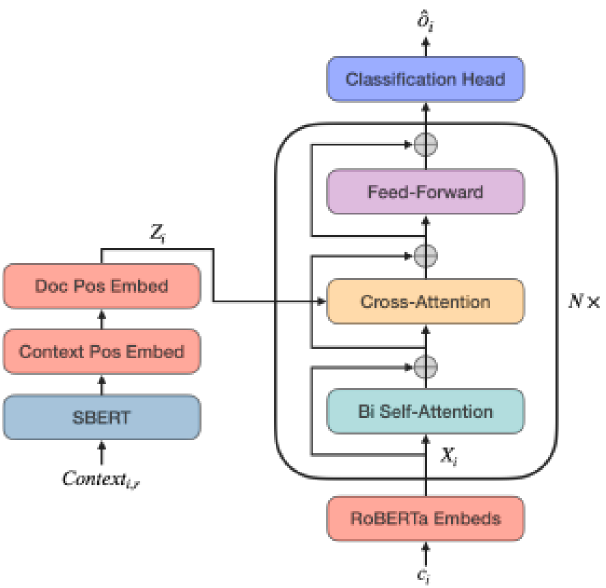
- layers initialised with weights from a context-independent classifier

Internal structure

- ***Token level*** encoder for c_i

Context

- fixed window of Sentence level embedding (SBERT) for ***surrounding sentences***



Planning Model

RoBERTa classifier with cross-attention over the context

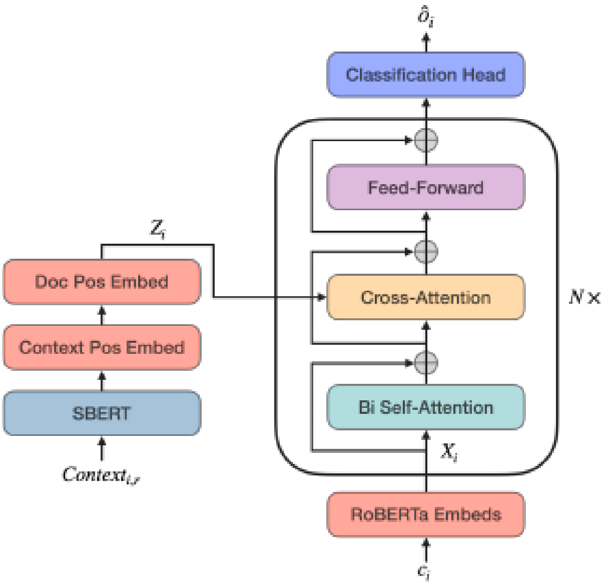
- layers initialised with weights from a context-independent classifier

Internal structure

- ***Token level*** encoder for c_i

Context

- fixed window of Sentence level embedding (SBERT) for ***surrounding sentences***
- The left context is ***dynamically*** updated with previously simplified sentences



Planning Model

RoBERTa classifier with cross-attention over the context

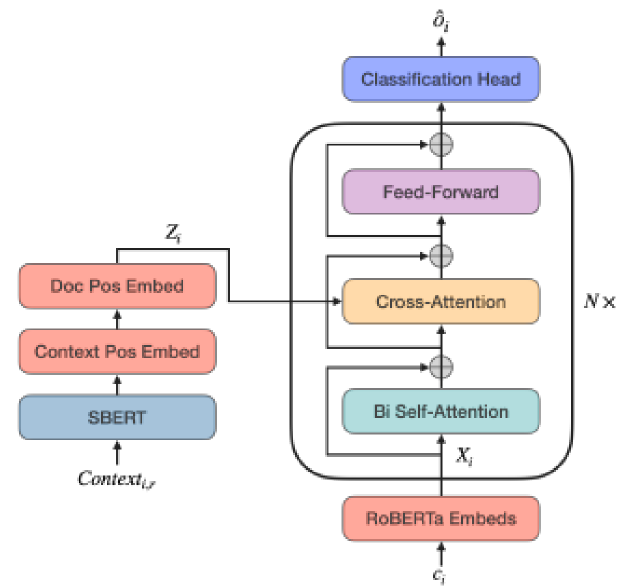
- layers initialised with weights from a context-independent classifier

Internal structure

- ***Token level*** encoder for c_i

Context

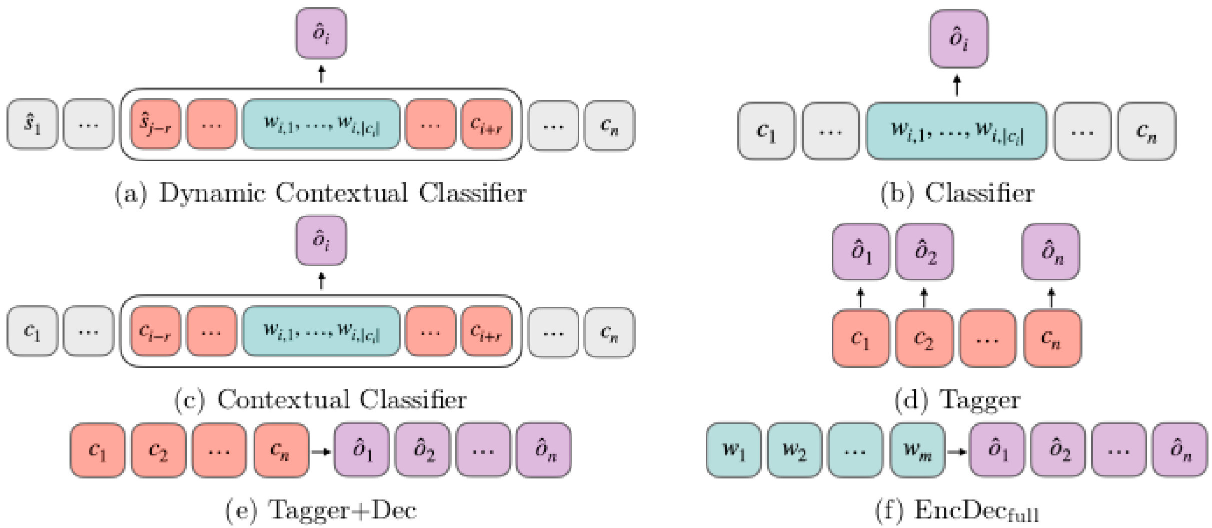
- fixed window of Sentence level embedding (SBERT) for ***surrounding sentences***
- The left context is ***dynamically*** updated with previously simplified sentences



Context positional embedding: relative distance of a given sentence from the input sentence c_i

Document positional embedding: the document quintile (1-5) that a given sentence falls into

Alternative Models

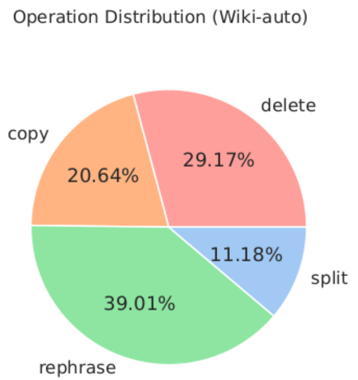


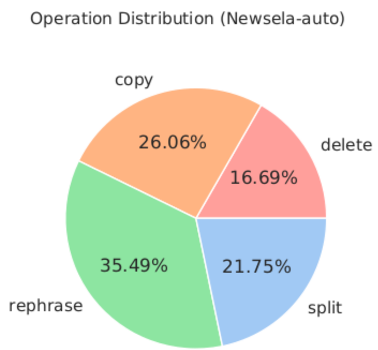
Dynamic Contextual Classifier: our model
Contextual Classifier: Static left context
Classifier: no context
Tagger: Sequence tagging on SBERT representations (no internal structure)

Tagger-Decoder: Each prediction is conditioned on the input document and on the previously predicted operation tags. SBERT encodings.
EncDec$_{full}$: Same as Tagger-Decoder but with token encodings

Data

(C, S) pairs with C a complex document and S its simplification (sentences are aligned)





	Wiki-auto	Newsela-auto
# Doc Pairs	85,123	18,319
# Sent Pairs	461,852	707,776
Avg. $\lvert C\rvert$	155.51	868.98
Avg. $\lvert S\rvert$	97.72	674.94
Avg. $\lvert c_i\rvert$	28.64	22.49
Avg. $\lvert s_i\rvert$	21.57	15.84
Avg. n	5.43	38.64
Avg. k	4.53	42.60

- n: the number of sentences in C
- k: the number of sentences in S

Labeling the data

$(C, S) \rightarrow (C, S, o)$

Delete

- c_i is not aligned to any s_j .
  The complex sentence c_i is not aligned to any sentence s_j in the simplified version.

Labeling the data

$(C, S) \rightarrow (C, S, o)$

Delete

- c_i is not aligned to any s_j .

Copy

- c_i is aligned to a single s_j with a Levenshtein similarity above 0.92.
  The complex sentence c_i is aligned to a similar sentence s_j in the simplified version

Labeling the data

$(C, S) \rightarrow (C, S, o)$

Delete

- c_i is not aligned to any s_j .

Copy

- c_i is aligned to a single s_j with a Levenshtein similarity above 0.92.

Rephrase

- c_i is aligned to a single s_j with a Levenshtein similarity below 0.92.
  The complex sentence c_i is aligned to a sentence s_j in the simplified version but differs from it.

Labeling the data

$(C, S) \rightarrow (C, S, o)$

Delete

- c_i is not aligned to any s_j .

Copy

- c_i is aligned to a single s_j with a Levenshtein similarity above 0.92.

Rephrase

- c_i is aligned to a single s_j with a Levenshtein similarity below 0.92.

Split

- c_i is aligned to multiple s_j
  The complex sentence c_i is aligned to several sentences in the simplified version.

Planning Accuracy Results

	Wiki-auto						Newsela-auto					
Model	C	R	S	D	Micro	**Macro**	C	R	S	D	Micro	**Macro**
$\text{EncDec}_{\text{full}}$	26.9	42.2	36.0	51.8	43.2	40.8	26.1	10.8	11.7	9.0	12.2	11.5
EncDec	29.3	54.5	30.0	51.8	47.7	41.4	72.2	73.9	75.9	79.7	75.0	75.4
Tagger	38.6	54.2	31.7	**58.5**	50.6	45.8	71.4	72.7	74.1	78.4	73.7	74.1
Classifier	42.1	52.9	42.6	49.0	48.4	46.7	77.0	75.6	80.0	78.5	77.4	77.8
Dyn. Context	**44.8**	**57.9**	42.4	54.8	**52.8**	**50.0**	79.3	77.3	82.8	81.4	79.7	80.2
+ docpos	43.7	55.4	**43.6**	56.7	52.3	49.9	**80.0**	**78.1**	**83.6**	**82.0**	**80.3**	**80.8**

- Our model (Dyn. Context) shows best results

Planning Accuracy Results

	Wiki-auto						Newsela-auto					
Model	C	R	S	D	Micro	**Macro**	C	R	S	D	Micro	**Macro**
EncDec$_{full}$	26.9	42.2	36.0	51.8	43.2	40.8	26.1	10.8	11.7	9.0	12.2	11.5
EncDec	29.3	54.5	30.0	51.8	47.7	41.4	72.2	73.9	75.9	79.7	75.0	75.4
Tagger	38.6	54.2	31.7	**58.5**	50.6	45.8	71.4	72.7	74.1	78.4	73.7	74.1
Classifier	42.1	52.9	42.6	49.0	48.4	46.7	77.0	75.6	80.0	78.5	77.4	77.8
Dyn. Context	**44.8**	**57.9**	42.4	54.8	**52.8**	**50.0**	79.3	77.3	82.8	81.4	79.7	80.2
+ docpos	43.7	55.4	**43.6**	56.7	52.3	49.9	**80.0**	**78.1**	**83.6**	**82.0**	**80.3**	**80.8**

- Our model (Dyn. Context) shows best results
- Deletion needs context
  the context-free classifier (Classifier) under-performs for deletions

Planning Accuracy Results

	Wiki-auto						Newsela-auto					
Model	C	R	S	D	Micro	**Macro**	C	R	S	D	Micro	**Macro**
EncDec$_{full}$	26.9	42.2	36.0	51.8	43.2	40.8	26.1	10.8	11.7	9.0	12.2	11.5
EncDec	29.3	54.5	30.0	51.8	47.7	41.4	72.2	73.9	75.9	79.7	75.0	75.4
Tagger	38.6	54.2	31.7	**58.5**	50.6	45.8	71.4	72.7	74.1	78.4	73.7	74.1
Classifier	42.1	52.9	42.6	49.0	48.4	46.7	77.0	75.6	80.0	78.5	77.4	77.8
Dyn. Context	**44.8**	**57.9**	42.4	54.8	**52.8**	**50.0**	79.3	77.3	82.8	81.4	79.7	80.2
+ docpos	43.7	55.4	**43.6**	56.7	52.3	49.9	**80.0**	**78.1**	**83.6**	**82.0**	**80.3**	**80.8**

- Our model (Dyn. Context) shows best results
- Deletion needs context
  the context-free classifier (Classifier) under-performs for deletions
- A token-level encoding of the complex sentence is important
  The encoder-decoder and the tagger, which both use a sentence level encoding of the complex sentence, underperform

Ablations

Model	Copy	Rephrase	Split	Delete	Micro	**Macro**
(a) Ablation on Best Model						
Dyn, $r = 13$, +init, +docpos	80.0	78.1	83.6	82.0	80.3	80.8
-docpos	79.3	77.3	82.8	81.4	79.7	80.2
-init	74.9	72.1	77.8	75.2	74.6	75.0
-init, -docpos	75.6	72.0	77.7	77.1	75.1	75.6
(b) Dynamic vs. Static Context						
Stat, $r = 9$	71.3	69.5	75.4	73.3	72.0	72.4
Stat, $r = 13$	72.2	65.3	69.9	68.3	68.5	68.9
Dyn, $r = 9$	73.1	70.1	75.5	75.9	73.1	73.6
Dyn, $r = 13$	75.6	72.0	77.7	77.1	75.1	75.6
(c) With vs without Initialisation						
Dyn, $r = 9$	73.1	70.1	75.5	75.9	73.1	73.6
Dyn, $r = 9$ +init	79.3	78.0	82.7	79.8	79.7	80.0
Dyn, $r = 13$	75.6	72.0	77.7	77.1	75.1	75.6
Dyn, $r = 13$ +init	79.3	77.3	82.8	81.4	79.7	80.2
(d) Window Size						
Stat, $r = 9$	71.3	69.5	75.4	73.3	72.0	72.4
Stat, $r = 13$	72.2	65.3	69.9	68.3	68.5	68.9
Dyn, $r = 9$	73.1	70.1	75.5	75.9	73.1	73.6
Dyn, $r = 13$	75.6	72.0	77.7	77.1	75.1	75.6
Dyn, $r = 9$ +docpos	73.8	72.9	77.2	75.8	74.6	74.9
Dyn, $r = 13$ +docpos	74.9	72.1	77.8	75.2	74.6	75.0
Dyn, $r = 9$ +init +docpos	79.4	78.0	83.1	82.0	80.1	80.6
Dyn, $r = 13$ +init +docpos	80.0	78.1	83.6	82.0	80.3	80.8

Plan-Guided Simplification

$$c_1, \ldots, c_n \Rightarrow \hat{o}_1, \ldots, \hat{o}_n$$

$$c_i, \hat{o}_i \Rightarrow s_i$$

Predict simplification operations

$$c_1, \ldots, c_n \Rightarrow \hat{o}_1, \ldots, \hat{o}_n$$

Simplify each input sentences using controls

$$c_i, \hat{o}_i \Rightarrow s_i$$

Models

Fine-tuned on ***sentence pairs*** and
iteratively applied to each input sentence

- Plan-Guided (PG): ***pipeline***

$$D \rightarrow \hat{o}_1, \ldots, \hat{o}_n$$

$$c_i, \hat{o}_i \Rightarrow s_i$$

Models

Fine-tuned on ***sentence pairs*** and
iteratively applied to each input sentence

- Plan-Guided (PG): ***pipeline***

$$D \rightarrow \hat{o}_1, \ldots, \hat{o}_n$$

$$c_i, \hat{o}_i \Rightarrow s_i$$

- Sent-BART: ***end-to-end***

$$c_i \Rightarrow s_i$$

Models

Fine-tuned on ***sentence pairs*** and iteratively applied to each input sentence

- Plan-Guided (PG): ***pipeline***

$$D \rightarrow \hat{o}_1, \ldots, \hat{o}_n$$

$$c_i, \hat{o}_i \Rightarrow s_i$$

- Sent-BART: ***end-to-end***

$$c_i \Rightarrow s_i$$

Fine-tuned on ***full document pairs***

- Doc-BART

$$DOC \Rightarrow SIMPLIFIED$$

Evaluation Metrics

Summarization metrics

- BARTScore (Yuan et al., 2021)
- SMART (Amplayo et al., 2022)

SARI (Xu et al., 2016)

- Most popular simplification metric.
- Computes n-gram edits between input, output, and references.

FKGL (Kincaid et al., 1975)

- Readibility metrics
- Uses surface-level statistics like syllable counts and sentence length.

Results

System	BARTScore ↑				SMART ↑			FKGL ↓	SARI ↑	Length	
	Faith. $(s \rightarrow h)$	P $(r \rightarrow h)$	R $(h \rightarrow r)$	F1	P	R	F1			Tokens	Sents
Input	-0.93	-2.47	-1.99	-2.23	63.2	62.7	62.8	8.44	20.52	866.9	38.6
Reference	-1.99	-0.93	-0.93	-0.93	100	100	100	4.93	99.99	671.5	42.6
Doc-BART	-2.48	-2.68	-2.76	-2.72	61.9	43.9	50.6	10.01	47.07	600.8	20.7
Sent-BART	**-1.86**	-1.63	-1.56	-1.60	78.9	80.1	79.3	5.03	73.02	666.4	42.6
PG_{Tag}	-1.95	-2.22	-2.18	-2.20	5.07	62.0	62.6	61.6	56.13	657.4	41.8
PG_{EncDec}	-1.94	-2.22	-2.18	-2.20	62.2	62.5	61.6	5.09	56.06	654.2	41.4
PG_{Clf}	-1.91	-1.68	**-1.53**	-1.60	77.8	**81.2**	79.3	**4.95**	73.83	688.8	44.5
PG_{Dyn}	-1.91	**-1.60**	-1.54	**-1.57**	**80.2**	81.0	**80.5**	4.98	**75.00**	667.2	42.6
PG_{Oracle}	-1.93	**-1.39**	**-1.40**	**-1.40**	**85.5**	**85.0**	**85.3**	**4.91**	**80.74**	655.6	42.1

- ***Pipeline*** (PG Dyn) achieves the highest results of all systems.

Results

System	BARTScore ↑				SMART ↑			FKGL ↓	SARI ↑	Length	
	Faith. $(s \to h)$	P $(r \to h)$	R $(h \to r)$	F1	P	R	F1			Tokens	Sents
Input	-0.93	-2.47	-1.99	-2.23	63.2	62.7	62.8	8.44	20.52	866.9	38.6
Reference	-1.99	-0.93	-0.93	-0.93	100	100	100	4.93	99.99	671.5	42.6
Doc-BART	-2.48	-2.68	-2.76	-2.72	61.9	43.9	50.6	10.01	47.07	600.8	20.7
Sent-BART	**-1.86**	-1.63	-1.56	-1.60	78.9	80.1	79.3	5.03	73.02	666.4	42.6
PG_{Tag}	-1.95	-2.22	-2.18	-2.20	5.07	62.0	62.6	61.6	56.13	657.4	41.8
PG_{EncDec}	-1.94	-2.22	-2.18	-2.20	62.2	62.5	61.6	5.09	56.06	654.2	41.4
PG_{Clf}	-1.91	-1.68	**-1.53**	-1.60	77.8	**81.2**	79.3	**4.95**	73.83	688.8	44.5
PG_{Dyn}	-1.91	**-1.60**	-1.54	**-1.57**	**80.2**	81.0	**80.5**	4.98	**75.00**	667.2	42.6
PG_{Oracle}	-1.93	**-1.39**	**-1.40**	**-1.40**	**85.5**	**85.0**	**85.3**	**4.91**	**80.74**	655.6	42.1

- ***Pipeline*** (PG Dyn) achieves the highest results of all systems.
- ***Improving planning*** (PG Oracle) would substantially increase performance (PG Oracle)

Results

System	BARTScore ↑				SMART ↑			FKGL ↓	SARI ↑	Length	
	Faith. ($s \rightarrow h$)	P ($r \rightarrow h$)	R ($h \rightarrow r$)	F1	P	R	F1			Tokens	Sents
Input	-0.93	-2.47	-1.99	-2.23	63.2	62.7	62.8	8.44	20.52	866.9	38.6
Reference	-1.99	-0.93	-0.93	-0.93	100	100	100	4.93	99.99	671.5	42.6
Doc-BART	-2.48	-2.68	-2.76	-2.72	61.9	43.9	50.6	10.01	47.07	600.8	20.7
Sent-BART	**-1.86**	-1.63	-1.56	-1.60	78.9	80.1	79.3	5.03	73.02	666.4	42.6
PG_{Tag}	-1.95	-2.22	-2.18	-2.20	5.07	62.0	62.6	61.6	56.13	657.4	41.8
PG_{EncDec}	-1.94	-2.22	-2.18	-2.20	62.2	62.5	61.6	5.09	56.06	654.2	41.4
PG_{Clf}	-1.91	-1.68	**-1.53**	-1.60	77.8	**81.2**	79.3	**4.95**	73.83	688.8	44.5
PG_{Dyn}	-1.91	**-1.60**	-1.54	**-1.57**	**80.2**	81.0	**80.5**	4.98	**75.00**	667.2	42.6
PG_{Oracle}	-1.93	**-1.39**	**-1.40**	**-1.40**	**85.5**	**85.0**	**85.3**	**4.91**	**80.74**	655.6	42.1

- ***Pipeline*** (PG Dyn) achieves the highest results of all systems.
- ***Improving planning*** (PG Oracle) would substantially increase performance (PG Oracle)
- E2E ***simplification of full document*** (Doc-BART) yields poor results

Context-Aware, Plan-Guided Document Simplification

$$c_1, \ldots, c_n \Rightarrow \hat{o}_1, \ldots, \hat{o}_n$$

$$Context, c_i, \hat{o}_i \Rightarrow s_i$$

Context-Aware, Plan-Guided Simplification

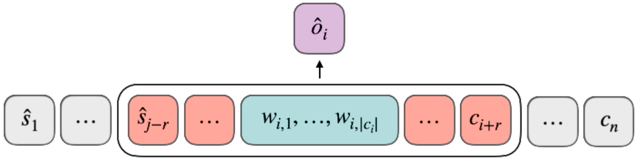
PG (plan-guided) pipeline

First PLAN,

Input D ⇒ Simplification Plan

$c1, \ldots, c_n \Rightarrow \hat{o}, \ldots, \hat{o}_n$

PLANNING is Context-Aware ...



then SIMPLIFY

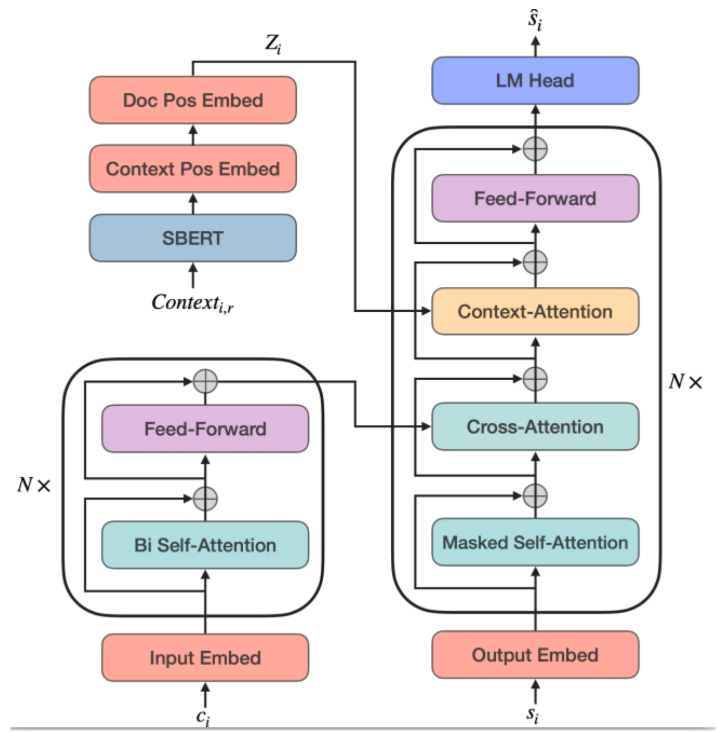
Input S + Simplification Operation ⇒ Simplified S

$c_i, \hat{o}_i \Rightarrow s_i$

... but SIMPLIFICATION is not

ConBART: Context-Aware Sentence Simplification

- Modification of the BART architecture
- Generation is ***conditioned on both an input sentence*** c_i ***and a representation of the document context*** Z_i of that sentence
- Same ***context modeling*** as for planner (SBERT encoding of the neighbouring sentences)



Models

Text-Only Models (BART, LED)

- input = sentence, paragraph or document
- Model: LongFormer (LED)

Models

Text-Only Models (BART, LED)

- input = sentence, paragraph or document
- Model: LongFormer (LED)

Contextual Model (ConBART)

- Input: sentence + context window of n sentences
- Model: Context-aware, controlled sentence simplification (ConBART)

Models

Text-Only Models (BART, LED)

- input = sentence, paragraph or document
- Model: LongFormer (LED)

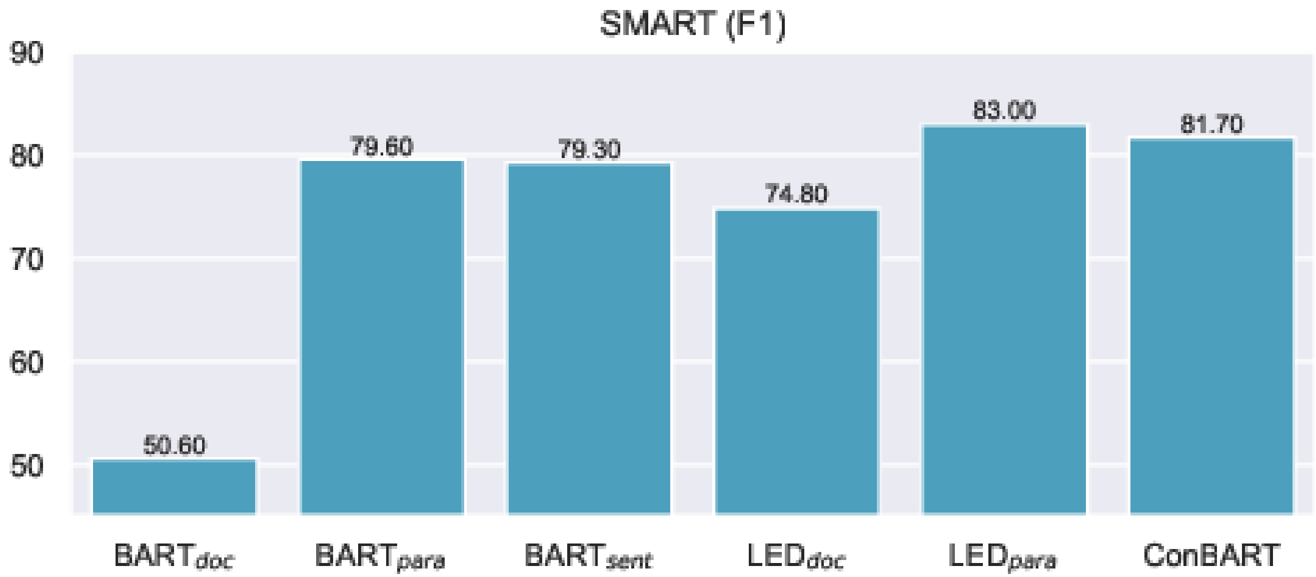
Contextual Model (ConBART)

- Input: sentence + context window of n sentences
- Model: Context-aware, controlled sentence simplification (ConBART)

Plan-Guided Pipelines ($\hat{O} \rightarrow M$)

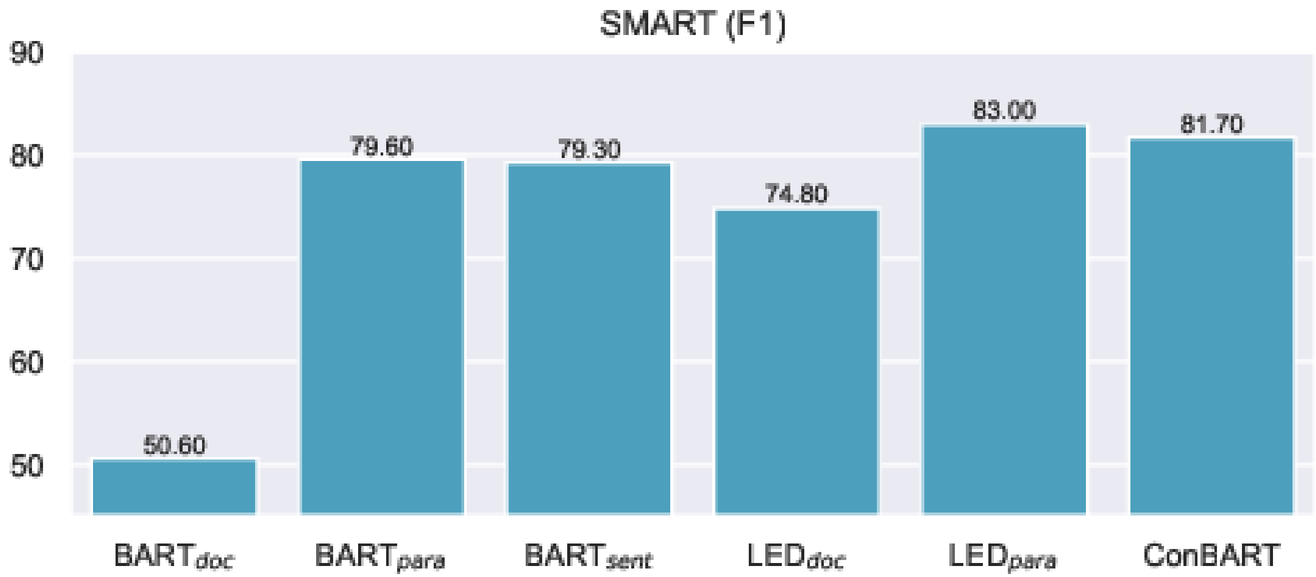
- $\hat{O}$, a predicted simplification plan
- M, a simplification model (BART, LED, ConBART)

Which context helps most ?



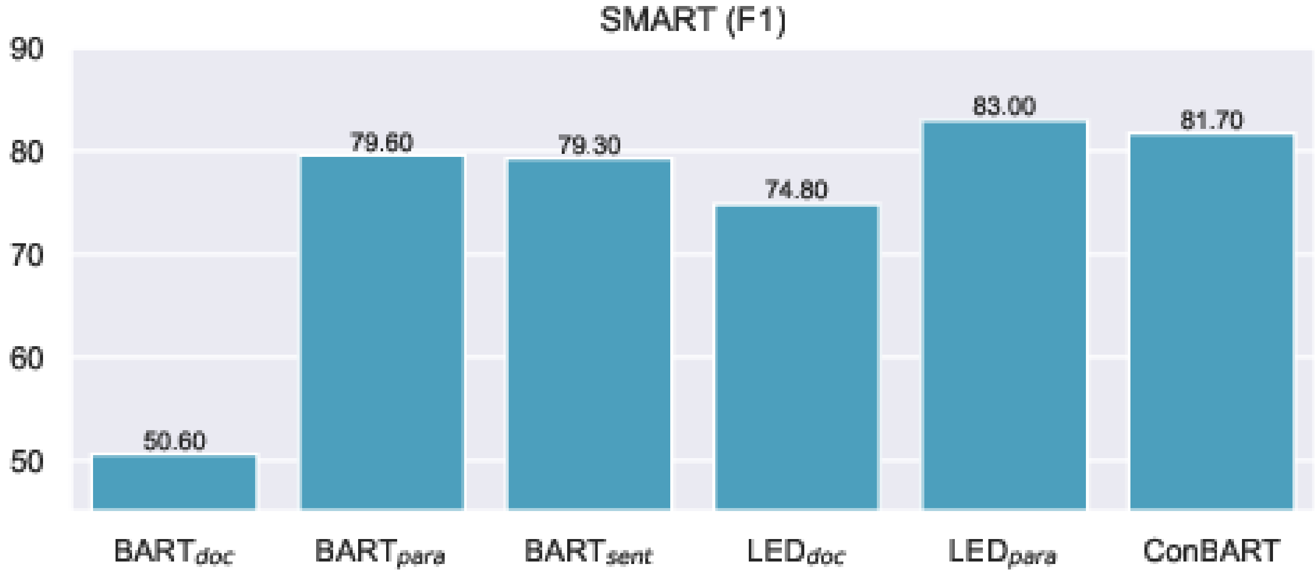
- The best two models use a medium size context (ConBART, LED$_{para}$)

Which context helps most ?



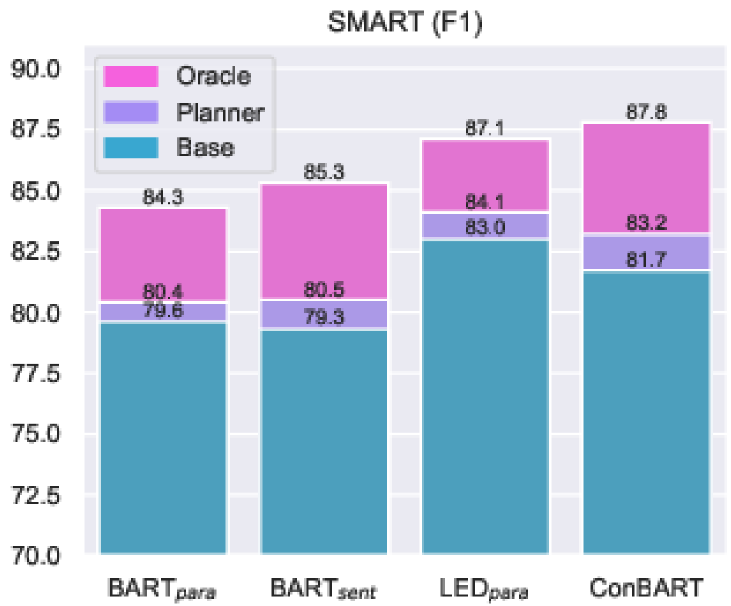
- The best two models use a medium size context (ConBART, LED$_{para}$)
- Full Document context does not work well (BART$_{doc}$, LED$_{doc}$)

Which context helps most ?



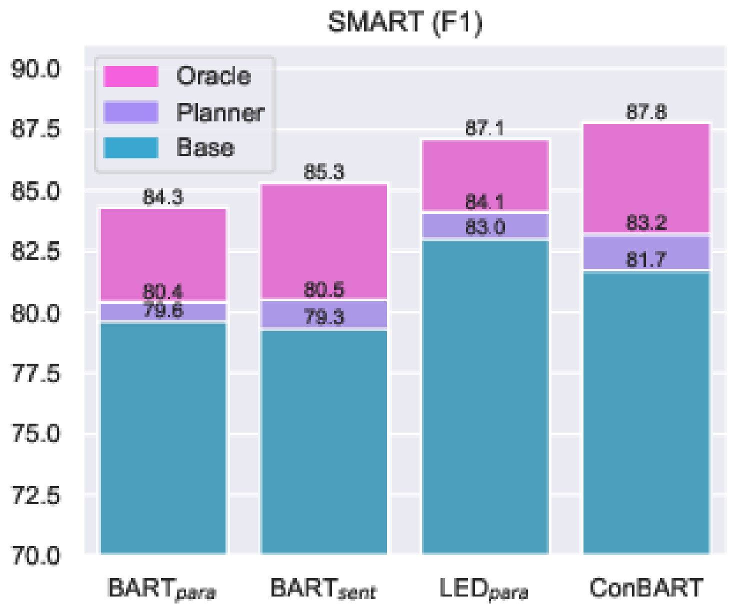
- The best two models use a medium size context (ConBART, LED$_{para}$)
- Full Document context does not work well (BART$_{doc}$, LED$_{doc}$)
- LongFormers improve results (BART$_{X}$ vs. LED$_{X}$)

Does planning help ?



- Planning systematically improves performance

Does planning help ?



- Planning systematically improves performance
- Planning needs improving

Models simplifying based on the oracle plan have much higher performance

Human Evaluation

On 198 paragraphs, Binary question, Score = Proportion of yes for each model
Comparison: $PG_{Dyn}, LED_{para}, \hat{O} \rightarrow LED_{para}, \hat{O} \rightarrow ConBART$

Fluency

All systems achieve high fluency (92.9% - 95.5%).

Semantic Adequacy

$\hat{O} \rightarrow LED_{para}$ is best (81.1%).

Simplificity

$\hat{O} \rightarrow LED_{para}$ and $\hat{O} \rightarrow ConBART$ achieve highest simplicity (89.4%).

Planning helps

Performance on OOD Data

System	Flu	Ad	Simp	**Mean**
LED_{para}	**0.932**	0.632	0.664	0.743
$\hat{O} \rightarrow \text{LED}_{\text{para}}$	0.890	0.684	0.760	0.778
$\hat{O} \rightarrow$ ConBART	0.890	**0.760**	**0.764**	**0.805**

Human Evaluation, training on Newsela, testing on Wikipedia

- Plan-guidance helps

Performance on OOD Data

System	Flu	Ad	Simp	Mean
$\mathrm{LED}_{\mathrm{para}}$	**0.932**	0.632	0.664	0.743
$\hat{O} \rightarrow \mathrm{LED}_{\mathrm{para}}$	0.890	0.684	0.760	0.778
$\hat{O} \rightarrow$ ConBART	0.890	**0.760**	**0.764**	**0.805**

Human Evaluation, training on Newsela, testing on Wikipedia

- Plan-guidance helps
- Paragraph-level models likely less adaptable to unseen domains because biased towards paragraph length of training data.

Example Output

Input

Glenn Edward Greenwald (born March 6, 1967) is an American journalist and author.

He is best known for a series of reports published from June 2013 by "The Guardian" newspaper detailing the United States and British global surveillance programs, and based on *classified documents disclosed* by Edward Snowden. Greenwald and the team he worked with won both a George Polk Award and a Pulitzer Prize for those reports.

He has written several best-selling books, including "No Place to Hide". Before the Snowden file *disclosures*, Greenwald was *considered one of the most influential* opinion columnists in the United States. After working as a constitutional attorney for ten years, he began *blogging* on national security issues before becoming a "Salon" *contributor* in 2007 and *then* for "The Guardian" in 2012. He now writes for (and has co-edited) "The Intercept", which he founded in 2013 with Laura Poitras and Jeremy Scahill.

Greenwald's work on the Snowden story was featured in the documentary "Citizenfour", which won *the 2014 Academy Award for Best Documentary Feature*. Greenwald *appeared on-stage with director Laura Poitras and Snowden's girlfriend, Lindsay Mills, when the Oscar was given*. In the 2016 Oliver Stone feature film "Snowden", Greenwald was played by actor Zachary Quinto.

Output

Glenn Greenwald is an American journalist and author.

He is best known for a series of reports published from June 2013 by the Guardian newspaper. <SPLIT> They are based on *documents leaked* by Edward Snowden.

He has written several best-selling books, including "No Place to Hide." Before the Snowden file *leaks*, Greenwald *was one of the most respected* opinion columnists in the United States. He began writing about national security issues before becoming a "Salon" *writer* in 2007 and *a writer* for "The Guardian" in 2012. He now writes for The Guardian.

Greenwald's work on the Snowden story was featured in the documentary "Citizenfour". <SPLIT> The movie won *an Academy Award*. Greenwald *worked with director Laura Poitras and Snowden's girlfriend, Lindsay Mills, to make the documentary*. The 2016 Oliver Stone feature, "Snowden," was played by Zachary Quinto.

Key Takeaways and Perspectives

- Controlled and context-aware sentence simplification can effectively be used to simplify documents

Key Takeaways and Perspectives

- Controlled and context-aware sentence simplification can effectively be used to simplify documents
- Document simplification is under-studied

Key Takeaways and Perspectives

- Controlled and context-aware sentence simplification can effectively be used to simplify documents
- Document simplification is under-studied
    - Few training and test data for languages other than English

Key Takeaways and Perspectives

- Controlled and context-aware sentence simplification can effectively be used to simplify documents
- Document simplification is under-studied
  - Few training and test data for languages other than English
  - Few work on User adaptation

    Simplification models for different types of users (native vs. non native, adult vs. children, expert vs layman, people with reading disability)

Graph $\Rightarrow$ Text

Graphs are frequent datastructures

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

Graph $\rightarrow$ Text

AMR $\Rightarrow$ Text

- Generating into 21 EU Languages

RDF $\Rightarrow$ Text

- WebNLG Challenge
- Generating into Low Resource Languages (Breton, Welsh, Irish)

AMR $\Rightarrow$ Text

Multilingual Generation

Fan and Gardent EMNLP 2020

Challenges of Multilingual Graph-to-Text

- Structured input has a different surface form

Challenges of Multilingual Graph-to-Text

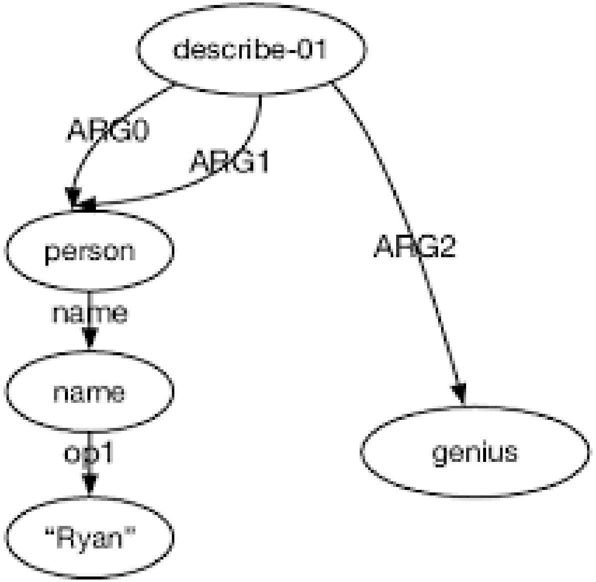
- Structured input has a different surface form
- Structured Input is often very underspecified

Challenges of Multilingual Graph-to-Text

- Structured input has a different surface form
- Structured Input is often very underspecified
- Decoding into languages with varied morphology and word order

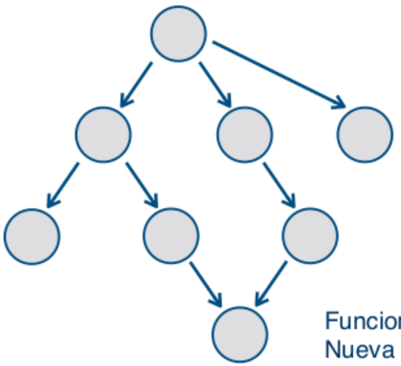
Abstract Meaning Representation (AMR)

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.

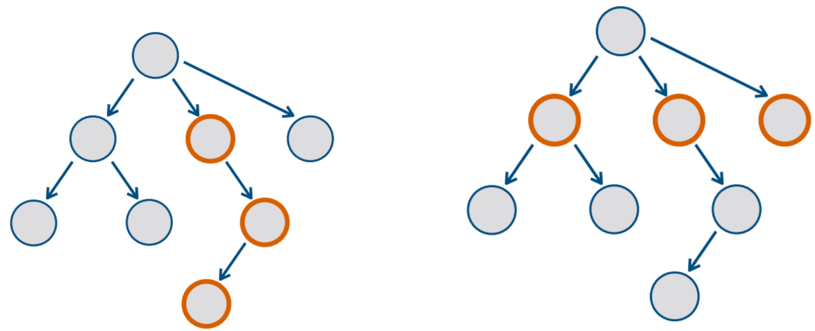
Des responsables américains ont tenu une réunion d'un groupe d'experts en janvier 2002 à Ne
York.

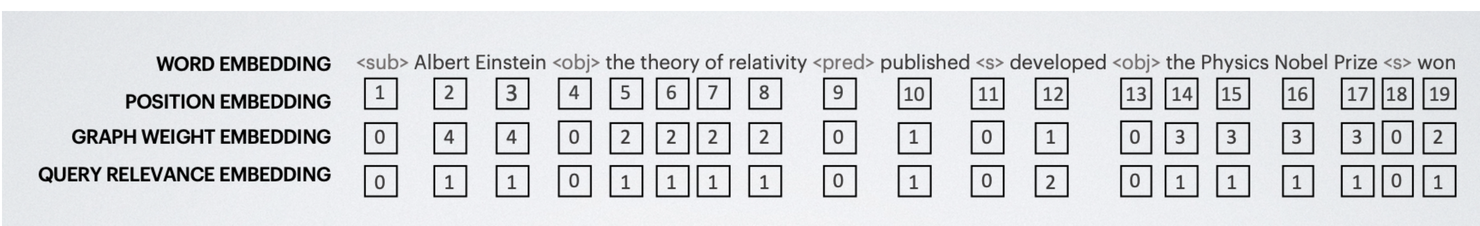
Funcionarios estadounidenses celebraron una reunión de un grupo de expertos en enero de 2002 en Nueva York.

Romance, Germanic, Slavic, Uralic

Encoding AMRs

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

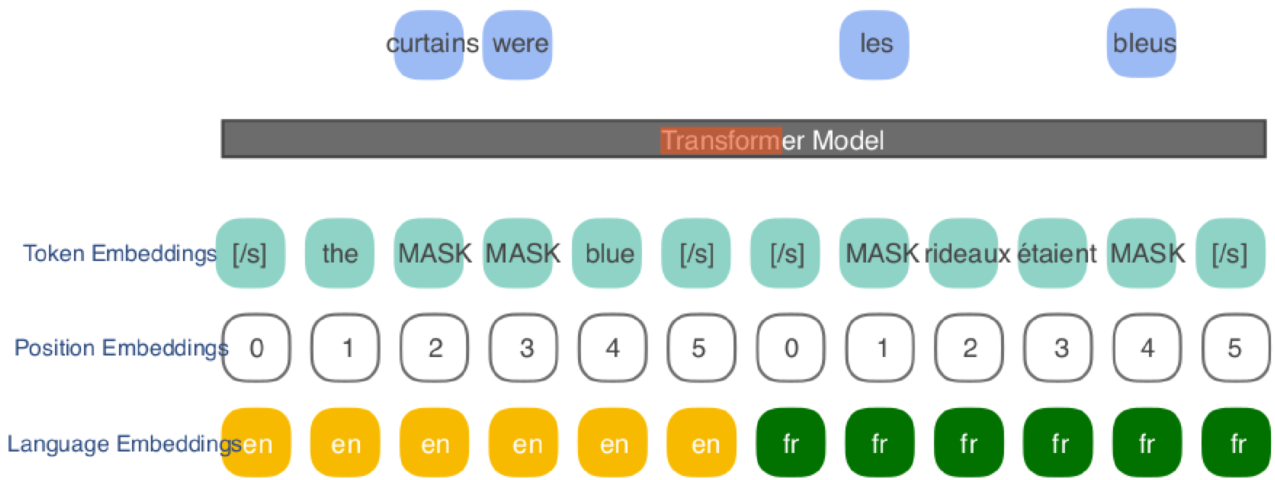




AMR Encoding

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

Multilingual Decoding



- Crosslingual embeddings (XLM Sentence Piece Model and Vocabulary)
- Language Models pretrained on 30M sentences (for each language)
- Multilingual decoding
  - Prefix each training instance with a control token
  - Trained on multilingual Europarl data

Multilingual Decoding

Decoding into 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

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

Decoding into 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.

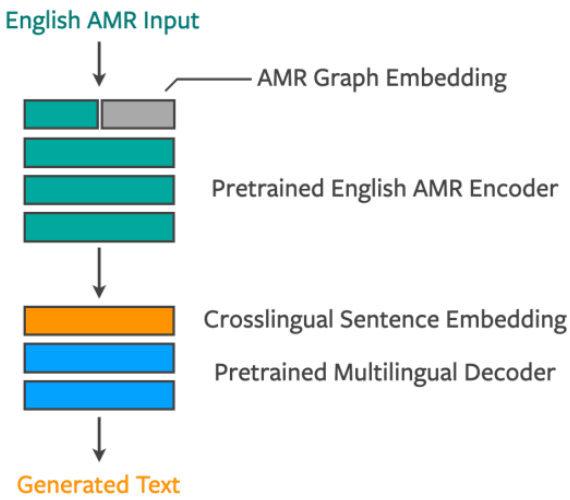
Training Data

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 expertgruppsmöt e i januari 2002 i New York.

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

Multilingual AMR-to-NL Model



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

Test Data

- Silver AMR: 21 languages, Europarl
- Gold AMR: 4 languages

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.

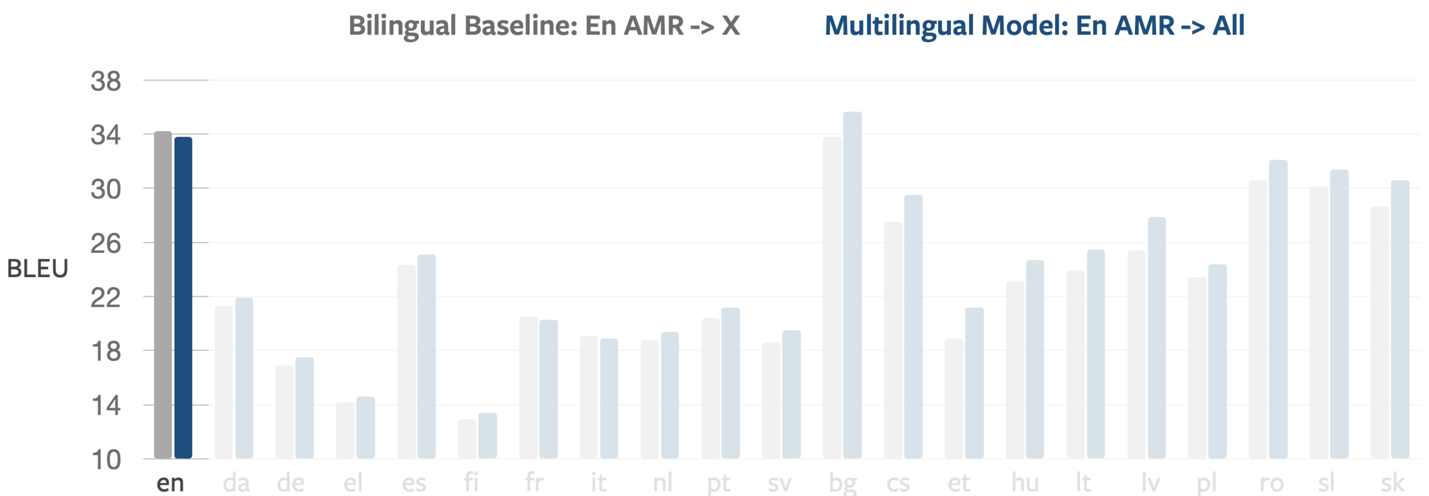
Multilingual Model

SV

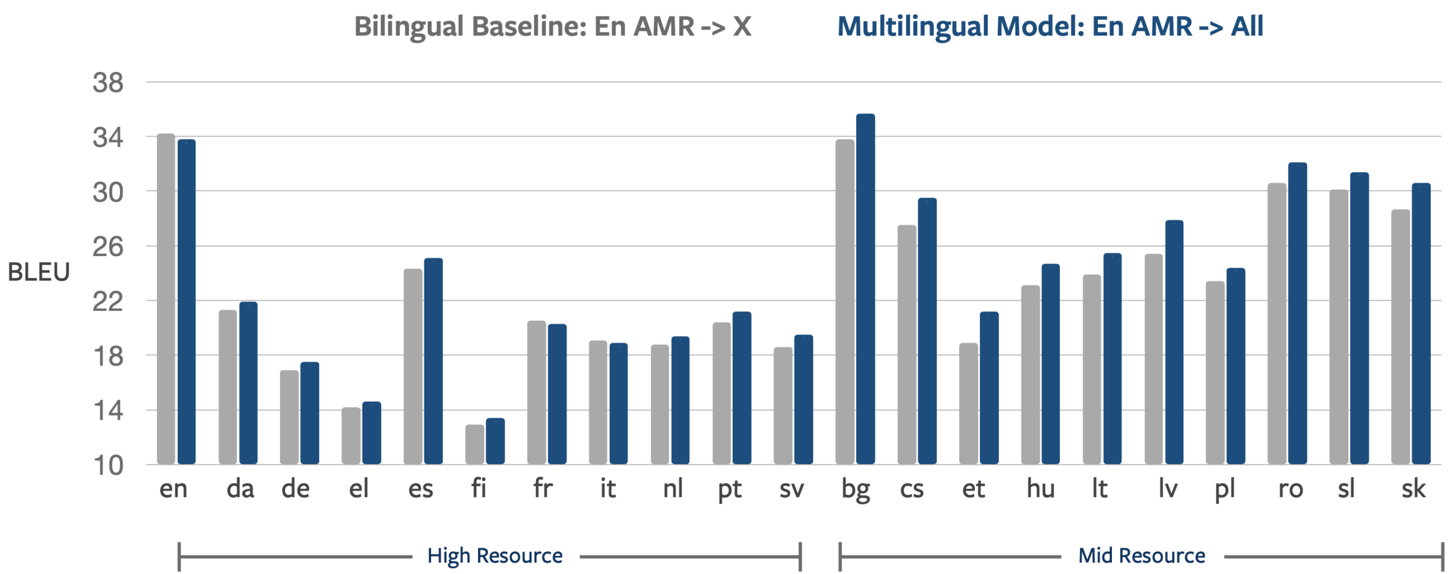
hold
:ARG0 person : ARG0-of have-org-role :ARG1 :op1
United :op2 States :ARG2 official
:ARG1 meet :ARG0 person :ARG1-of expert :ARG2-of group
:time date-entity :year 2002 :month 1
:location city :op1 New :op2 York

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

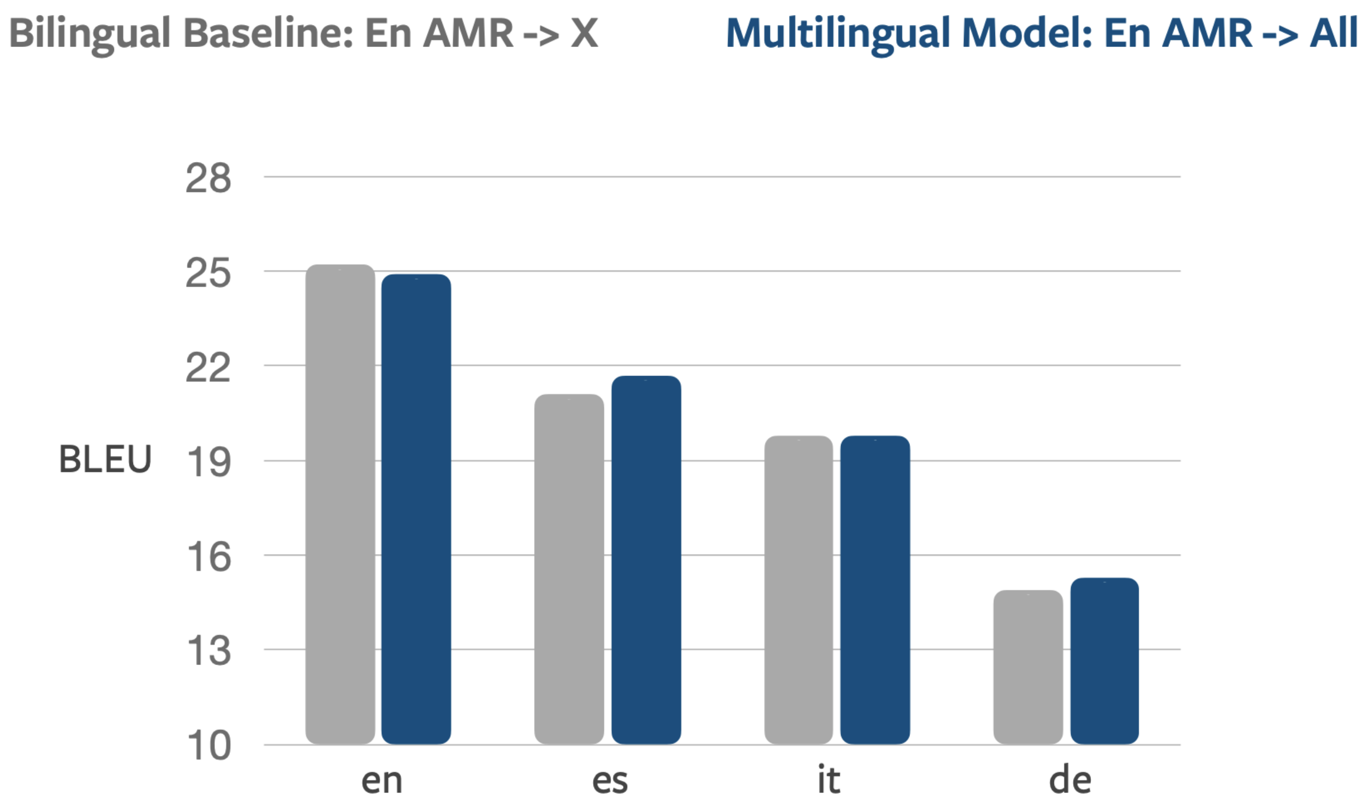
Results: Europarl



Results: Europarl



Results: Gold AMR



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.

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

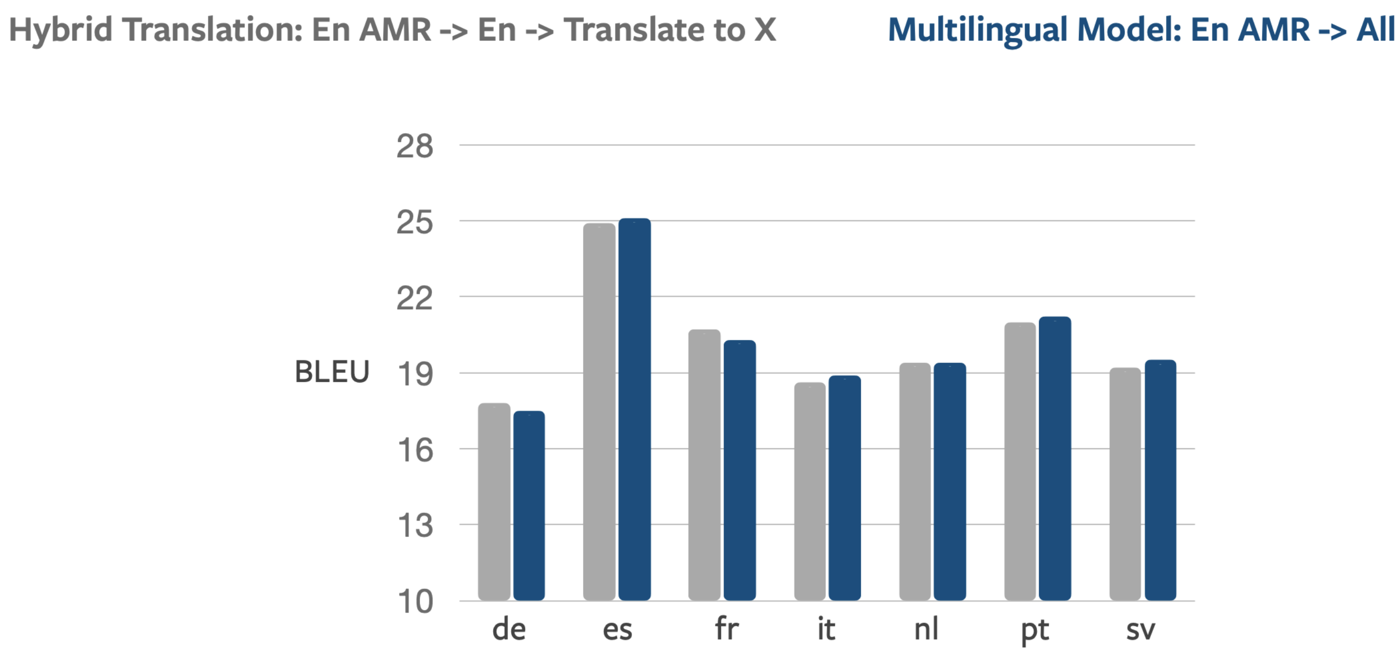
fr

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.

Comparison: Hybrid vs Multilingual



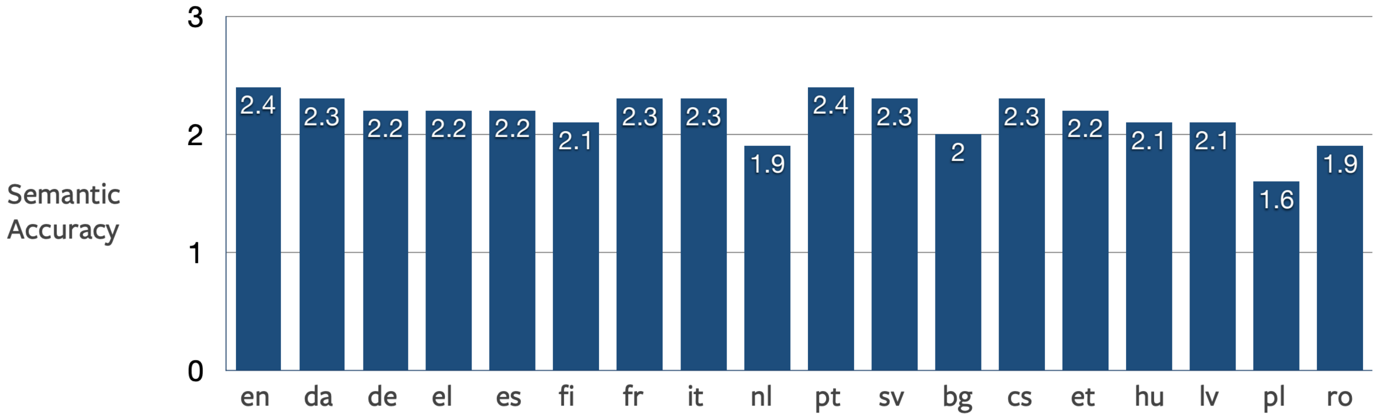
Human Evaluation

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

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: Semantic Accuracy



Key Takeaways

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

Key Takeaways

- Pre-training and Multilingual techniques permits bridging the gap between English-Centric AMRs and target languages with varied syntax and morphology
    - Pretrained LMs and AMRs, Crosslingual embeddings, Multilingual training

Key Takeaways

- Pre-training and Multilingual techniques permits bridging the gap between English-Centric AMRs and target languages with varied syntax and morphology
    - Pretrained LMs and AMRs, Crosslingual embeddings, Multilingual training
- Multilingual models benefits from increased training data and perform better on average than bilingual

Key Takeaways

- Pre-training and Multilingual techniques permits bridging the gap between English-Centric AMRs and target languages with varied syntax and morphology
  - Pretrained LMs and AMRs, Crosslingual embeddings, Multilingual training
- Multilingual models benefits from increased training data and perform better on average than bilingual
- Multilingual End-to-End models outperform NLG+MT models

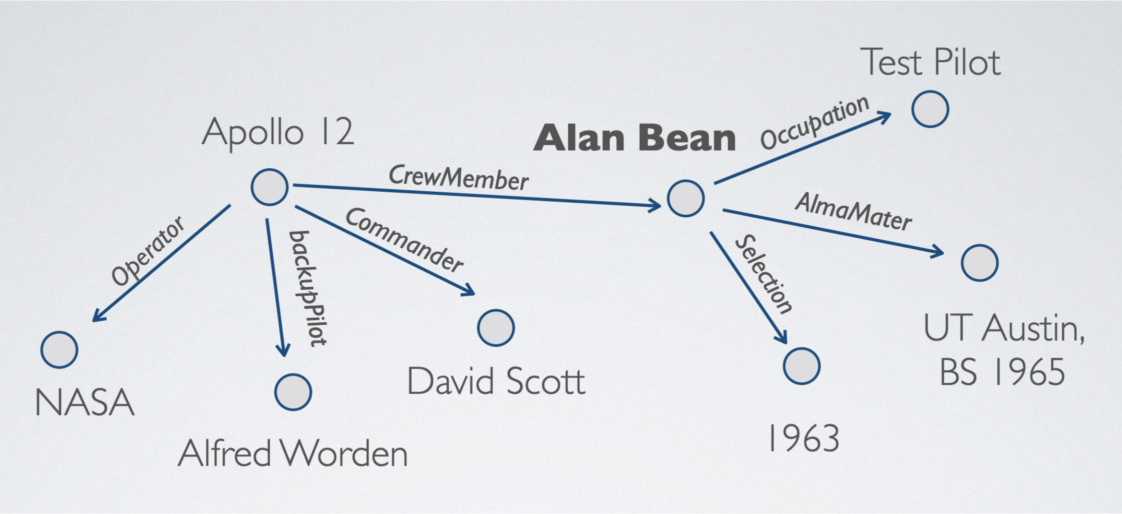
RDF $\Rightarrow$ Text

The WebNLG Challenge

Gardent et al. ACL 2017, Castro-Ferreira et al. 2020, Cripwell et al. 2023

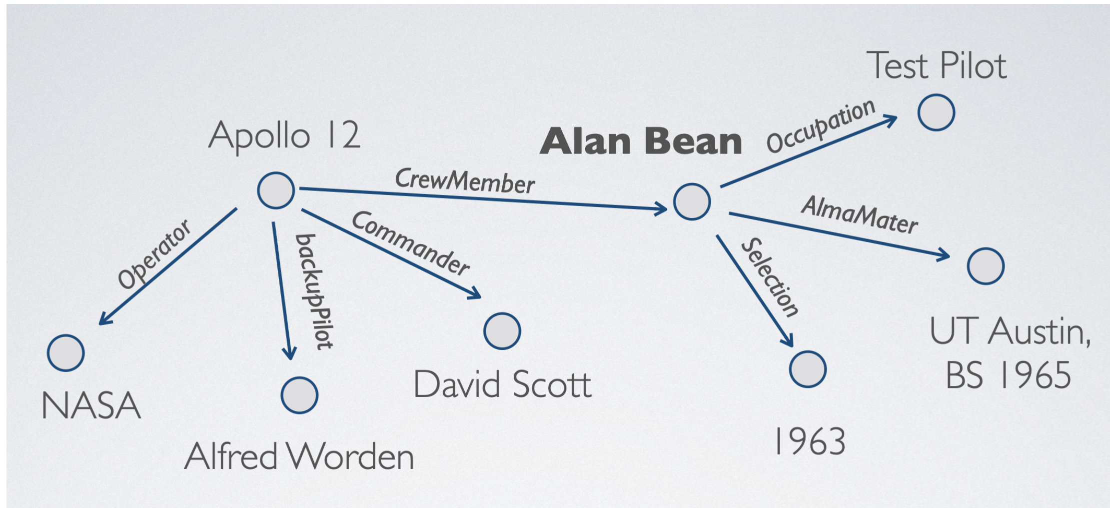
The WebNLG Challenge

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 ⇒ 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

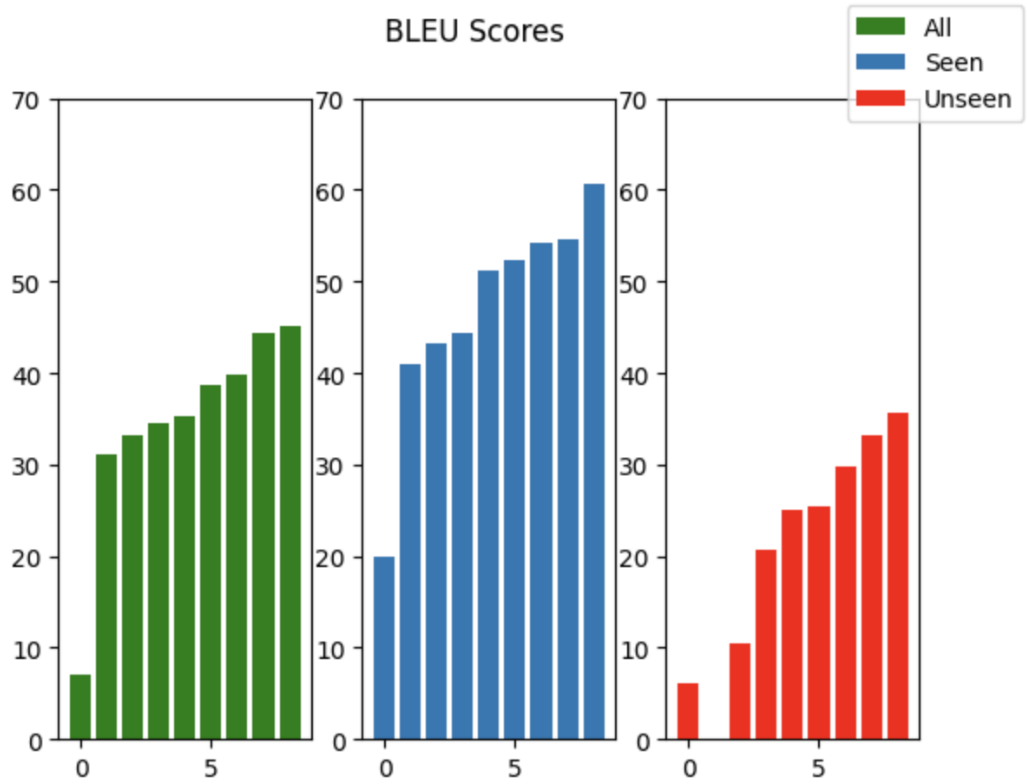
5 **unseen** categories:

- Athlete, Artist, MeanOfTransportation, CelestialBody, Politician

WebNLG 2017: RDF $\Rightarrow$ 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 2020

Natural Language Generation

- RDF $\Rightarrow$ English

WebNLG 2020

Natural Language Generation

- RDF $\Rightarrow$ English

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

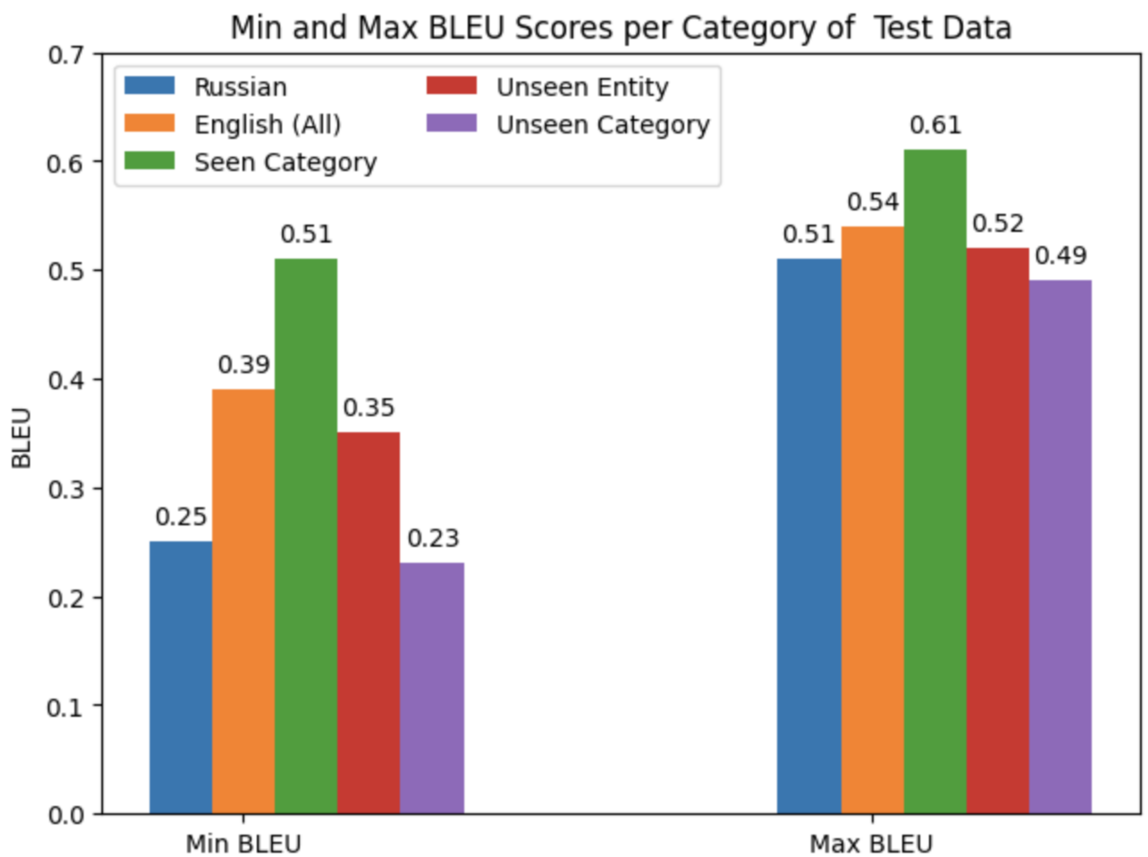
E.g., *Nie Haisheng* in category *Astronaut*

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 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: 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	-	✓	✓	✓	-

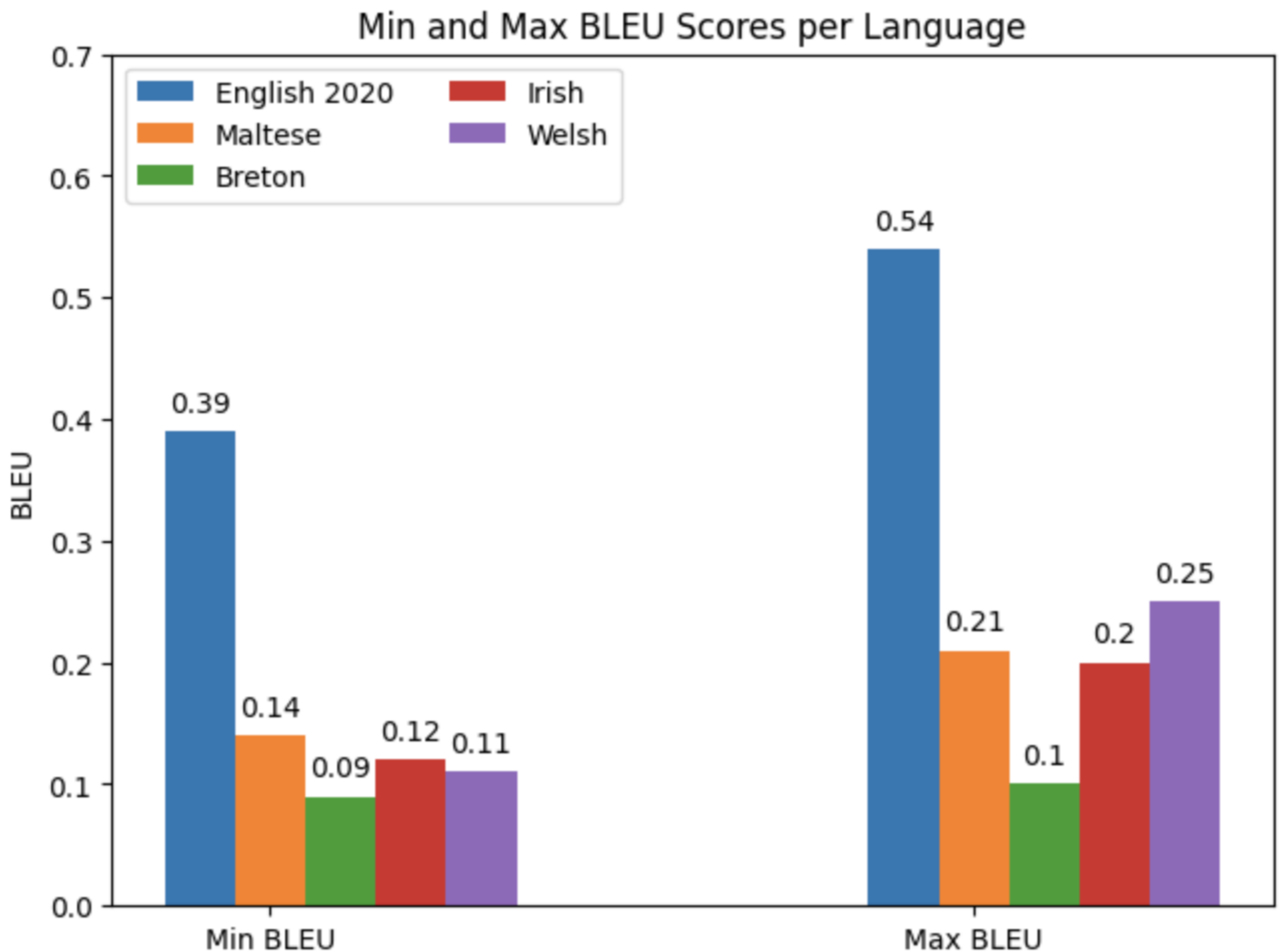
RDF $\Rightarrow$ English

- 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



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

Pipeline vs. End-to-End

For Breton, there is no (good) MT system

✗ NLG+MT pipeline

Pipeline vs. End-to-End

For Breton, there is no (good) MT system

✗ NLG+MT pipeline

✗ Full-fine tuning (BLEU: 0.10)

Pipeline vs. End-to-End

For Breton, there is no (good) MT system

✗ NLG+MT pipeline

✗ Full-fine tuning (BLEU: 0.10)

✓ Parameter Efficient Fine Tuning (PEFT)

Pipeline vs. End-to-End

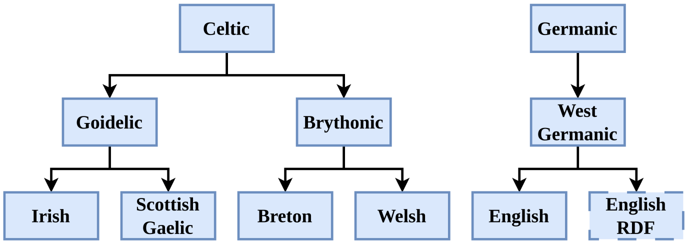
For Breton, there is no (good) MT system

✗ NLG+MT pipeline

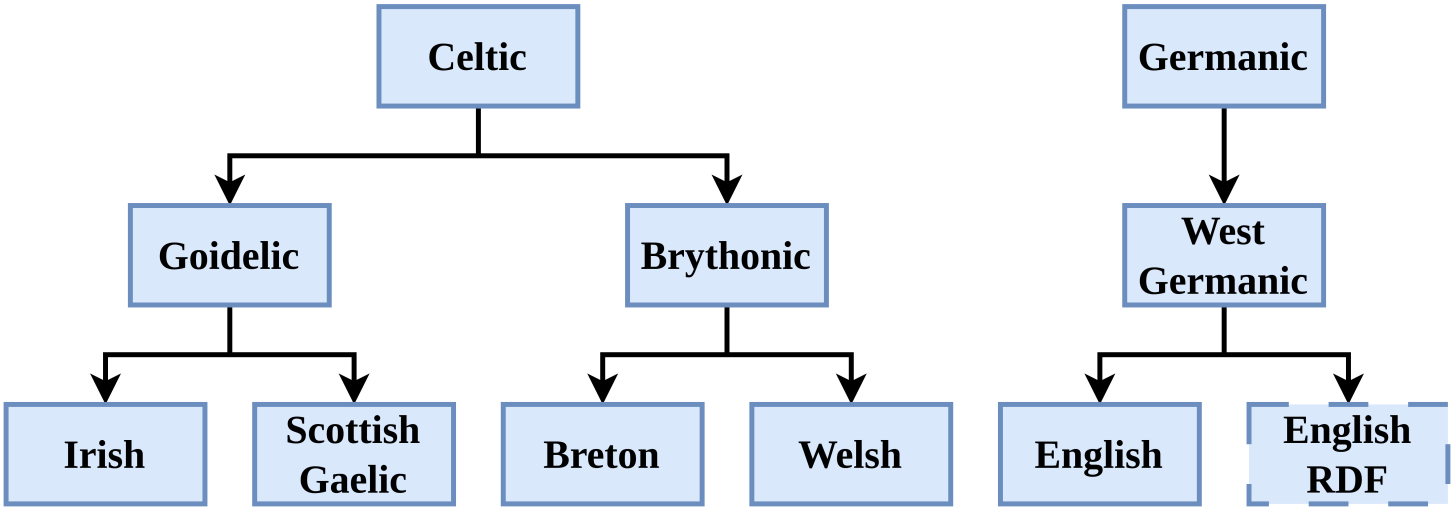
✗ Full-fine tuning (BLEU: 0.10)

✓ Parameter Efficient Fine Tuning (PEFT)

- Soft-Prompt
- Structured to capture language relatedness and various tasks



Phylogenetic Tree



Soft Prompt

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

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>

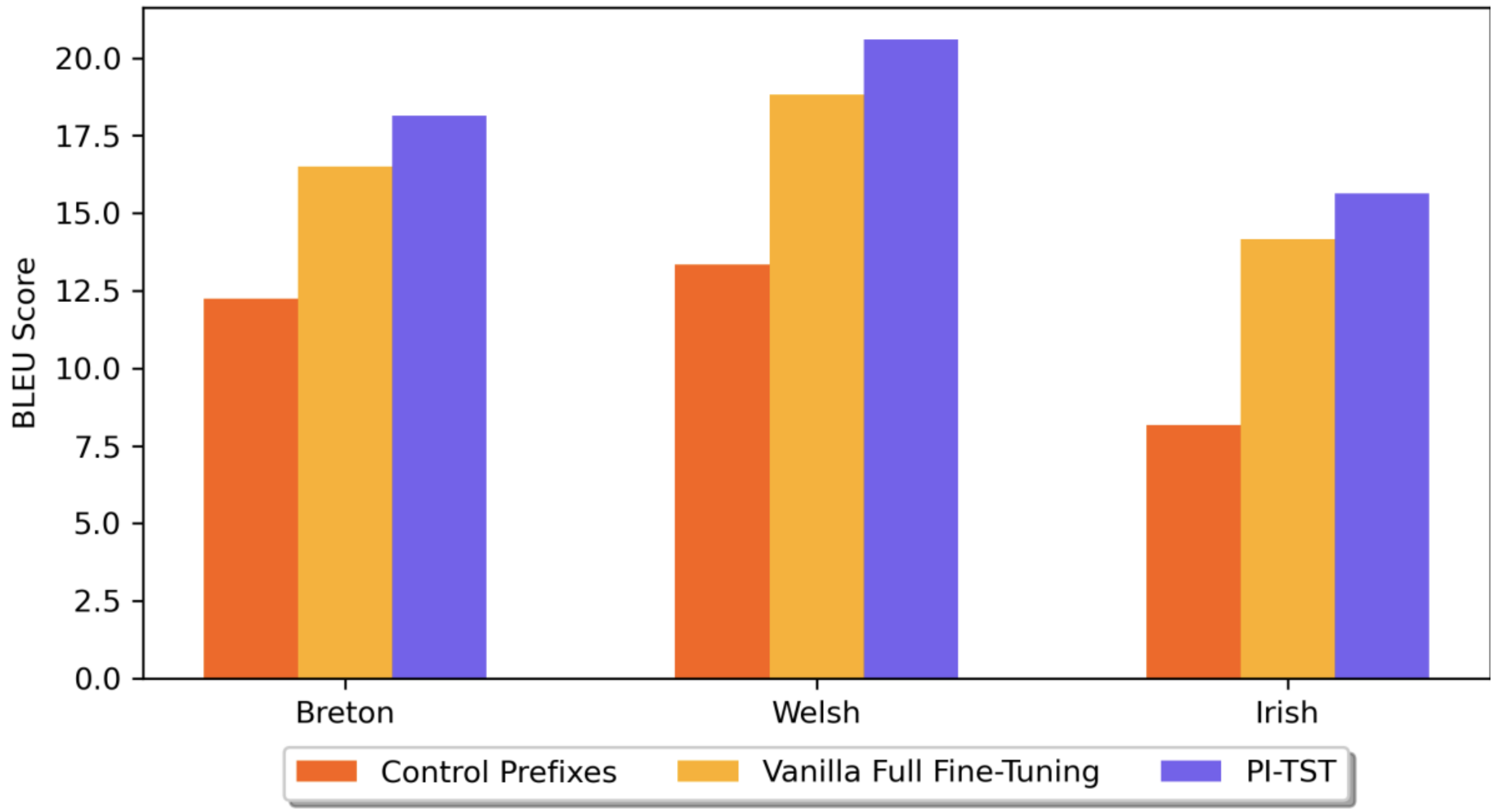
Step 2: Fine-Tuning on Dev RDF-to-Text data (RDF-to-Text Models)

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

Inference

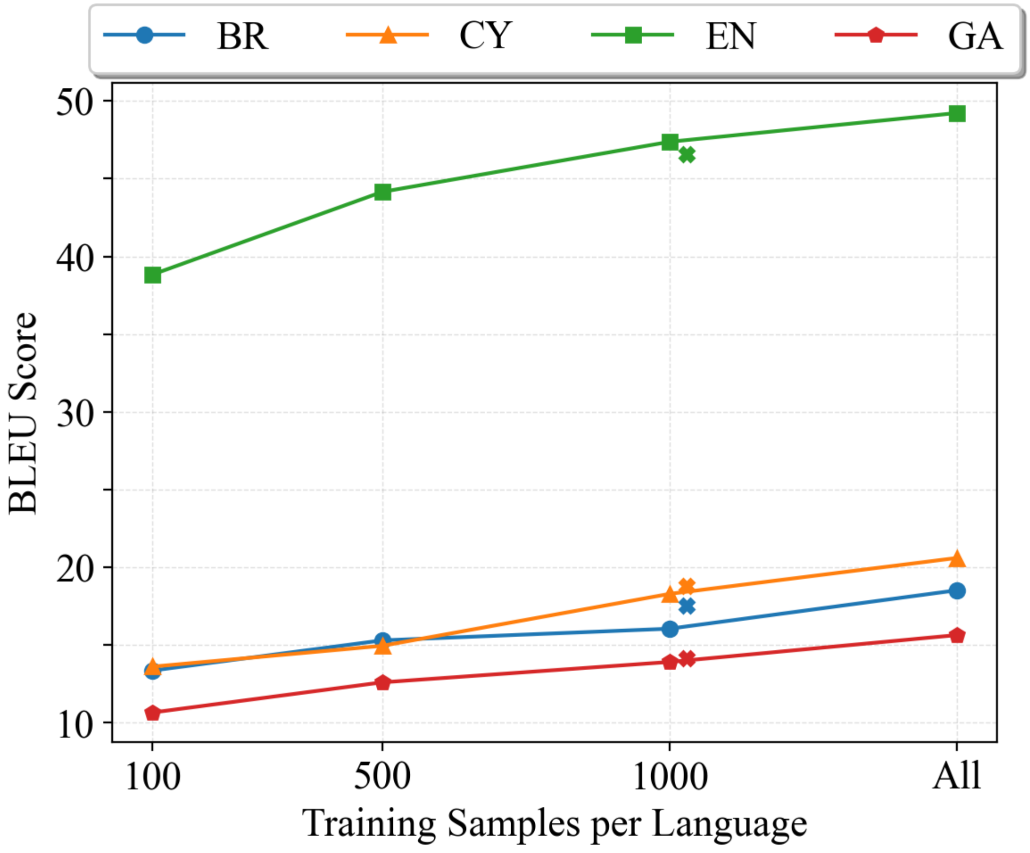
The Language sub-prompt is set to the target language.

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

Impact of Data Size



Key Takeaways

- Pretraining (2017 vs 2020) improves performance

Key Takeaways

- Pretraining (2017 vs 2020) improves performance
- Performance is lower for Russian than for English

Key Takeaways

- Pretraining (2017 vs 2020) improves performance
- Performance is lower for Russian than for English
- Performance degrades on out of domain data (unseen)

Key Takeaways

- Pretraining (2017 vs 2020) improves performance
- Performance is lower for Russian than for English
- 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 is lower for Russian than for English
- Performance degrades on out of domain data (unseen)
- Performance is very poor for Low Resource Languages (2023)
- Most LRL Models are NLG+MT pipelines

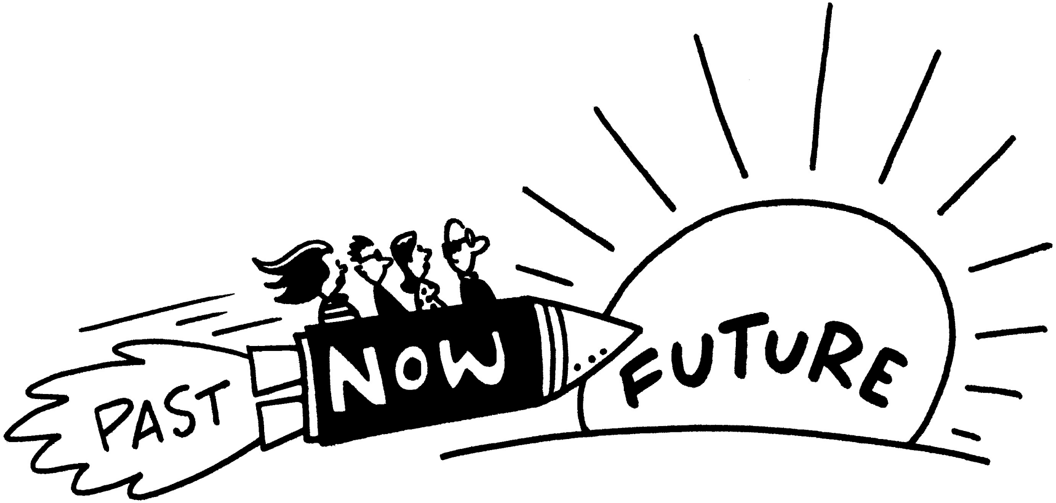
Key Takeaways

- Pretraining (2017 vs 2020) improves performance
- Performance is lower for Russian than for English
- Performance degrades on out of domain data (unseen)
- Performance is very poor for Low Resource Languages (2023)
- Most LRL Models are NLG+MT pipelines
- LRL for which no MT models exist (Breton) show very poor results for these models

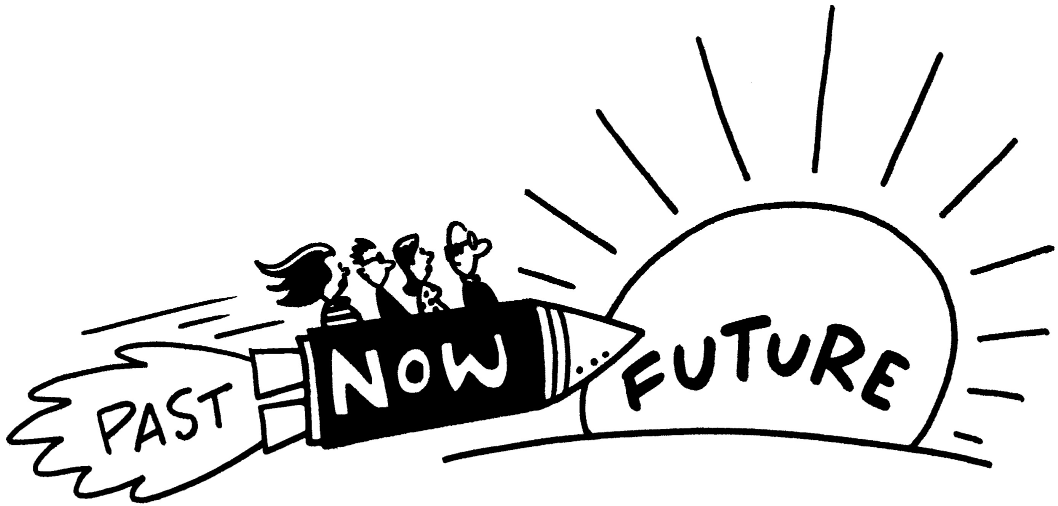
Key Takeaways

- Pretraining (2017 vs 2020) improves performance
- Performance is lower for Russian than for English
- Performance degrades on out of domain data (unseen)
- Performance is very poor for Low Resource Languages (2023)
- Most LRL Models are NLG+MT pipelines
- LRL for which no MT models exist (Breton) show very poor results for these models
- PEFT techniques help improve performance for these languages

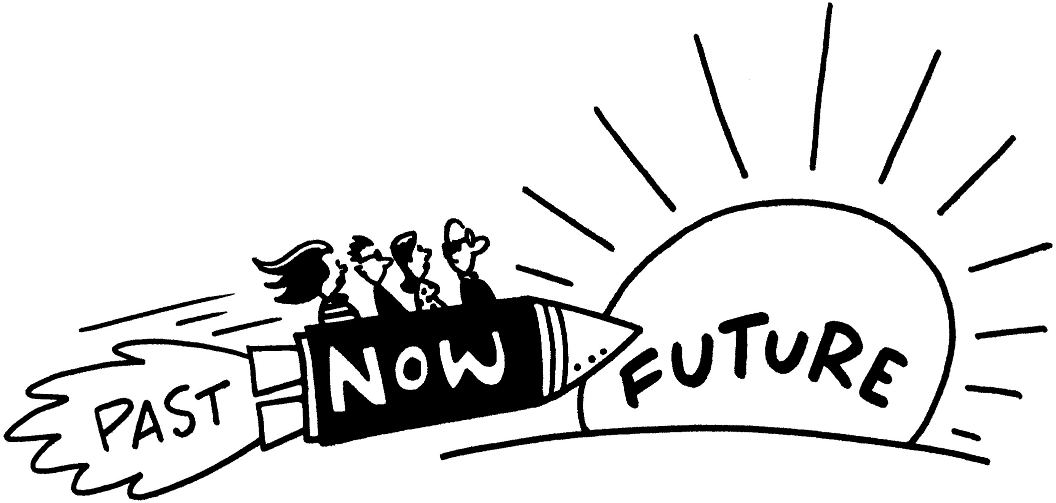
    BLEU for Breton: 10 (NLG+MT) $\rightarrow$ 18.15 (PEFT E2E Model)



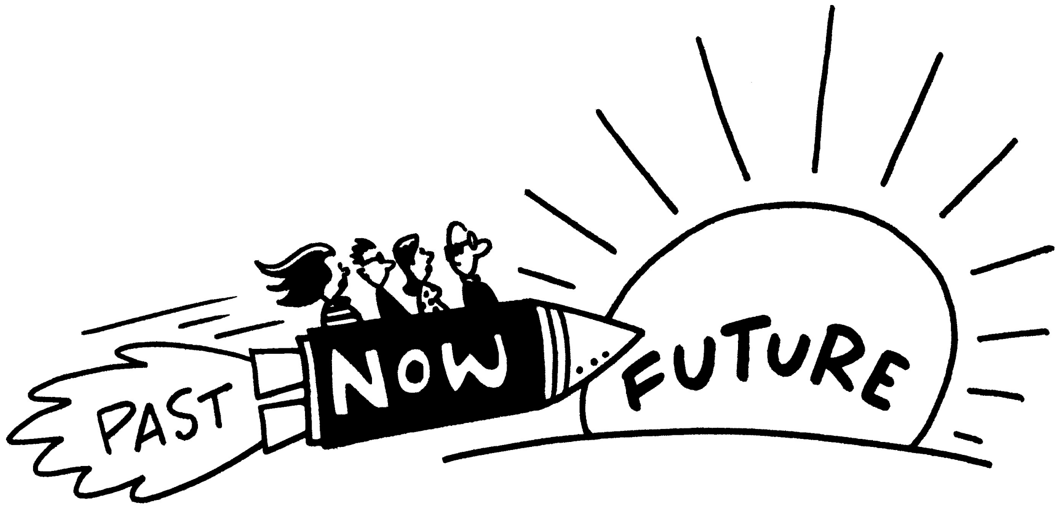
- LLMs for simplification, graph-to-text



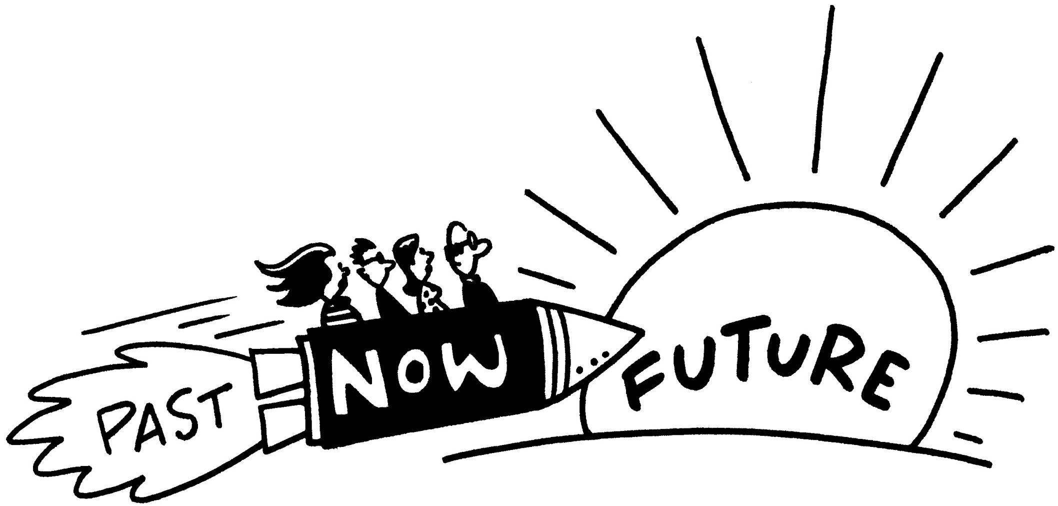
- LLMs for simplification, graph-to-text
    - Prompt engineering, chain-of-thought decomposition



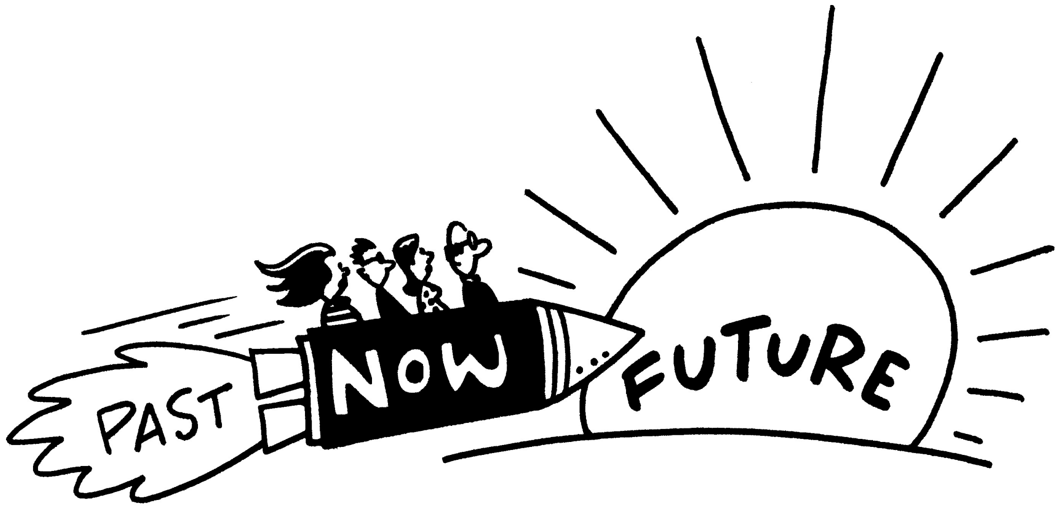
- LLMs for simplification, graph-to-text
    - Prompt engineering, chain-of-thought decomposition
    - LLMs for evaluation



- LLMs for simplification, graph-to-text
  - Prompt engineering, chain-of-thought decomposition
  - LLMs for evaluation
- Generalising RDF-to-Text Models
  - to other languages and other domains



- LLMs for simplification, graph-to-text
  - Prompt engineering, chain-of-thought decomposition
  - LLMs for evaluation
- Generalising RDF-to-Text Models
  - to other languages and other domains
- Evaluation



- LLMs for simplification, graph-to-text
  - Prompt engineering, chain-of-thought decomposition
  - LLMs for evaluation
- Generalising RDF-to-Text Models
  - to other languages and other domains
- Evaluation
  - Metrics for simplificity and Meaning Preservation

Le Scao and Gardent AACL-IJCNLP Findings 2023, Cripwell et al. EMNLP 2023

Questions ?