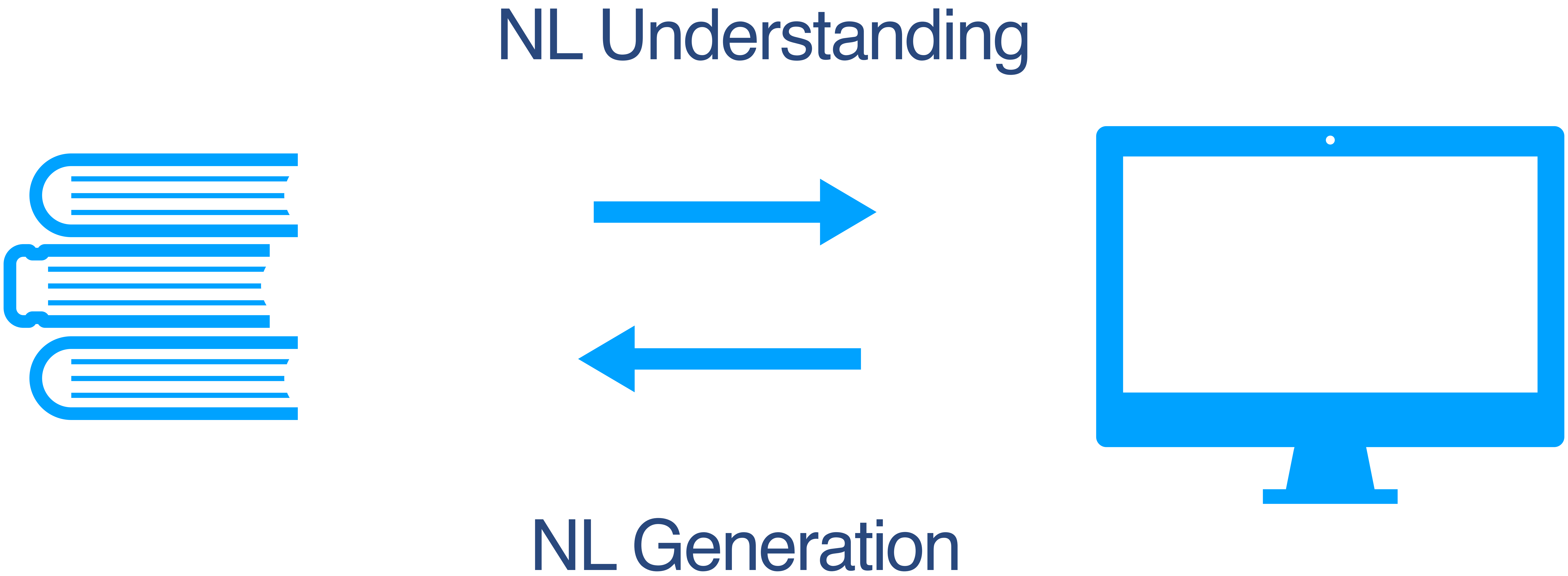
Generating Text

Claire Gardent
Joint work with Angela Fan (Facebook), Antoine Bordes (Facebook) and Chloé Braud (CNRS/IRIT)

cnrs dépasser les frontières

Loria Laboratoire lorrain de recherche en informatique et ses applications

Natural Language Processing



LORIA NLP

Multispeech

- Multimodal Speech, Speech Recognition, Speech-to-Speech Translation

Orpailleur

- Mining Knowledge from Text

Semagramme

- Logic-based models, methods and tools for the semantic analysis of natural Language

Smart

- Machine Translation and Speech Recognition

SYNALP

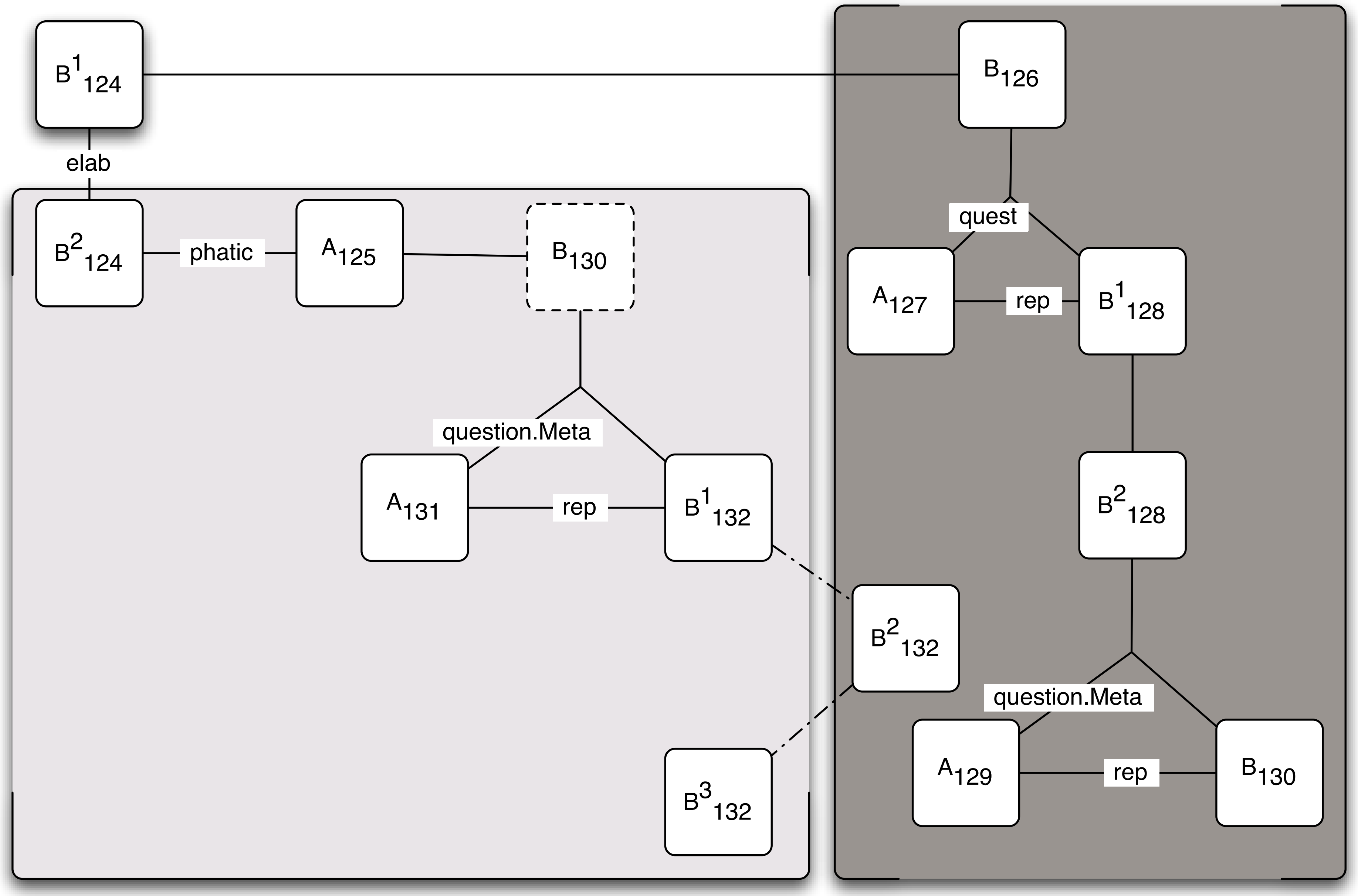
- Natural Language Generation, Human-Machine Dialog, Clustering

Semagramme

NLU for mental health

Maxime Amblard, Michel Musiol

- Incoherence detection in Schizophrenia speech
- Using formal Semantics and AI
- Useful to support
  - Diagnosis
  - Early illness detection
  - Relaps prevention
  - Long term monitoring



Multispeech

Franco-German ANR-DFG project

M-PHASIS (2018-2022)

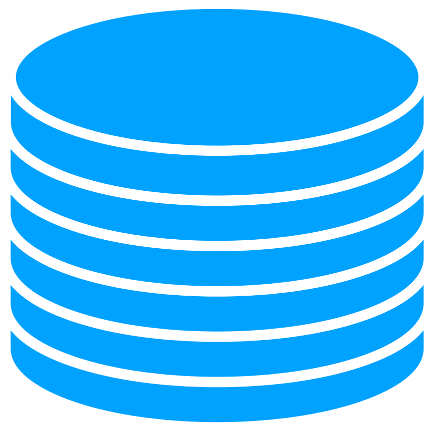
Irina Illina, Dominique Fohr, 3 PhDs

Neural Models for Hate Speech Detection

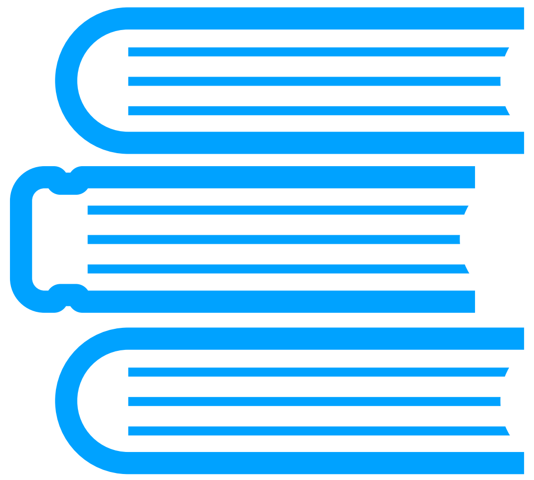
- Un- and weakly supervised models
- Generalizing well to different abusive corpora
- Integrating multi-word expressions
- Devising data augmentation techniques to compensate for the lack of labeled data

Applications

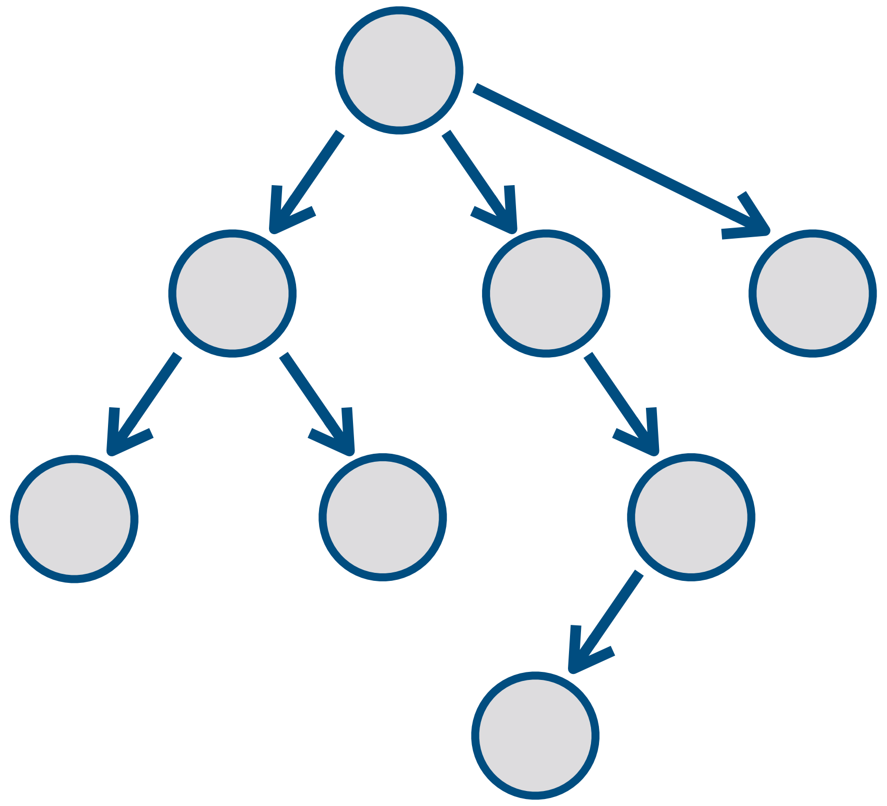
What is NLG useful for ?



Verbalising, Summarising, Querying Knowledge-Bases



Summarising, Simplifying, Paraphrases one or more Text(s)



Converting Graphs into Text

Outline

- Neural Networks

- Neural Generation
  - Embeddings: Representing words
  - Language Models: Generating words

- Four Challenges for Neural NLG

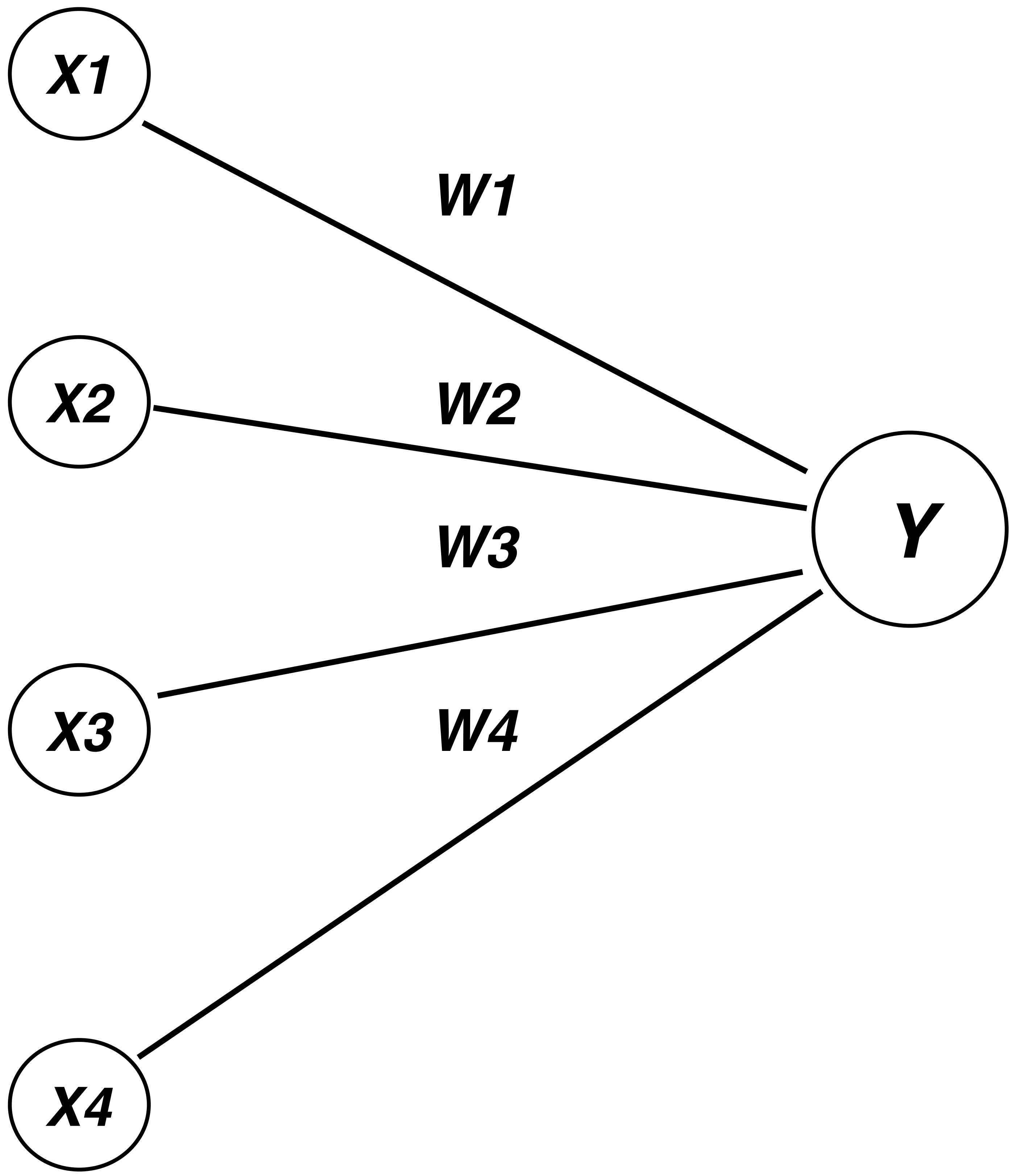
Neural Networks

Neuron

A neuron computes an activation

value $y = g(\sum_{i=1}^{n} w_i x_i)$

g is an activation function



Activation Functions

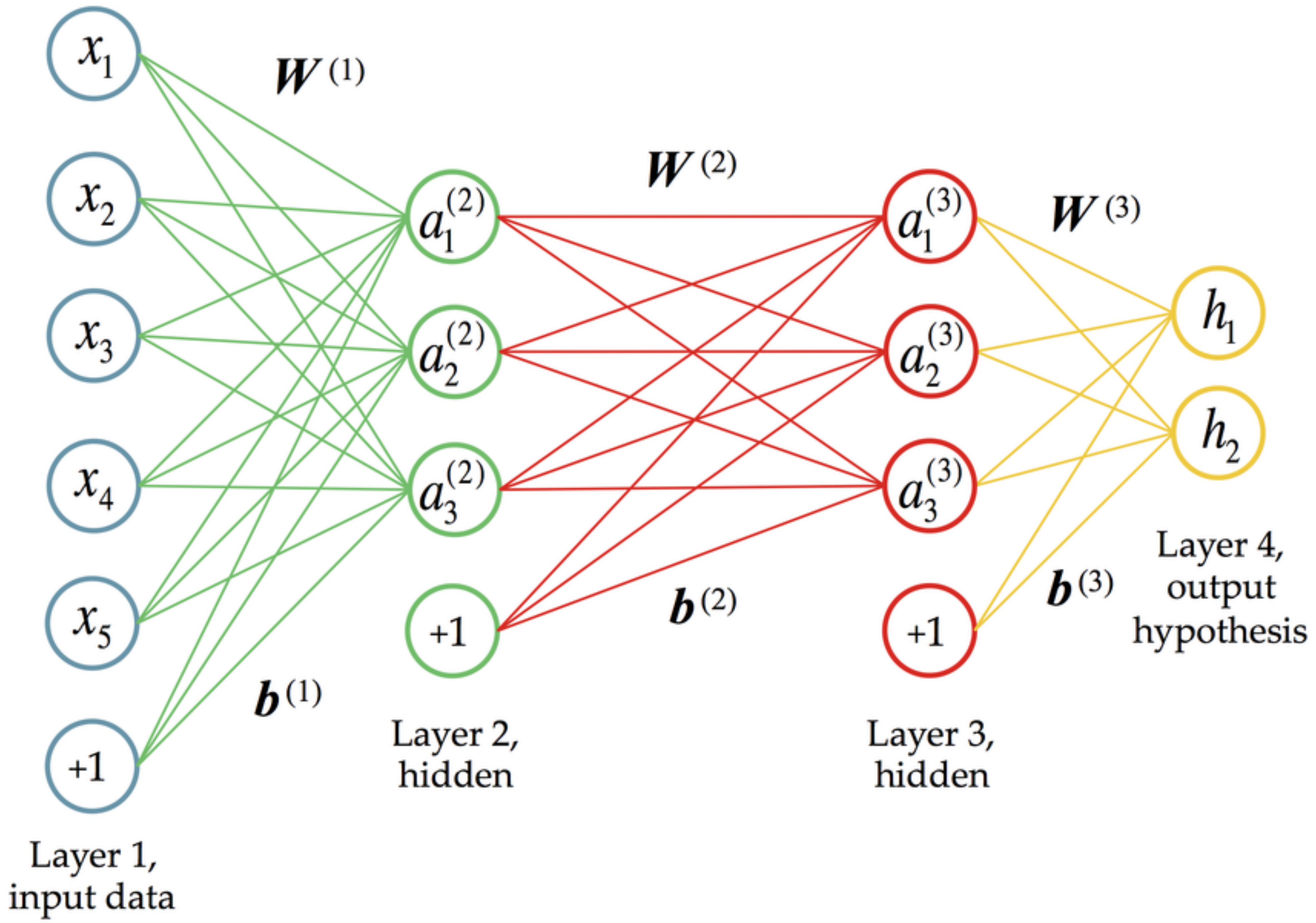
Sigmoid	$g(x) \in (0,1)$	$\sigma(x) = \dfrac{1}{1 + e^{-x}}$
Tanh	$g(x) \in (-1,1)$	$tanh(x) = \dfrac{2}{1 + e^{-2x}} - 1$
ReLu	$g(x) \in (0,\infty)$	$ReLu(x) = 0 \ if \ x < 0 \ else \ x$

Neural Network

- Each neuron produces a value which is the input to the next layer

- ***Weights are learned*** using back propagation and stochastic gradient descent

- “Good” weights allow the model to correctly predict the output given a new input



Neural Generation

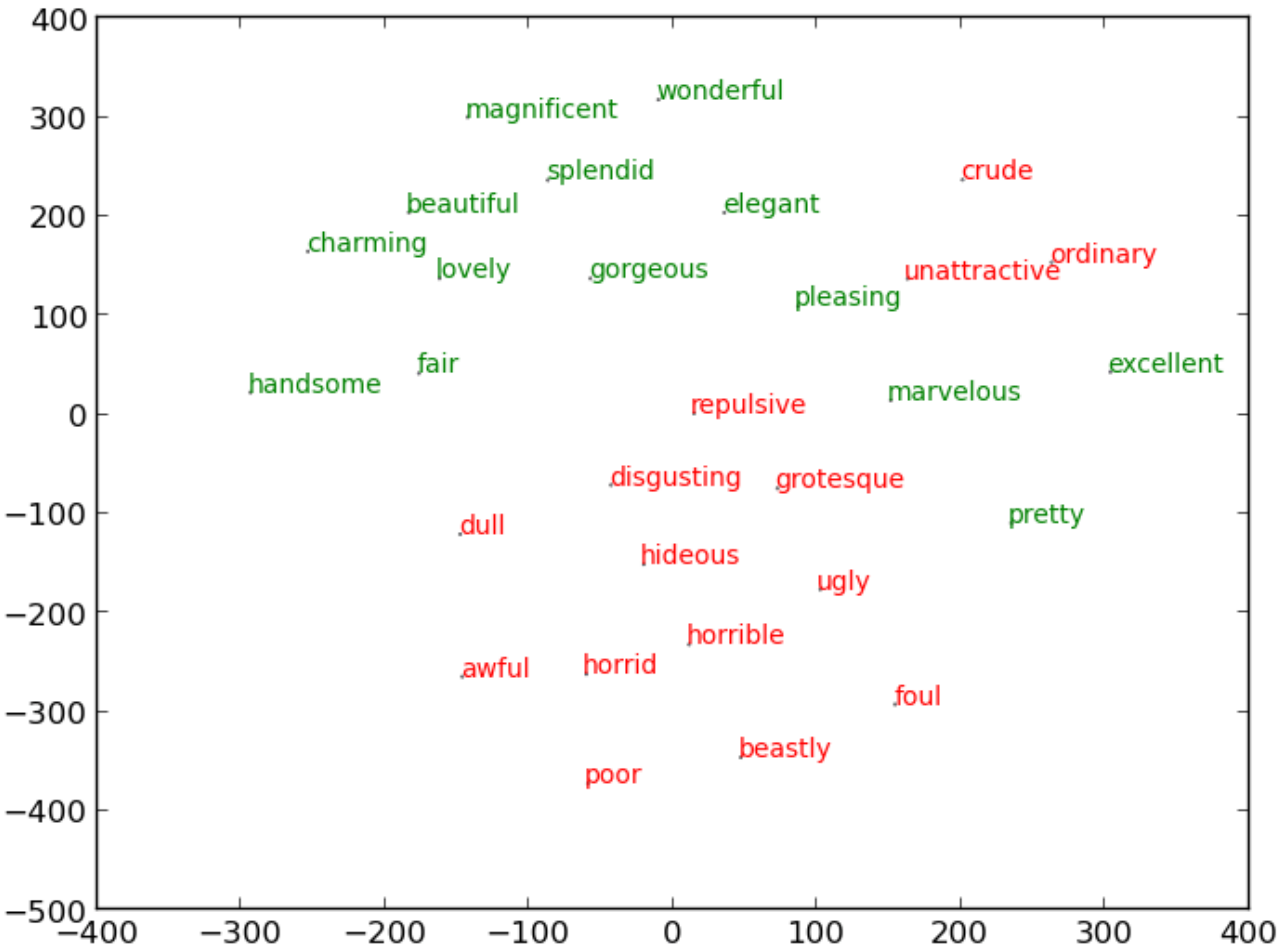
Neural Generation

- Represents words as vectors of real numbers called ***embeddings***
- Generates text by ***predicting the next most probable word***

Embeddings

Neural Word Representations

- Words are represented by vectors of real numbers called ***embeddings***
- Embeddings are learned on very large text corpora
- The embedding of words with similar meaning (contexts) are close in the vector space



Language Model

How probable is a sequence of words ?

Determines the probability of a sequence of words

$$P(W) = P(w_1, w_2, w_3 \ldots w_n)$$

Ex (in english)

$$P_1 = P(\text{"a quick brown dog"})$$
$$P_2 = P(\text{"dog quick a brown"})$$
$$P_3 = P(\text{"un chien quick brown"})$$
$$P_4 = P(\text{"un chien brun rapide"})$$

- $P_1 > P_2 > P_3 > P_4$

The probability of a sequence of words can be computed using the chain Rule of Probability

$$P(w_1, w_2, \ldots, w_n) = \prod_i P(w_i \mid w_1, w_2, \ldots, i_{i-1})$$

Example

$P(\text{its water is so transparent}) =$
$\quad P(\text{its}) \times P(\text{ water} \mid \text{its}) \times P(\text{ is} \mid \text{its water }) \times P(\text{so} \mid \text{its water is }) \times (\text{ transparent} \mid \text{its water is so})$

Language Models

Generating Words

Neural Models generate a sentence by auto-regressively predicting the next word given a previous context

France is where I grew up and where I now work. I speak fluent ***??***

p(***French*** | France is where I grew up and where I now work. I speak fluent)

>

p(*English* | France … fluent)

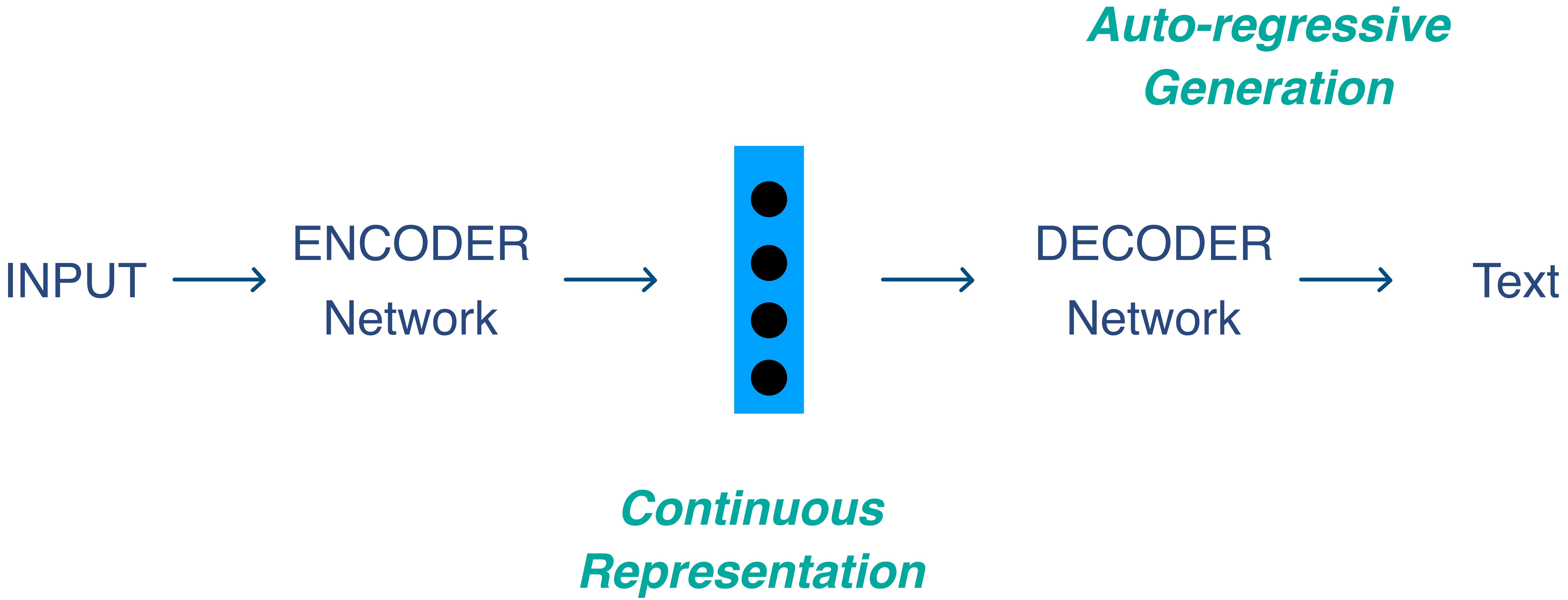
>

p(*Pizza* | France … fluent)

>

p(*the* | France … fluent)

The Encoder-Decoder Model



Encoders

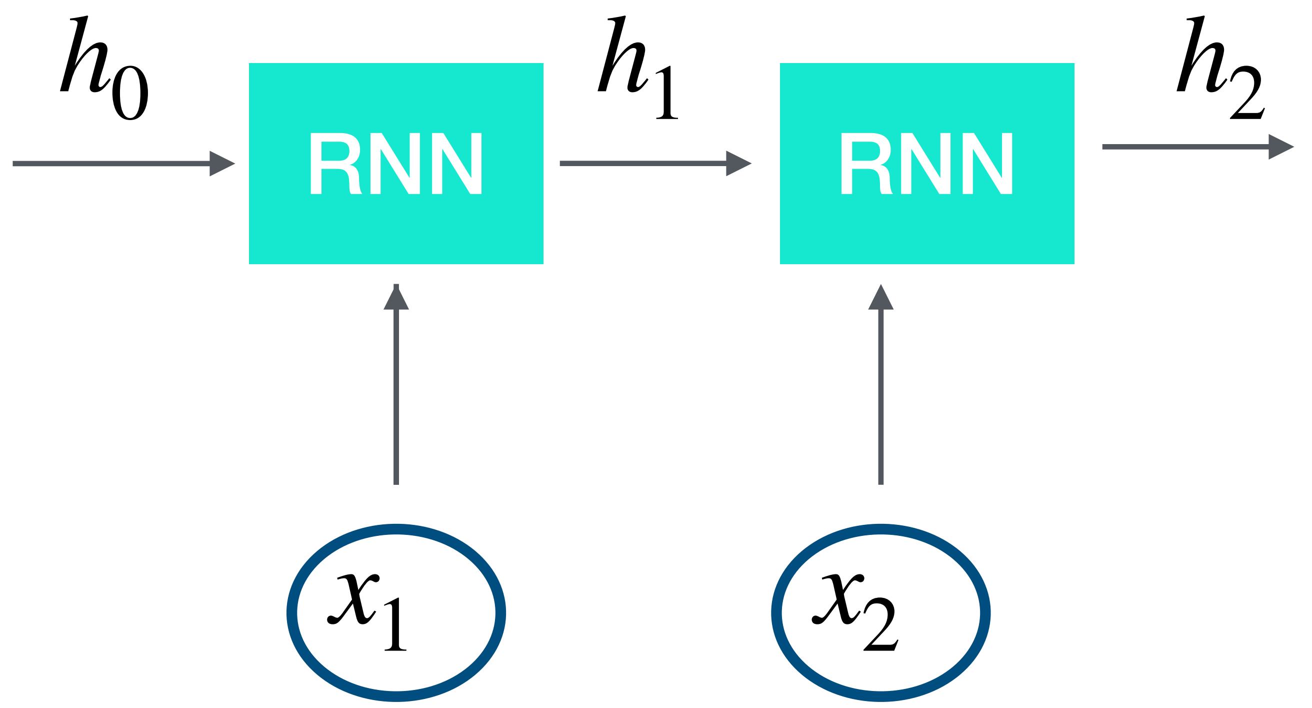
- Recurrent Neural Network (sequences)
- Convolutional Neural Network (Images and Text)
- Graph Encoder (Knowledge Bases, Tabular Data, RDF store)
- Transformer

Decoders

- Recurrent Neural Network
- Transformer

Encoding the Input with a Recurrent Neural Network

- For sequences
- Recurs over the input
- Outputs a new hidden state at each step



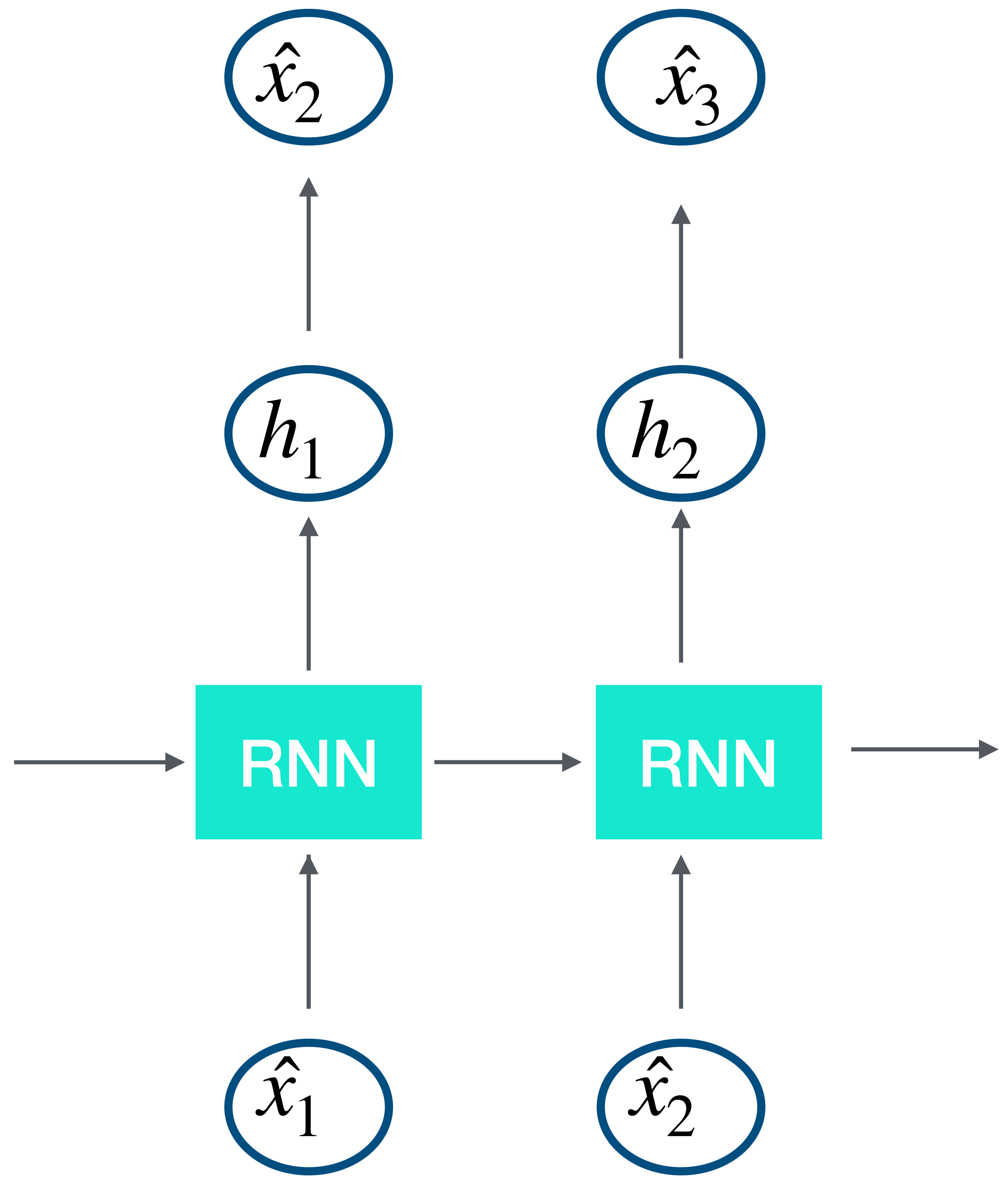
$$h_t = tanh(W_1 h_{t-1} + W_2 x_t)$$

The last hidden state is the input representation

Decoding with a Recurrent NN

Outputs a word at each step

- Softmax over the output vocabulary
- Sample from the output probability distribution
- The predicted word is the input to the next decoding step



$$x_t = softmax(W_t h_t)$$

Four Challenges for Neural Generation

Challenges for Neural NLG

- Generating from long Input

Challenges for Neural NLG

- Generating from Dealing long Input
- Retrieving and Integrating Relevant Knowledge

Challenges for Neural NLG

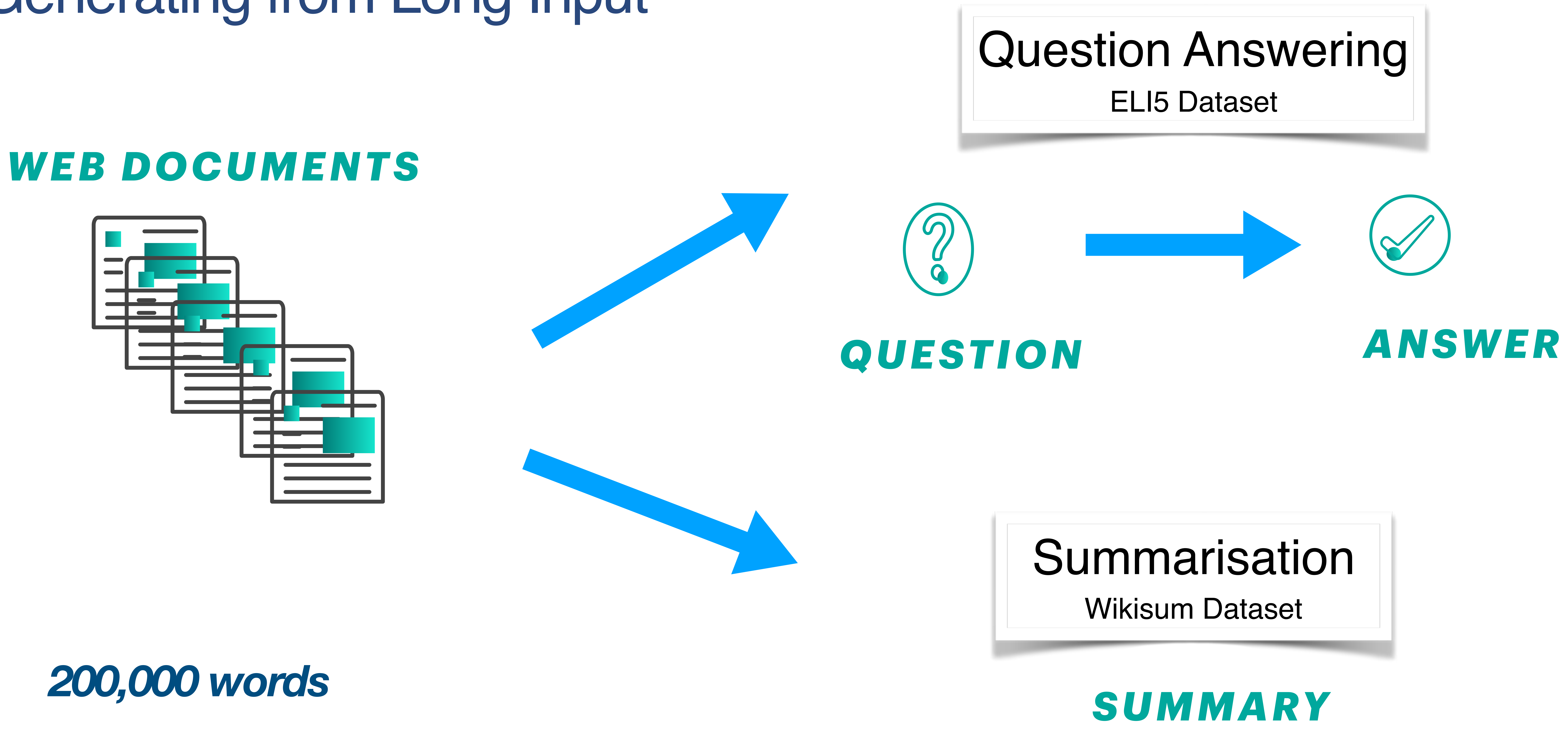
- Generating from long Input
- Retrieving and Integrating Relevant Knowledge
- Generating into Languages other than English

Challenges for Neural NLG

- Generating from Long Input
- Retrieving and Integrating Relevant Knowledge
- Generating into Languages other than English
- Generating Long Form Text

Handling Long Input

Generating from Long Input

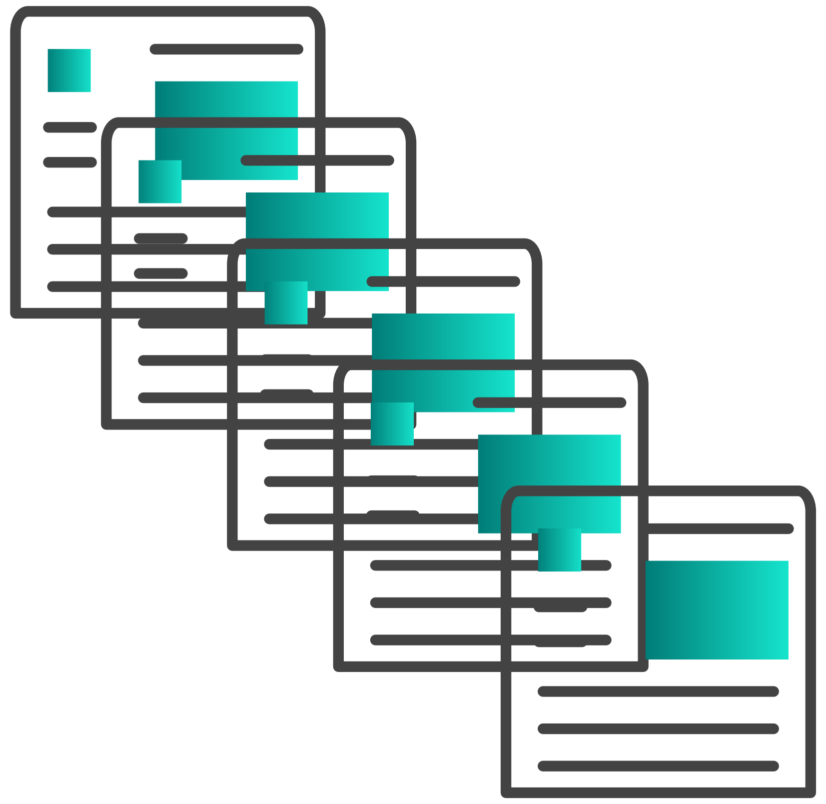


Question Answering

Explain Like I'm Five Dataset

270,000 TRAINING INSTANCES

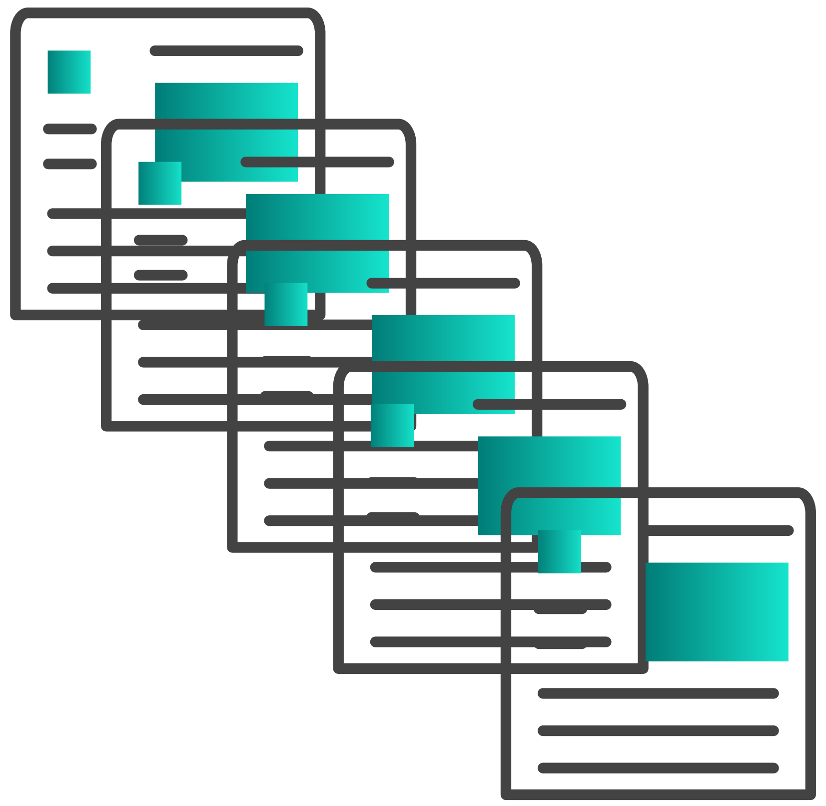
QUESTION + **WEB DOCUMENTS** + **ANSWER**



200,000 words

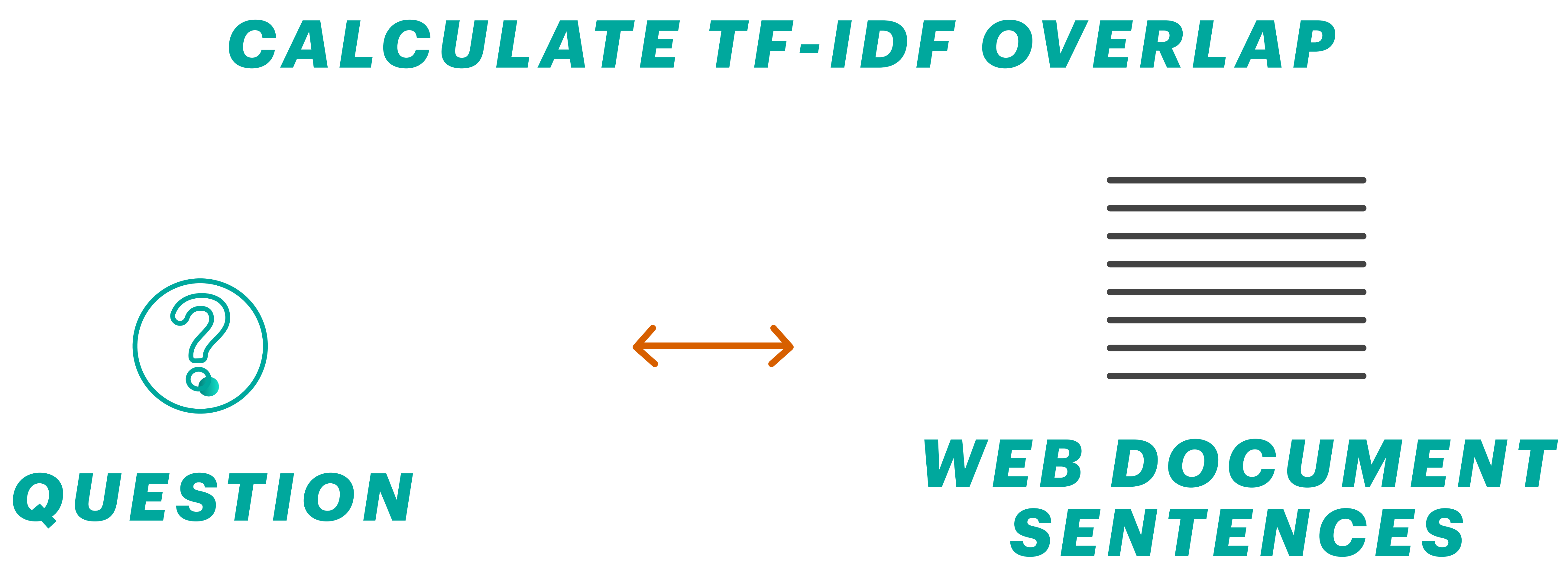
Dealing with Long Web Input

WEB DOCUMENTS

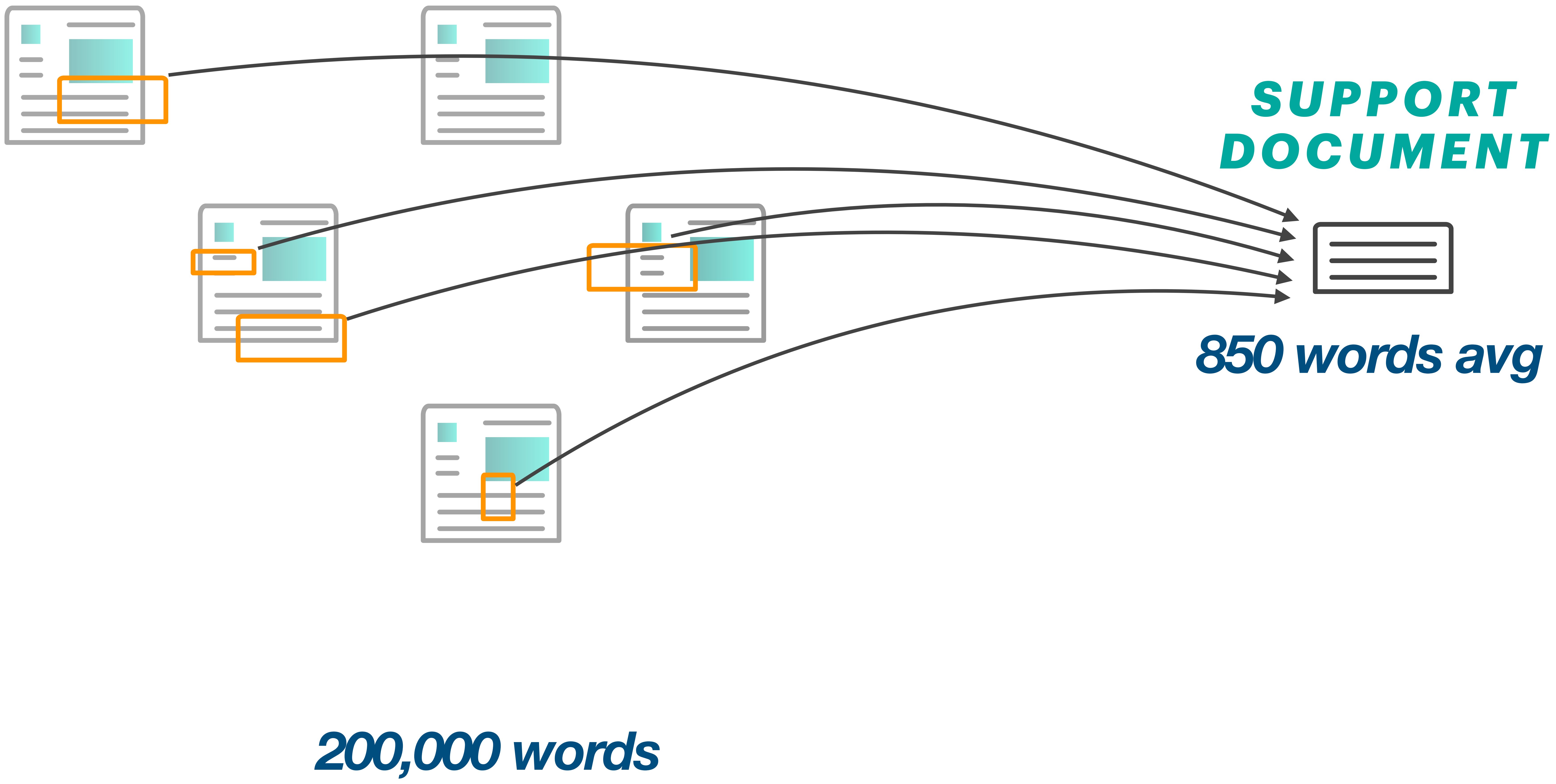


Over 200,000 words long

Creating a Shorter Support Document



Creating a Shorter Support Document



Downsides of Short Support Document

SUPPORT DOCUMENT

850 words avg

40% of the Answer Tokens are Missing

Downsides of Short Support Document

SUPPORT DOCUMENT

850 words avg

40% of the Answer Tokens are Missing

Information selected is Redundant

Downsides of Extractive Support Document

SUPPORT DOCUMENT

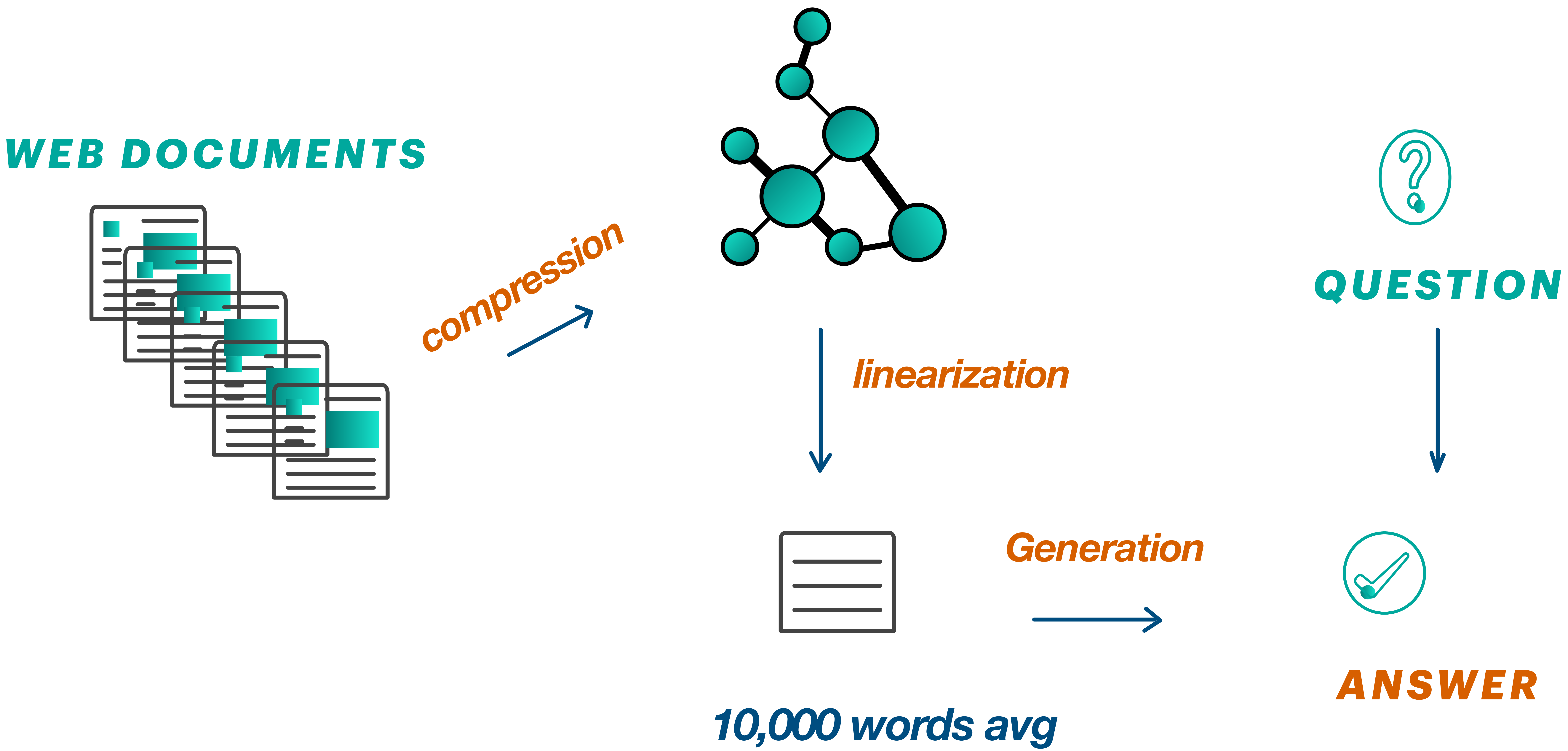
850 words avg

40% of the Answer Tokens are Missing

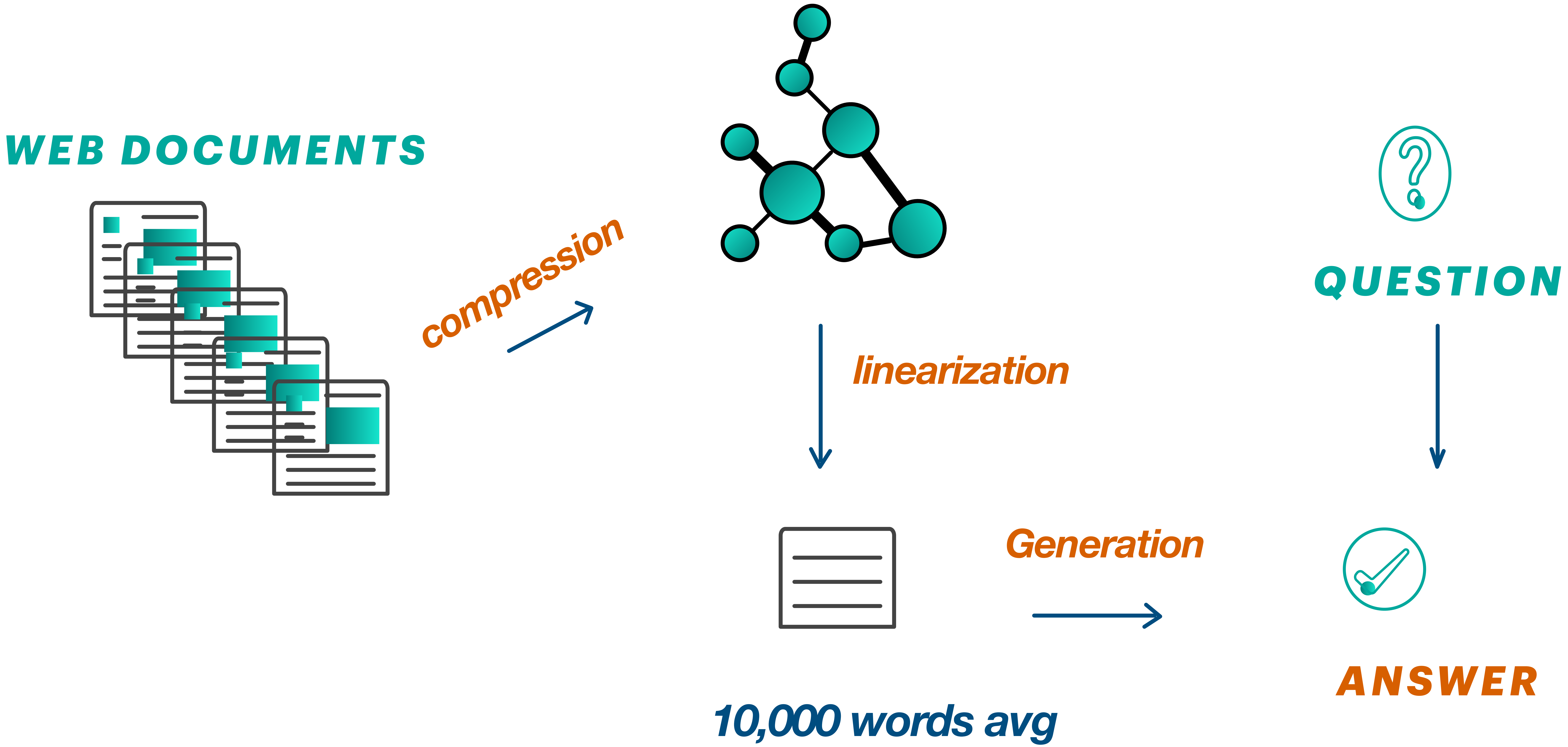
Information selected is Redundant

Web Input is Noisy, Selection is Hard

Knowledge Graph Construction

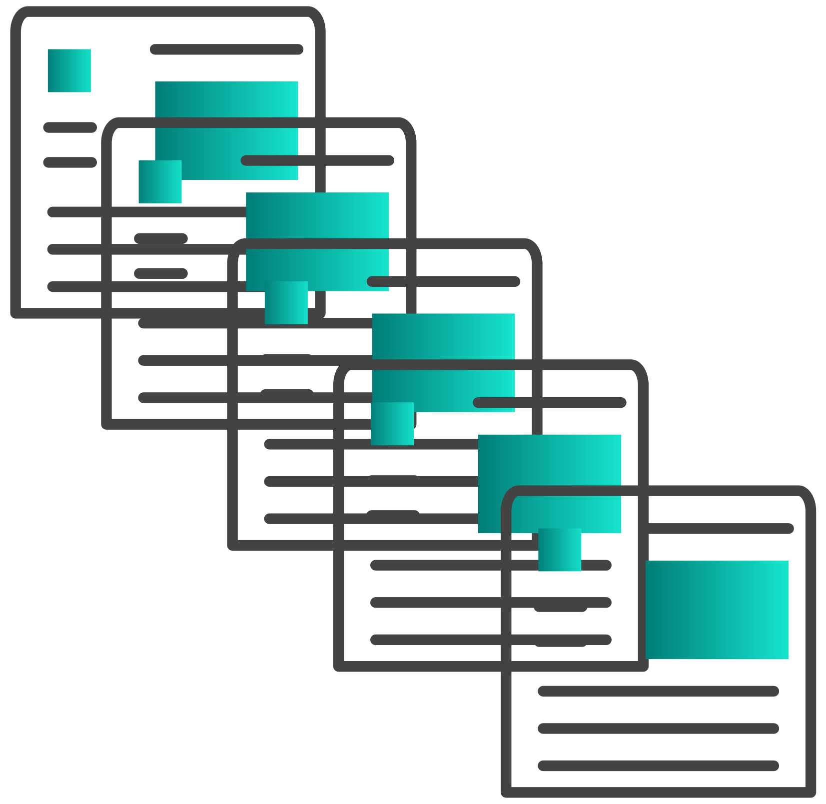


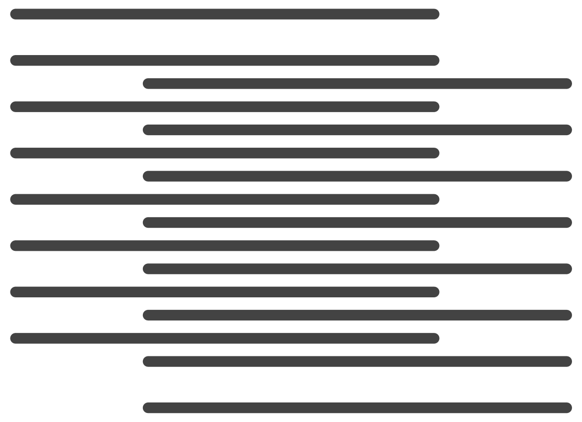
Knowledge Graph Construction



Converting a Text to a Graph

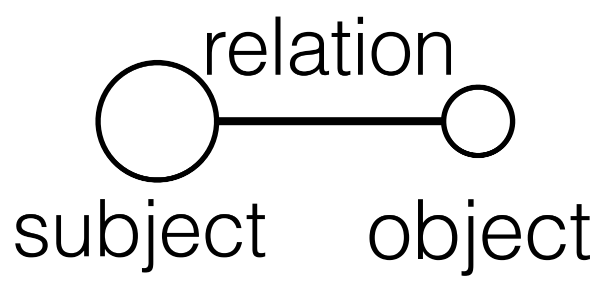
WEB DOCUMENTS





WEB DOCUMENT SENTENCES

open information extraction



coreference Resolution

Merge nodes
Increment
Nodes Weight

Tf-idf filtering

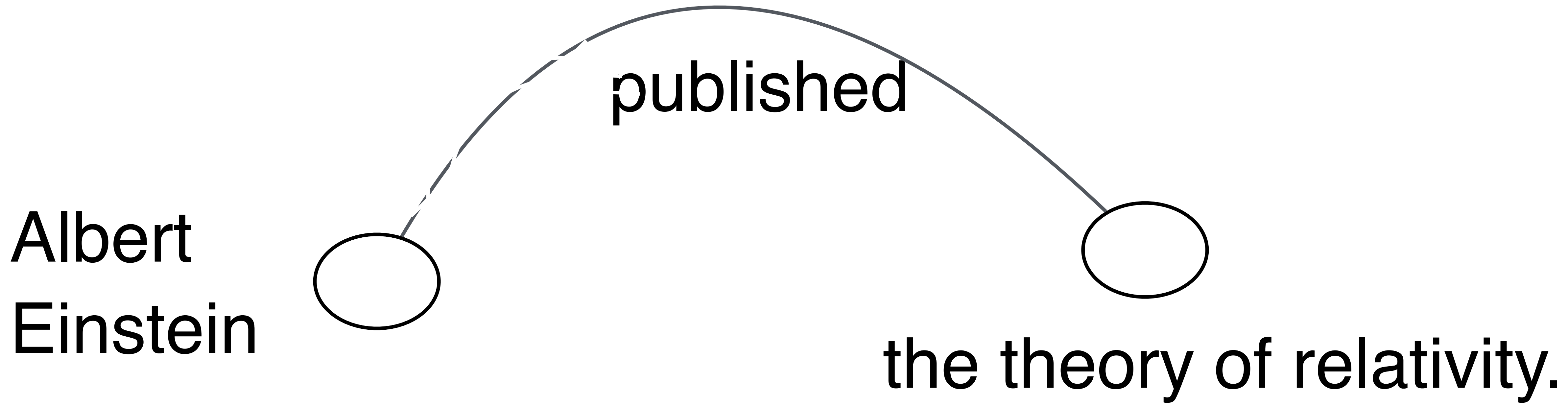
Filter Irrelevant
Input

Open Information Extraction

Converting text to edges

Can someone explain the theory of relativity ?

Albert Einstein, a German theoretical physicist , published the theory of relativity.



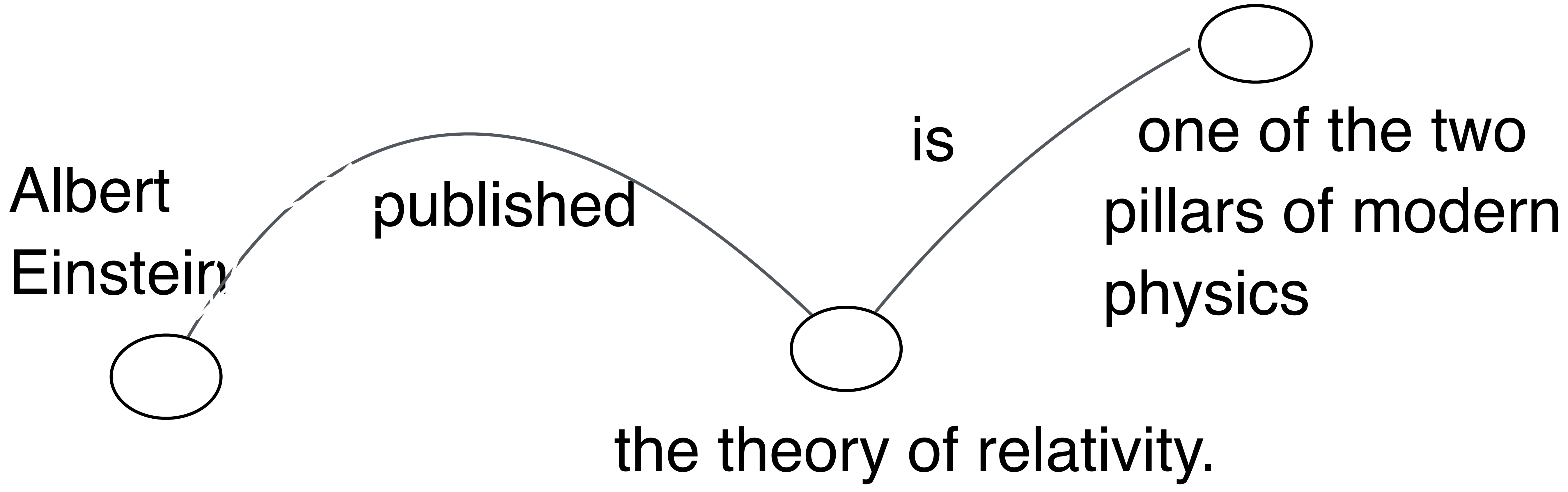
Coreference

Can someone explain the theory of relativity ?

Albert Einstein, a German theoretical physicist , published the theory of relativity.

The theory of relativity is one of the two pillars of modern physics

node weight +1

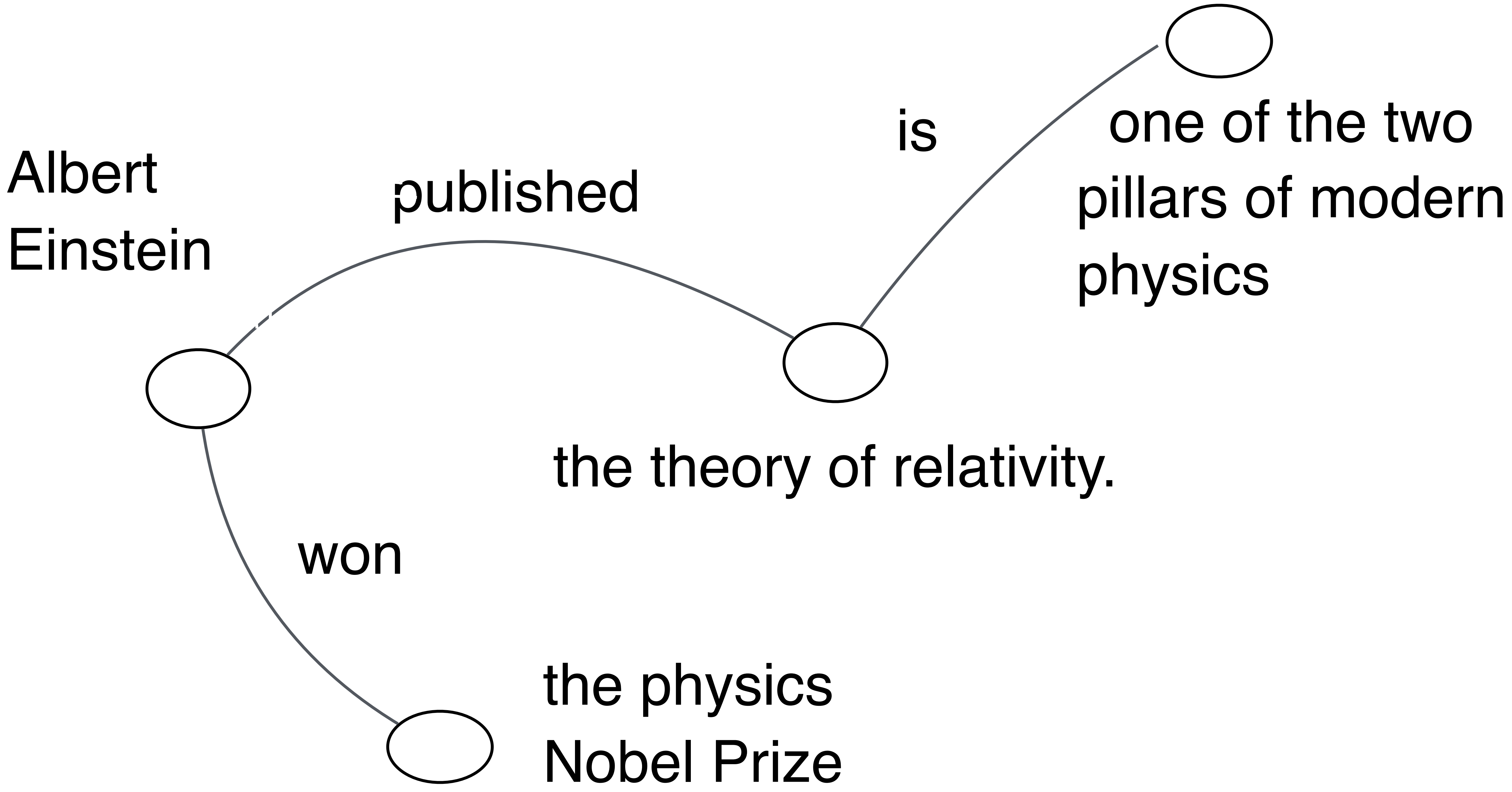


Coreference

Albert Einstein, a German theoretical physicist , published the theory of relativity.

The theory of relativity is one of the two pillars of modern physics.

He won the physics Nobel Prize
node weight +1



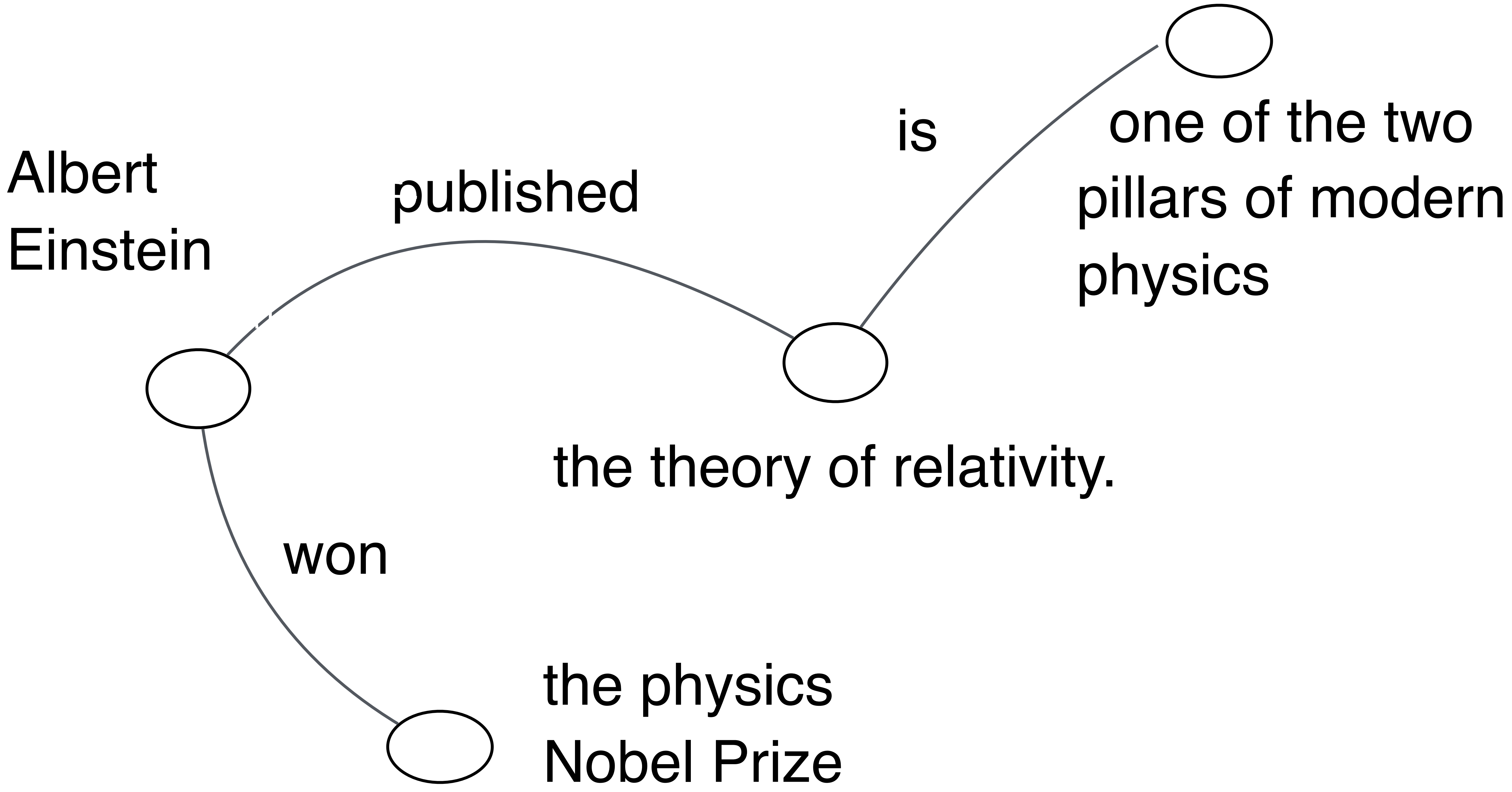
Relevance Filtering

Albert Einstein, a German theoretical physicist , published the theory of relativity.

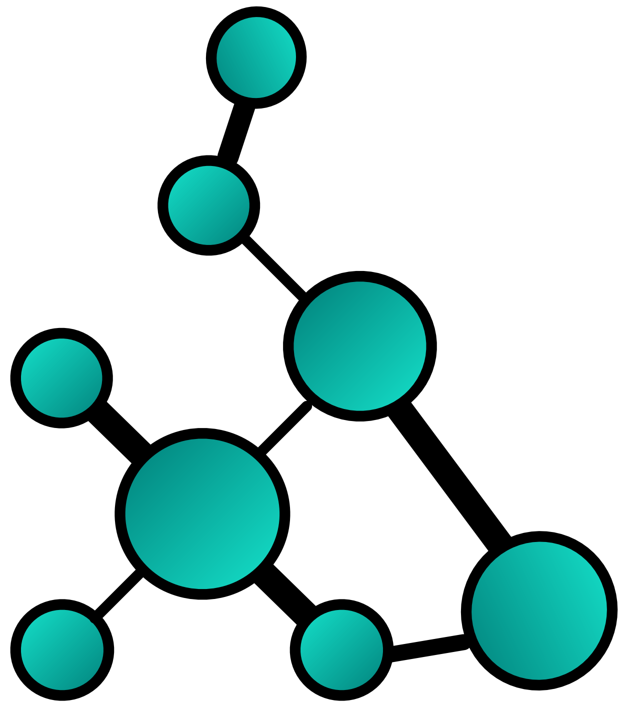
The theory of relativity is one of the two pillars of modern physics.

He won the physics Nobel Prize

Puppies are very cute.
Low TF-IDF overlap with query
Not added



Knowledge Graph Construction



Compresses the input by

- Merging redundant information
- Dropping words
- Filtering out irrelevant triples

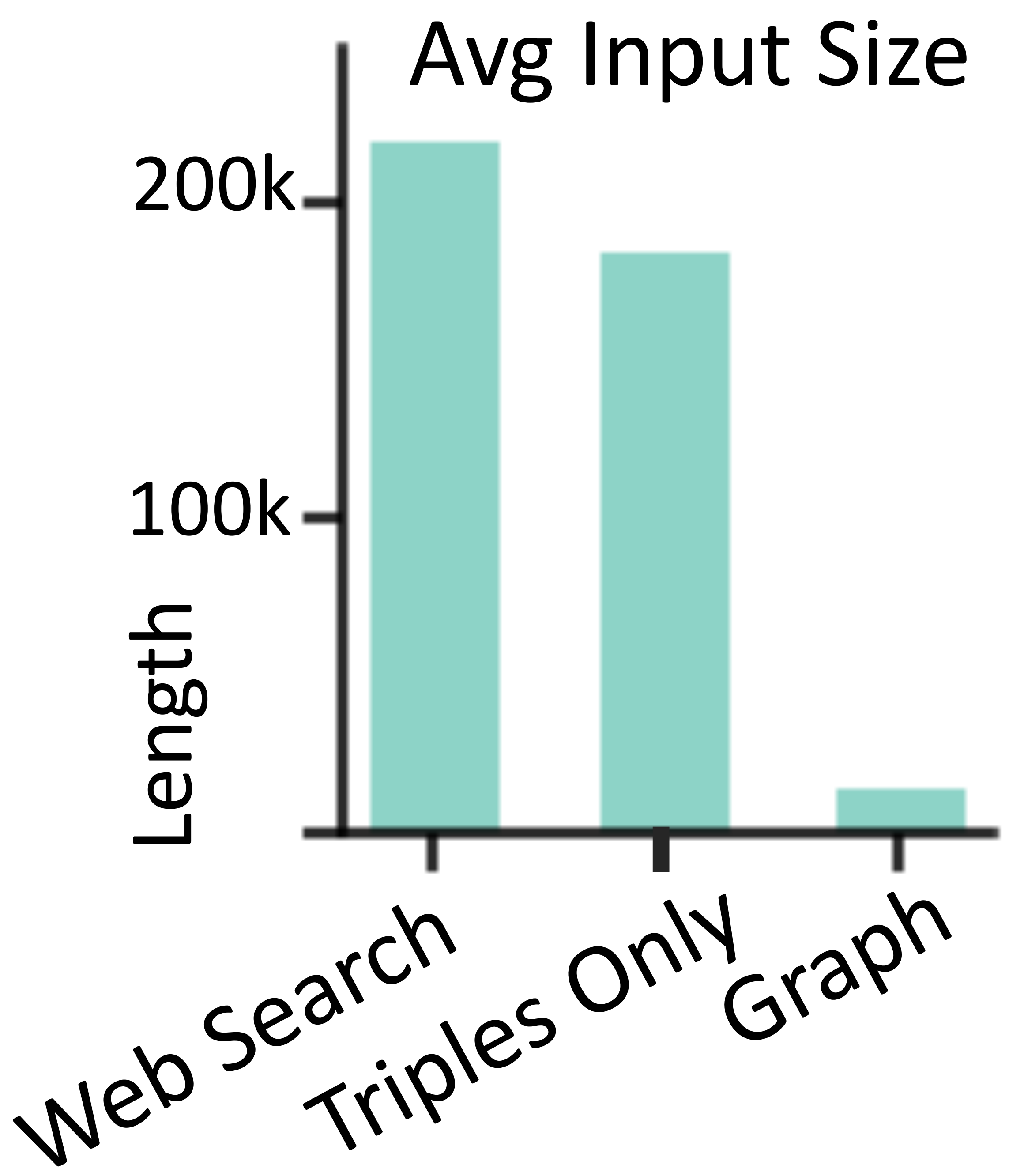
Reduces redundancy

- Merging nodes, edges and redundant triples

Filters out irrelevant content

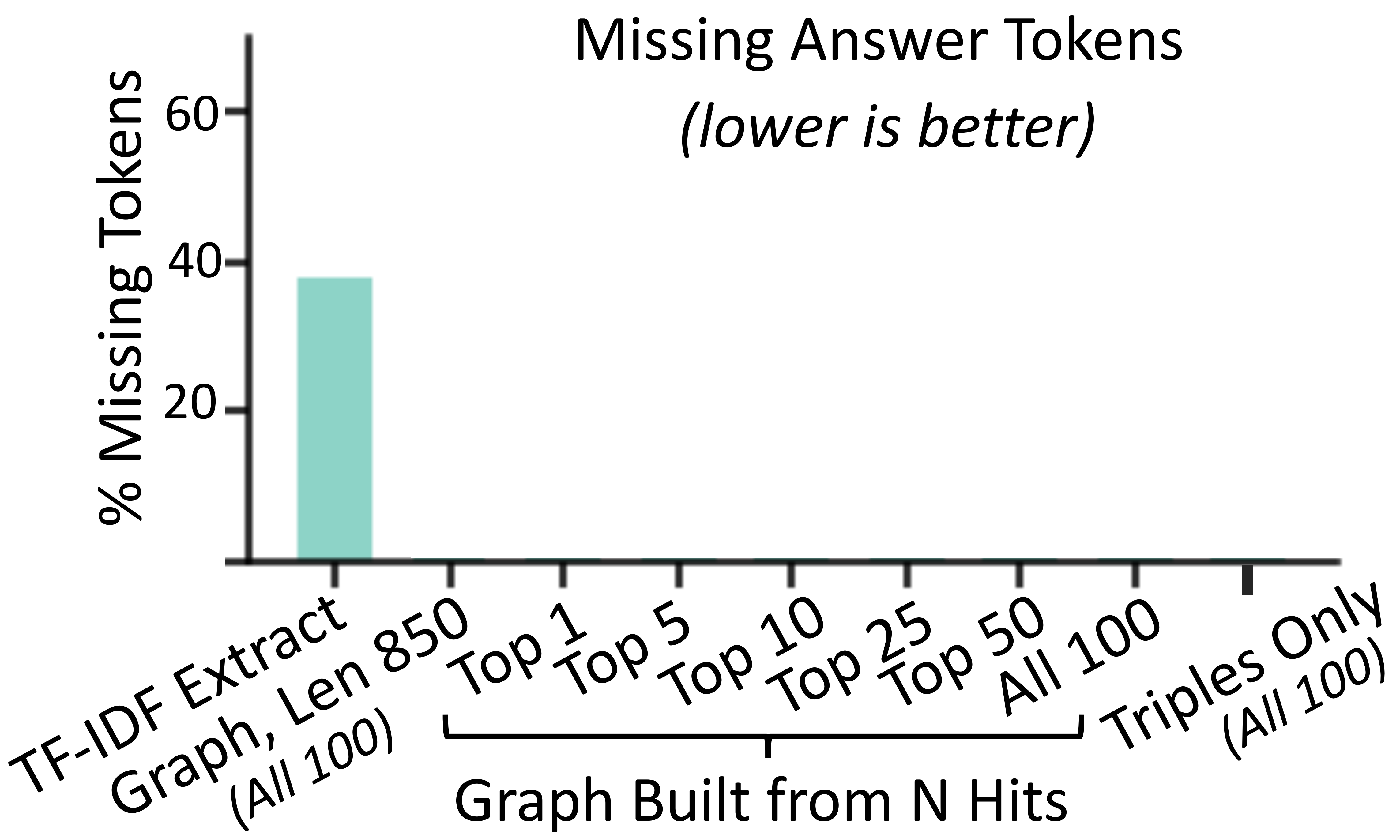
- Tf-idf overlap (Question,Triple)

How much does the graph manage to compress the input ?



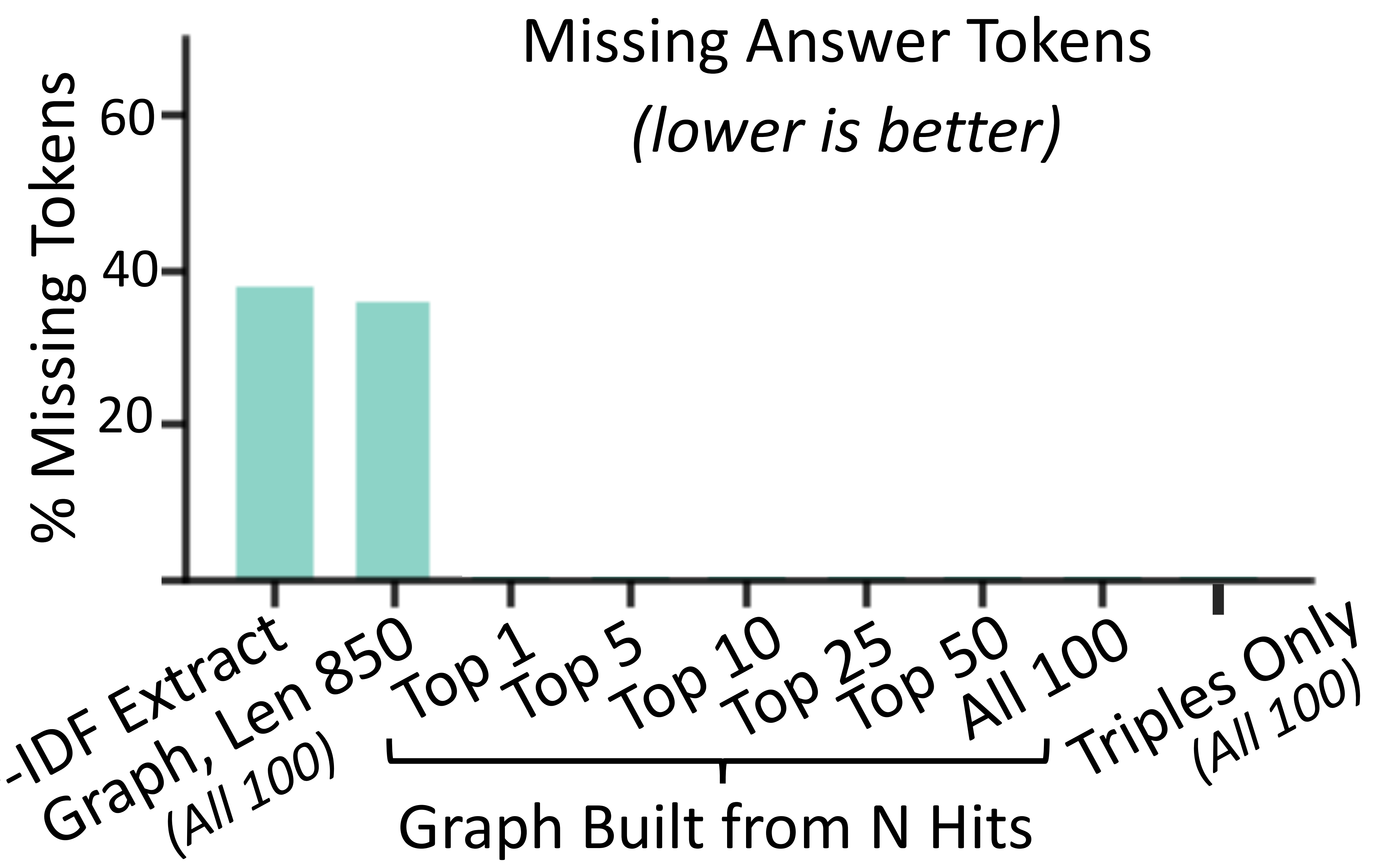
Knowledge Graph Construction drastically reduces the input size

How much does the graph preserve relevant information ?



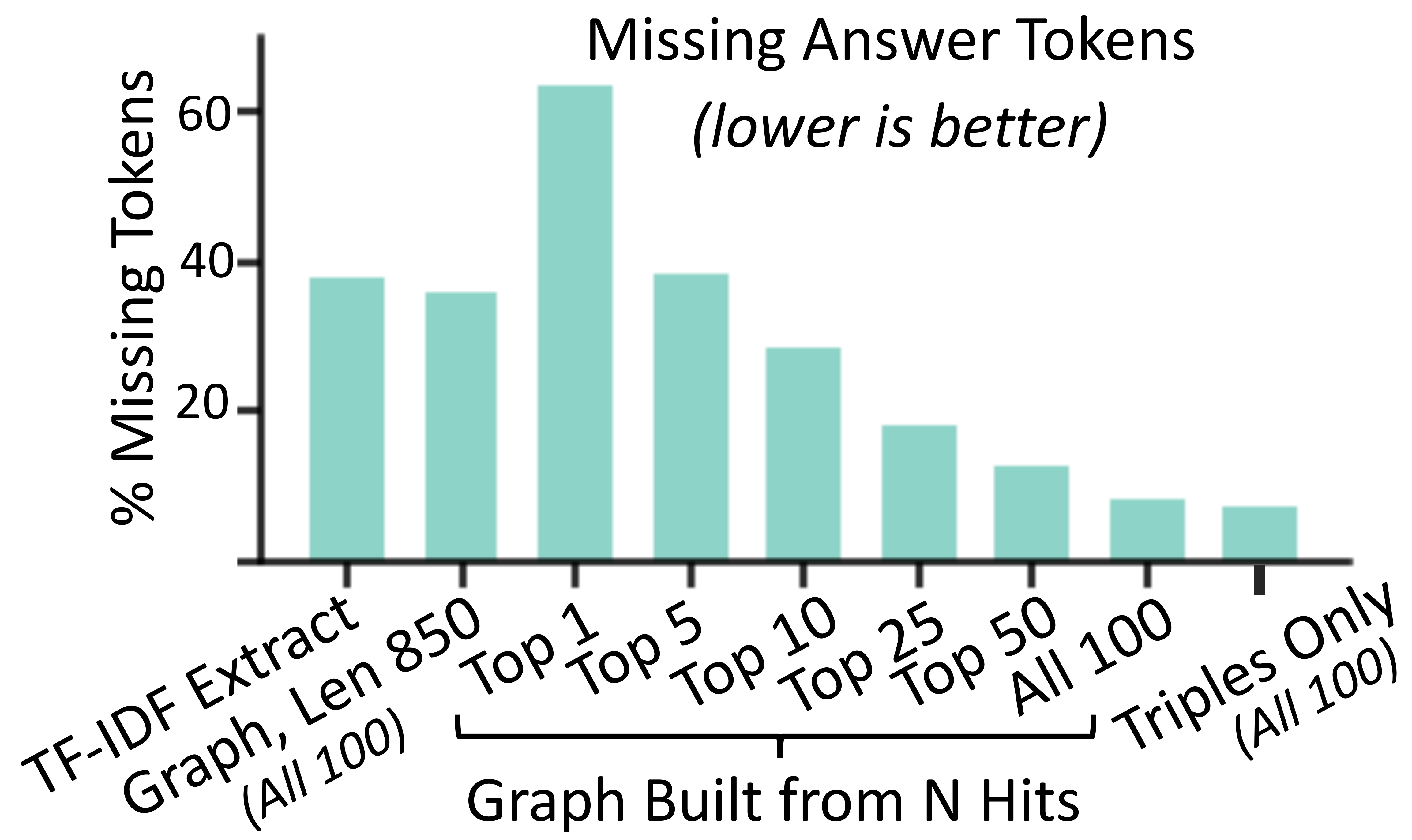
TF-IDF extraction is missing 38% of the answer tokens

Knowledge Graph Construction contains More Answer Tokens



The graph extracted for 850 tokens is missing 35% of the answer tokens

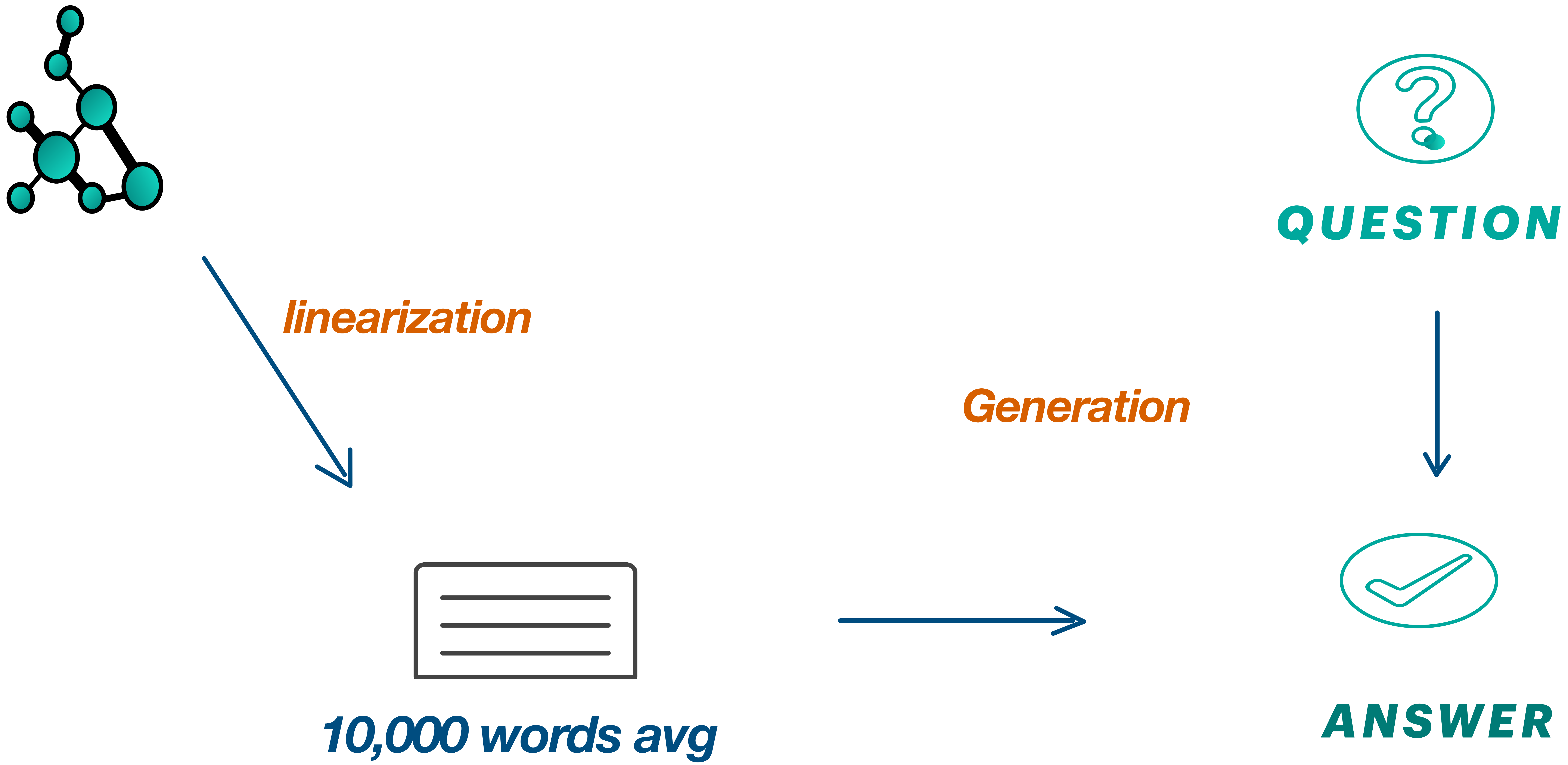
Knowledge Graph Construction contains More Answer Tokens



The graph for the full Input is missing only 8.7% of the answer tokens

Model

Generation Model



Encoding Graph Structure in a Seq2Seq Model

WORD EMBEDDING	<sub>	Albert	Einstein	<obj>	the	theory	of	relativity	<pred>	published	<s>	developed	<obj>	the	Physics	Nobel	Prize	<s>	won
POSITION EMBEDDING	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

Encoding Graph Structure in a Seq2Seq Model

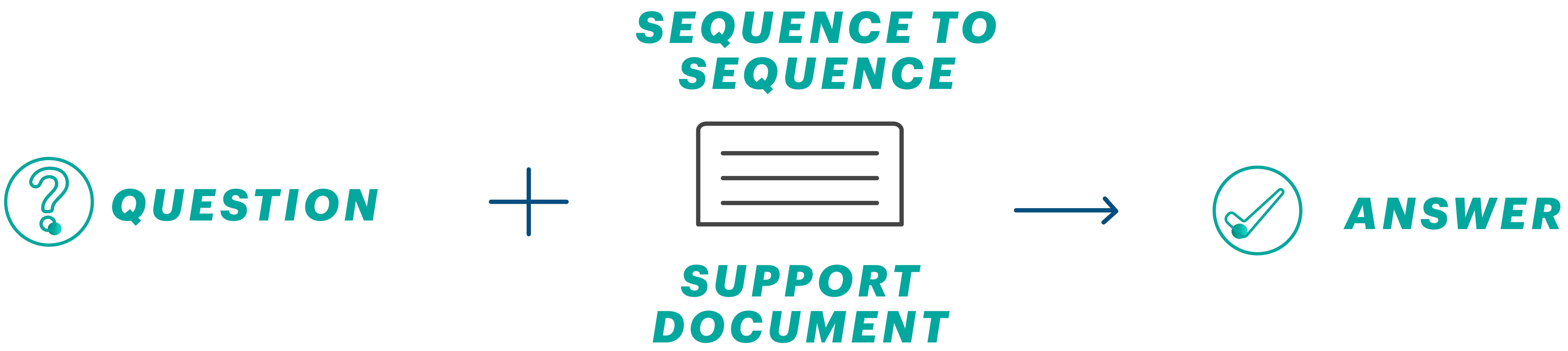
WORD EMBEDDING	<sub>	Albert	Einstein	<obj>	the	theory	of	relativity	<pred>	published	<s>	developed	<obj>	the	Physics	Nobel	Prize	<s>	won
POSITION EMBEDDING	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
GRAPH WEIGHT EMBEDDING	0	4	4	0	2	2	2	2	0	1	0	1	0	3	3	3	3	0	2

Encoding Graph Structure in a Seq2Seq Model

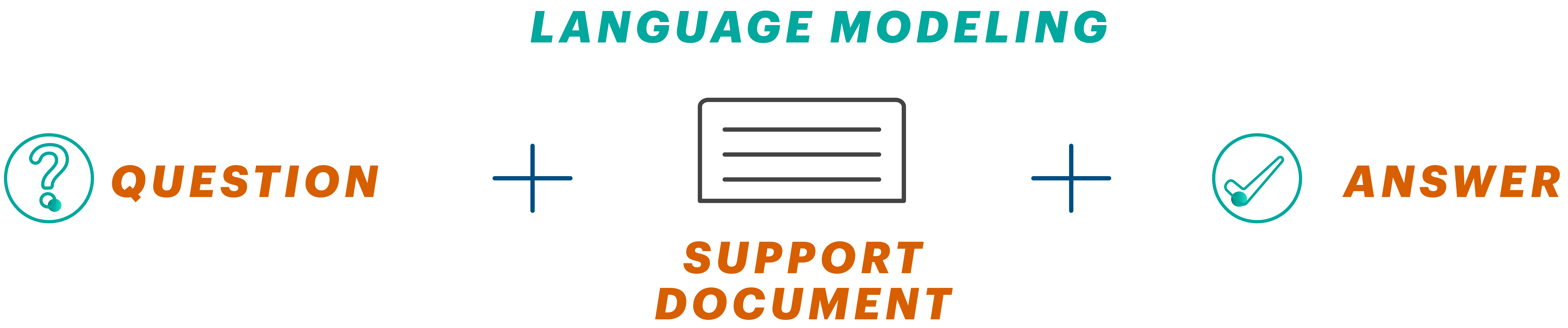
WORD EMBEDDING	<sub>	Albert	Einstein	<obj>	the	theory	of	relativity	<pred>	published	<s>	developed	<obj>	the	Physics	Nobel	Prize	<s>	won
POSITION EMBEDDING	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
GRAPH WEIGHT EMBEDDING	0	4	4	0	2	2	2	2	0	1	0	1	0	3	3	3	3	0	2
QUERY RELEVANCE EMBEDDING	0	1	1	0	1	1	1	1	0	1	0	2	0	1	1	1	1	0	1

Sequence-to-Sequence Model

Generate each word of the answer



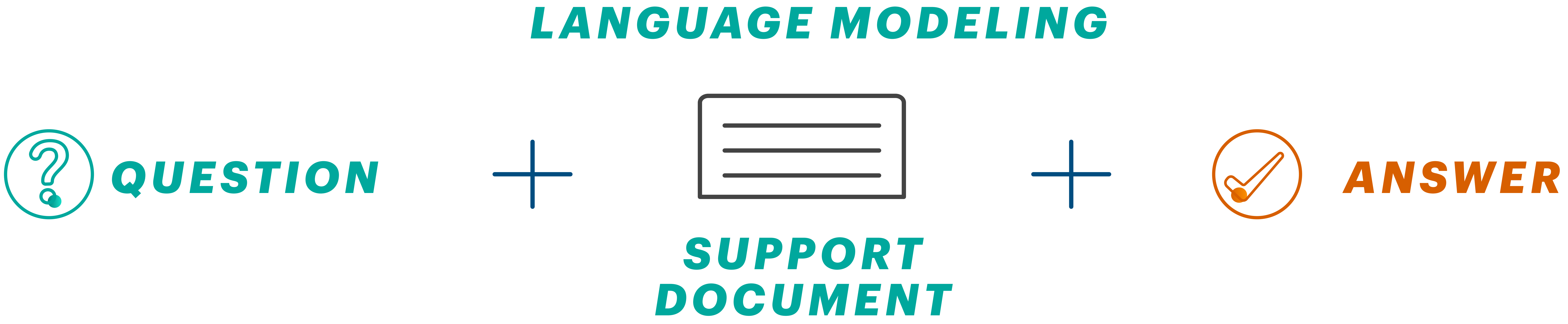
Language Model



Language Modeling Model

Inference time: provide true question and support document

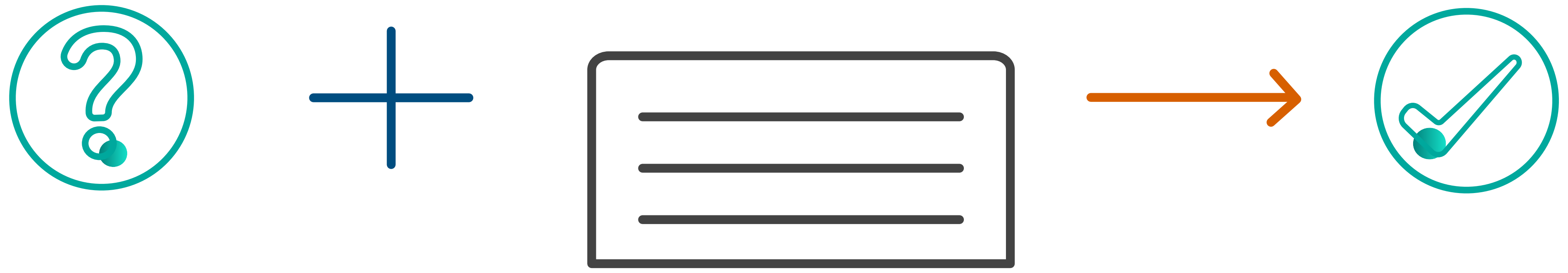
evaluate answer



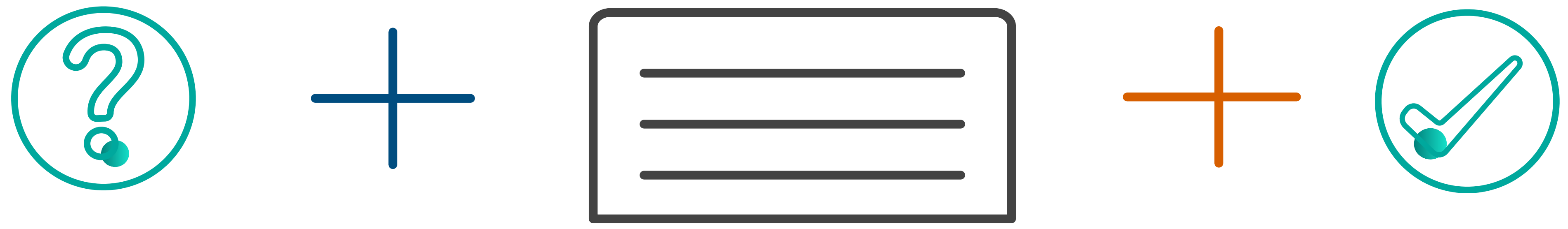
MULTITASK LEARNING

training time: train on many tasks

SEQUENCE TO SEQUENCE

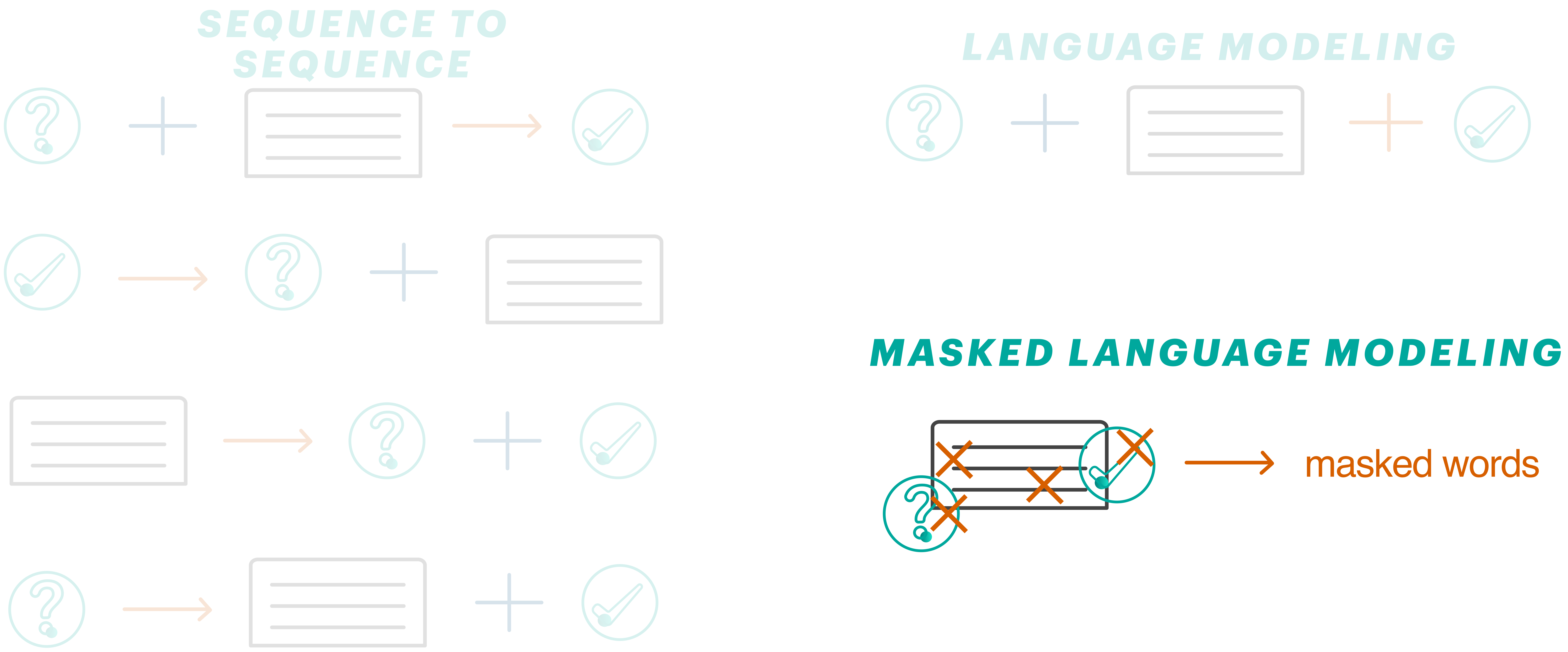


LANGUAGE MODELING

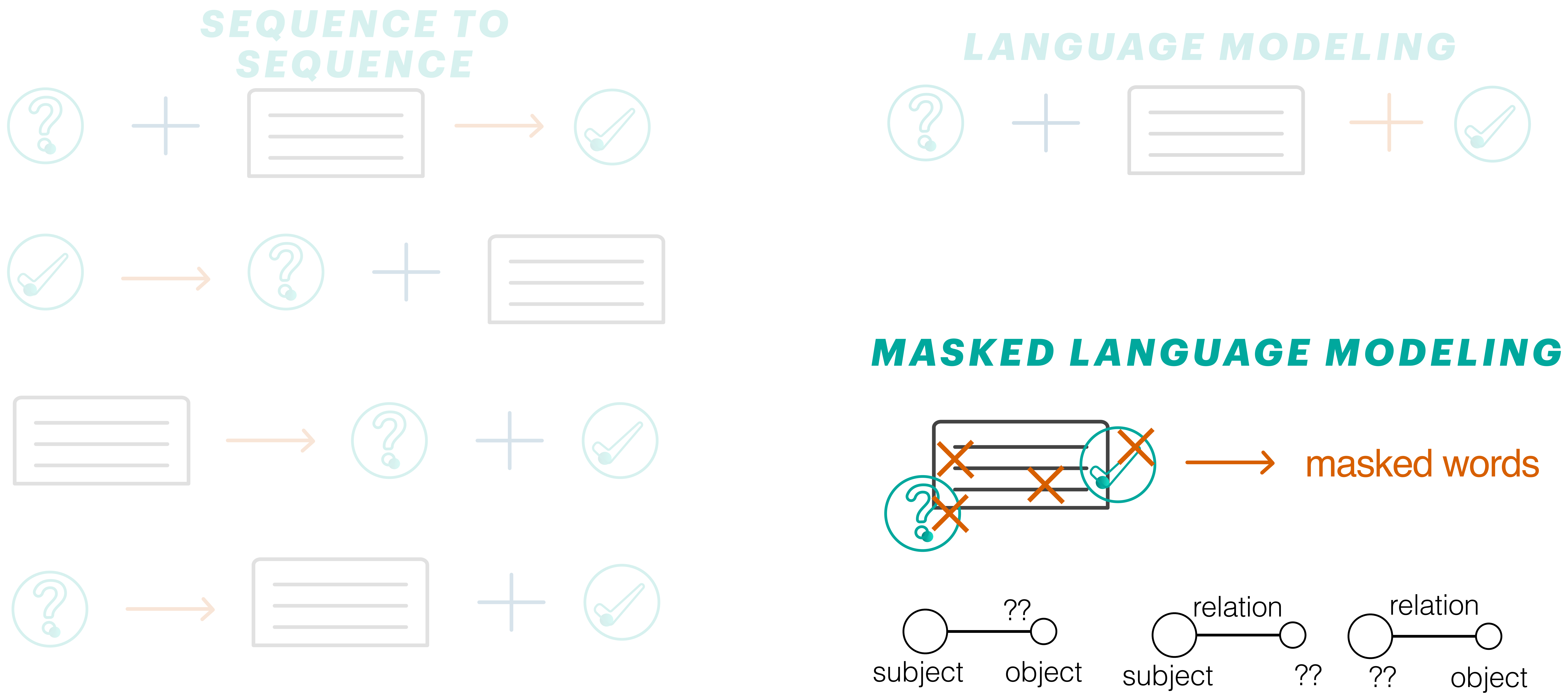


training time: train on many tasks

training time: train on many tasks

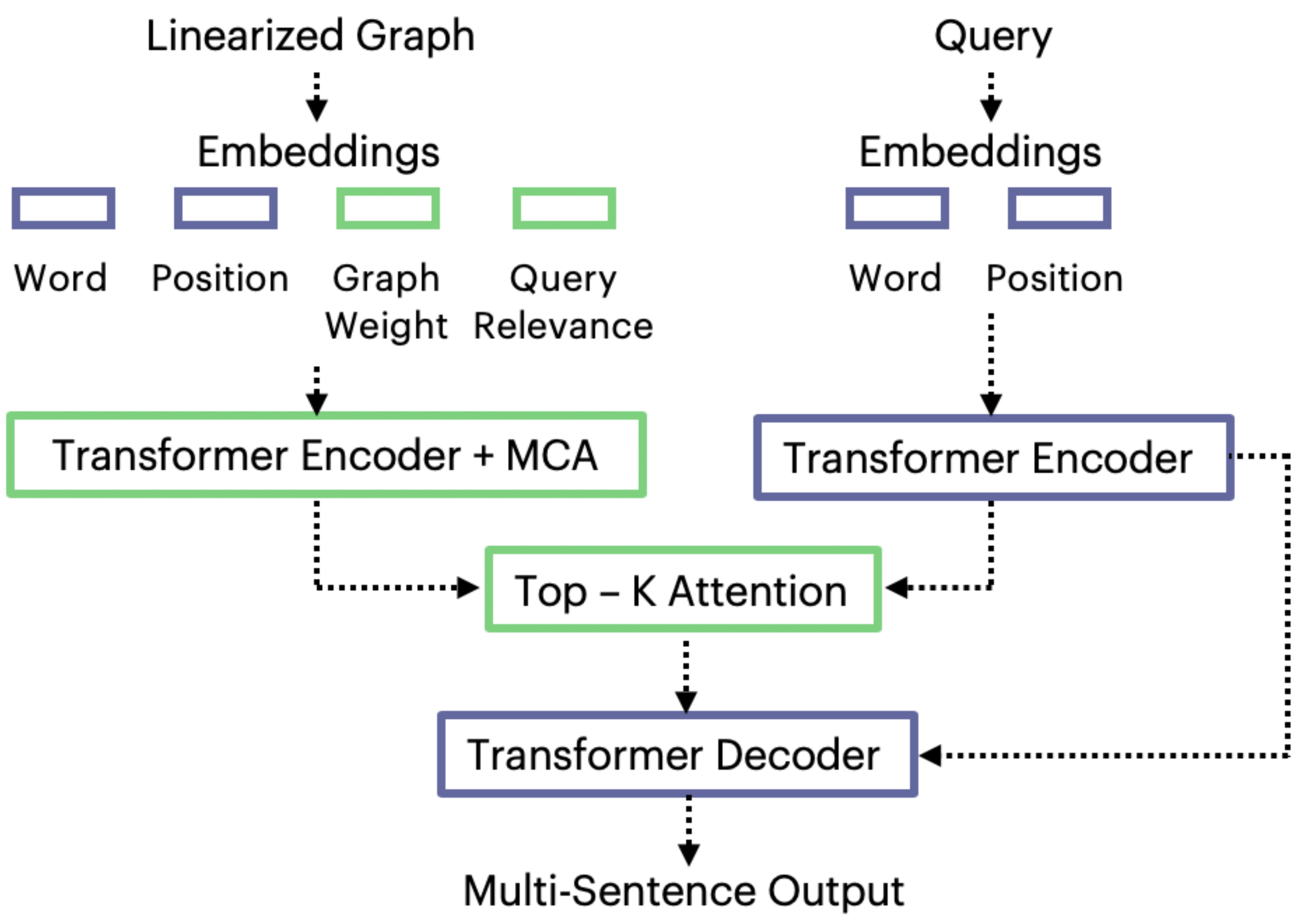


training time: train on many tasks



Handling Long Input

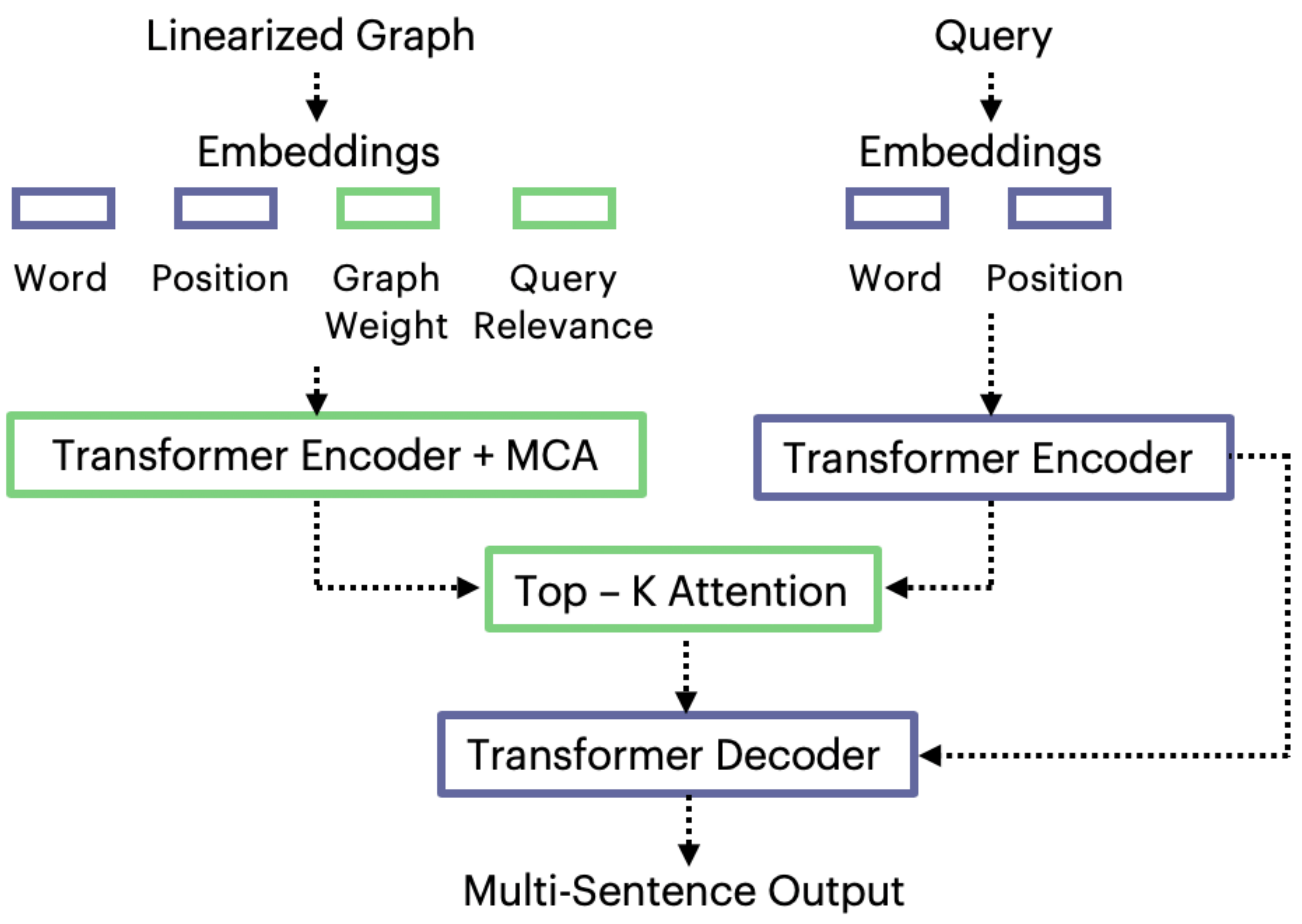
How do we encode 10K tokens in a Transformer ?



- MCA in Encoder

  Memory Compressed Attention

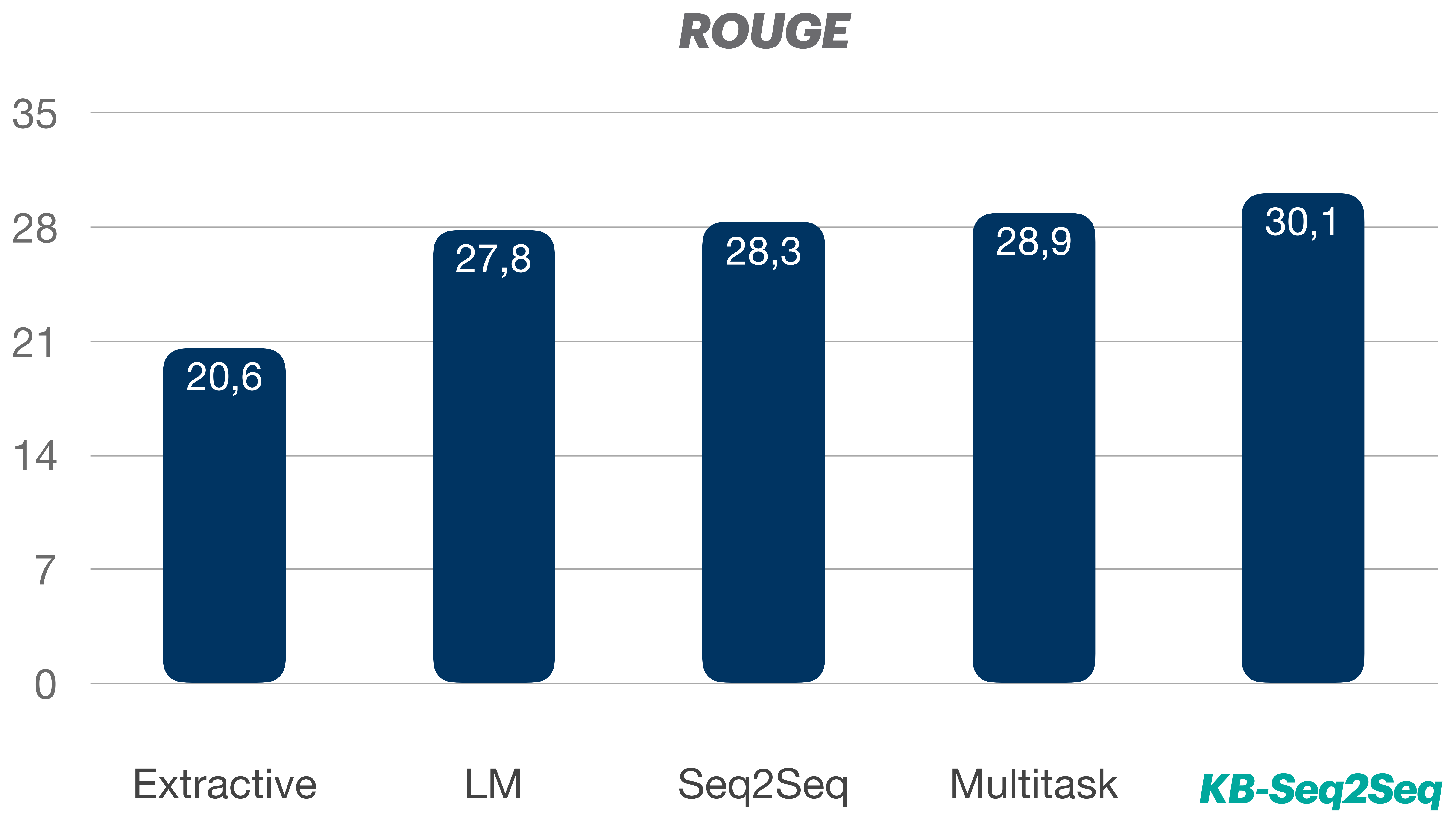
Handling Long Input



- MCA in Encoder
  Memory Compressed Attention
- Hierarchical Top-k Attention

Evaluation

Automatic Evaluation



Human Evaluation: Preference

Multi-task

KB-Seq2Seq

Generalising to other Datasets: WikiSum

Title: dwight h. perkins (economist)

Beginning of Web Search: memorial minute adopted by the faculty of arts and sciences, harvard university: abram bergson, john kenneth galbraith, carl kaysen, raymond vernon, dwight h. perkins, chairman. ed was a generous and understanding man who could see the good qualities in people and bring them out. he was always determined and persistent, which meant that he also saw to completion what he started . the list of projects, large and small, that he led is long. in 1946 he was one of the authors of the speech of secretary of state james byrnes in which the secretary announced the return of responsibility for the german economy to the germans. in 1956, he, together with ray vernon, did a pioneering study of the new york metropolitan region, a study that tried to identify the economic, political and social forces that were shaping that vast urban area. at the time the problems of urban areas were mainly the concern of architects and philosophers; almost no economist and only a few political scientists knew much or cared greatly about such issues. the studies that resulted provided rich fare for a generation of urban planners to follow. mason, edward s.(1899-1992) — harvard square library home biographies mason, edward s.(1899-1992) mason, edward s.(1899-1992) edward sagendorph mason was a member of the first parish unitarian church in harvard square who exemplified liberal religion in both thought and action. in addition to his notable contributions to the science of government, he served as chairman of the sloan commission on cable communication which issued recommendations for the future, on the cable. [...]

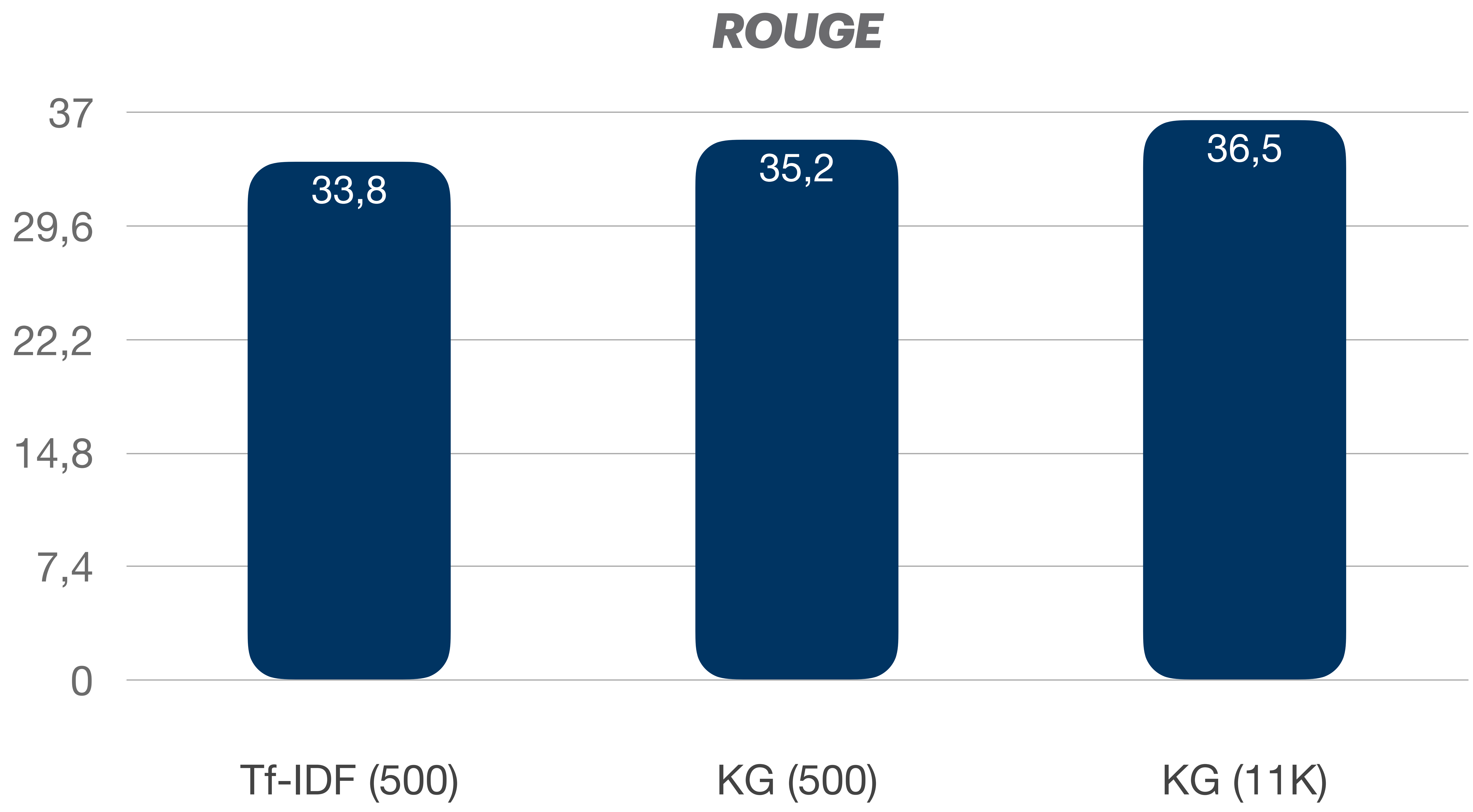
Target Lead Paragraph: dwight heald perkins (born in chicago, illinois in 1934) is an american academic, economist, sinologist and professor at harvard university. he is the son of lawrence bradford perkins, architect, and margery blair perkins and the grandson of dwight heald perkins, the architect, for whom he was named. he married julie rate perkins in 1957 and they have three adult children. == early life == perkins earned an undergraduate degree at cornell university in 1956. after two years military service in the us navy, perkins resumed his studies at harvard. he earned a ma in economics in 1961 and a ph.d. in economics in 1964. his doctoral thesis was "price formation in communist china". == academic career == perkins' teaching career at harvard began when he was still a graduate student, and continued uninterrupted through 2006 when he became a research and emeritus professor. he was a member of both the department of economics of the faculty of arts and sciences at harvard and of the harvard kennedy school. [...]

Title
= query for web search

Multi-document input
= Result from web search

Target
= Wikipedia Article lead paragraph

Automatic Evaluation



Example of Generated Text (ELI5)

Question: Why is touching microfiber towels such an uncomfortable feeling?

True Answer: Do you mean the kind of cloths used to clean glasses and lenses? I've never noticed any uncomfortable feeling myself, but I do find touching certain cleaning cloths can be quite uncomfortable. There's a brand called "e - cloth" which market themselves as not needing any cleaning supplies. Basically the threads have tiny little hooks and barbs in them, so dragging them over skin makes you feel like it 's catching against you.

Generation: The microfiber is made up of a bunch of tiny fibers that are attached to them. When you touch them, the fibers that make up the microfiber are attracted to each other. When they touch, they're attracted to the other end of the fiber, which is what makes them feel uncomfortable.

Fluent answer, factually correct

Some repetition ("attracted to each other")

Example of Generated Text (ELI5)

Question: Why do we have nightmares when stressed out?

True Answer: I can tell you that your mind, and in particular your dreams, are subject to the input that it the journey of your life, as you have interpreted it. Yours dreams tend to be your subconscious representing issues you have within, whether it be past or present, with things and people that are sort of "preinterpreted symbology" from your deep and ever power subconsciousness. All your nightmares are just dreams, and your inner mind knows exactly can mess with you the most, and sometimes it brings that to light, it usually being an issue you may be stressed about.

Generation: When you're stressed, your brain releases a hormone called cortisol. This hormone is released by your adrenal glands, and is released when you are stressed. This is why it's called "fight-or-flight" mode. When your body is stressed, it releases hormones that make you feel more stressed. The reason why you don't feel stressed is because your brain is trying to get rid of the stress hormone, which causes you to feel stressed, and you're less likely to get stressed. If you're really stressed, you're probably going to be more stressed than if you weren't stressed.

Repetitions ("when you are stressed")
Non sense fragments ("The reason ...get stressed")

Example of Generated Text (WikiSum)

Title: wylie rogers

True Paragraph: wylie rogers (born march 16, 1985) is an american ice hockey goaltender. he played with the rio grande valley killer bees of the central hockey league. prior to turning professional, rogers attended the university of alaska fairbanks where he played four seasons of ncaa men's division i ice hockey with the alaska nanooks men's ice hockey team. rogers played the 2010-11 season in the netherlands with the heerenveen flyers.

Generation: wylie rogers is an american football coach and former player. he is currently the head football coach at the university of north carolina at chapel hill. rogers played college football at north carolina state university. he was a member of the north carolina tar heels men's football team that won the ncaa division i football championship in 2012. he also played for the university's men's basketball team. after his playing career ended, rogers was hired as the head coach at unc chapel hill in 2013. he led the tar heels to their first ncaa division ii championship in 2015.

Fluent text with nice, elaborate discourse structure

Incorrect facts (Wylie Rogers is an ice hockey player not a football coach. University of Alaska not North Carolina. etc;)

Retrieving and Integrating Relevant Knowledge

Human-Machine Dialog

Retrieving Knowledge to generate good Responses

Human: I love watching Disney movies! Do you like Disney ?

Model: I love Disney, I love watching Disney movies and different animations and characters that makes them unique

Human: What's your favorite Disney movie ?

Model: I love the incredibles, they are my favorite Disney movie.

Human: omg, me too! The second one just came out. I don't watch a lot of sequels though, I think the last one I saw was Toy Story

Model: That's cool. I remember when Toy Story was first released in 1995

K-Nearest Neighbour Retrieval

Two retrieval sources

Human: What’s your favorite Disney movie ?

Model: I love the incredibles, they are my favorite Disney movie.

Fetched Knowledge: Disney announced intentions to develop additional superhero films after the success by the incredibles

Fetched Template: I love kiteboarding, it is one of my favorite activities on the water.

Wizard of Wikipedia

- Dialog about a topic

- Retrieval Corpus for KL
  - WKP passages
  - 34 per topic

- Retrieval Corpus for Template
  - Dialog turns
  - 170K

Image Chat

- Dialog about an image

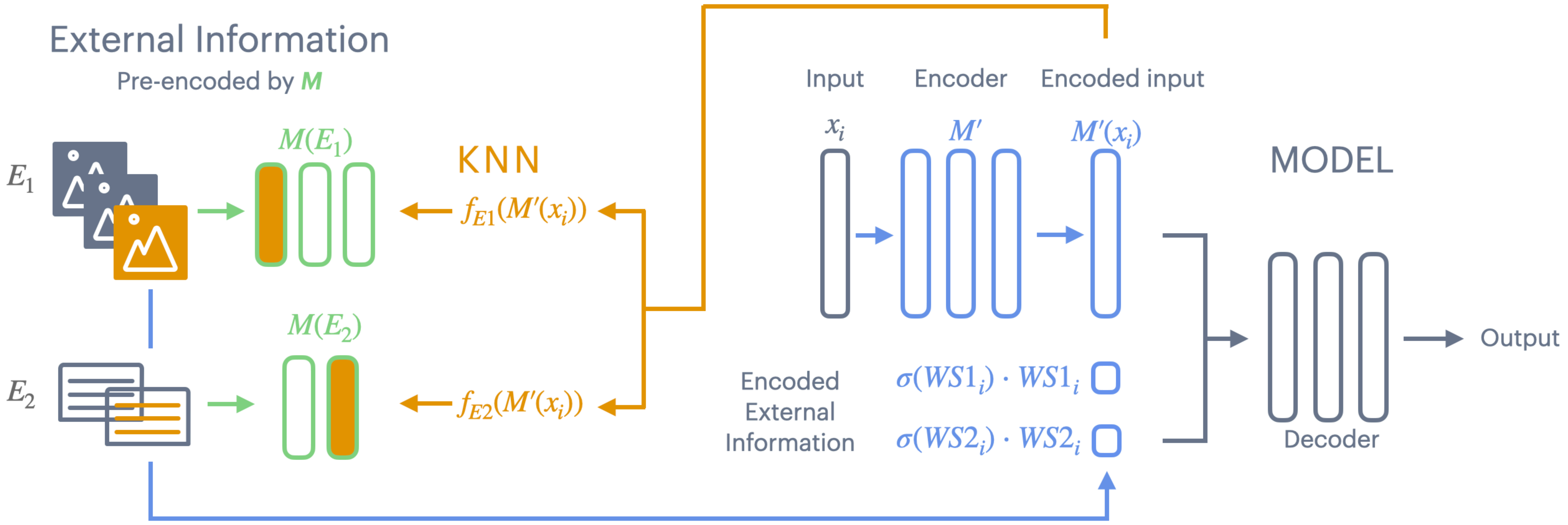
- Retrieval Corpus for KL
  - image + dialog
  - 184K images

- Retrieval Corpus for Template
  - Dialog turns
  - 350K dialog turns

Extending Human-Machine Dialog with External Retrieval

K-Nearest Neighbour Search

Two sources of information



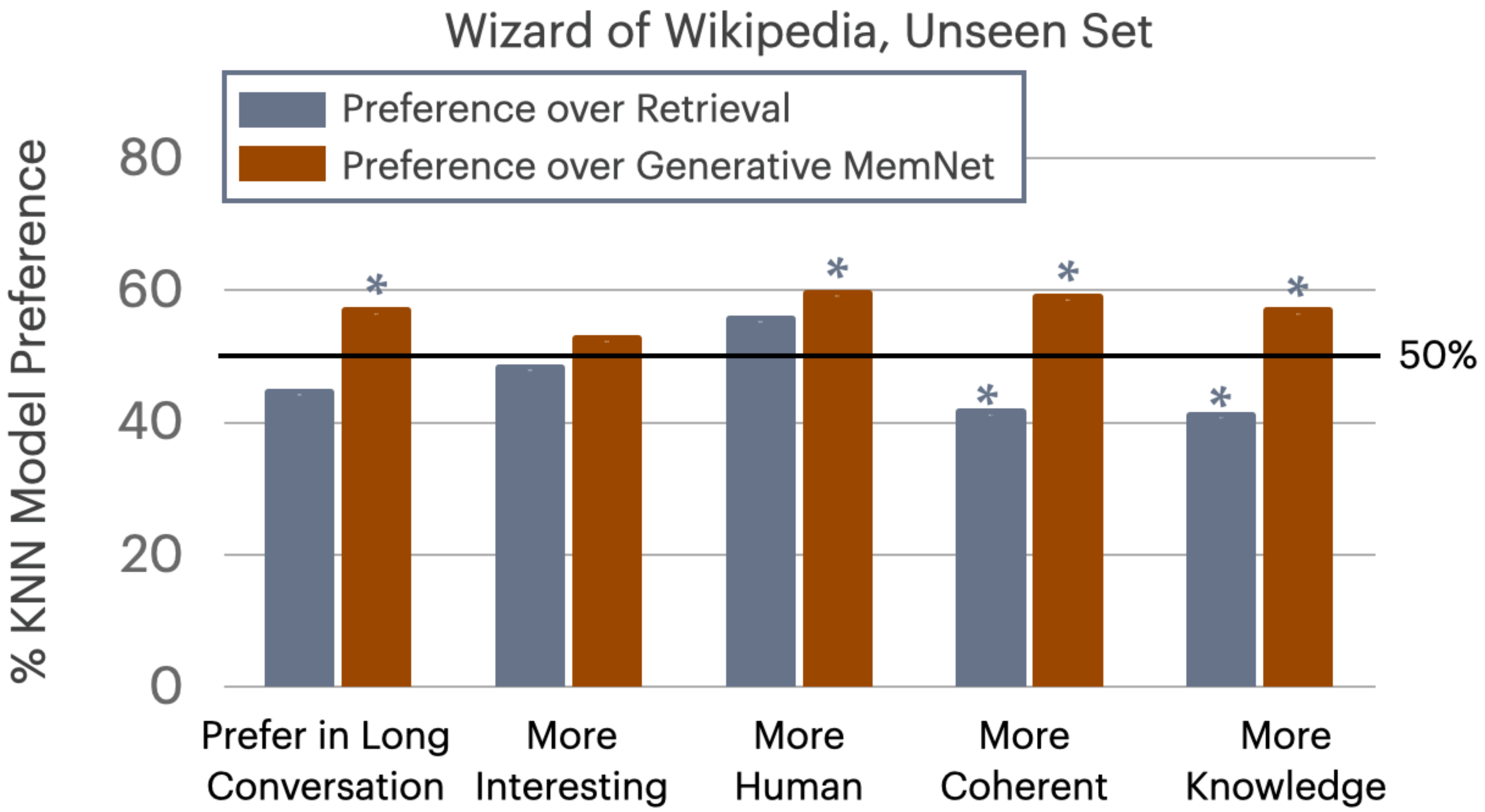
Effect of Fetched Text on Generation

Keeping the template fixed

Keeping the KL fixed

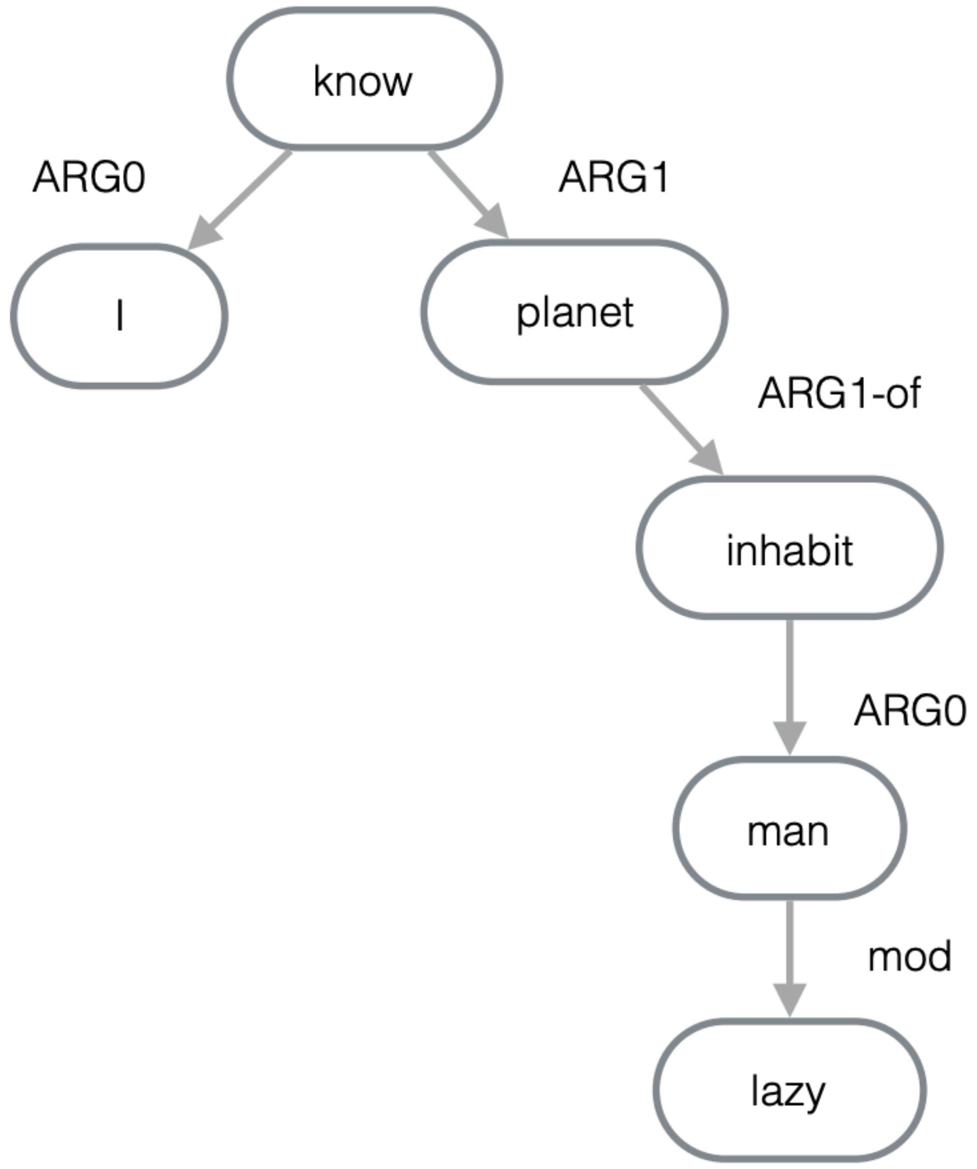
Knowledge	Training Utterance	Generation
buzz lightyear's name is in honor of astronaut edwin 'buzz' aldrin	my favorite character in that book series is hermione granger	cool! my favorite character in that movie is **buzz lightyear**
mr potato head is based on the **real-life mr. potato head** toy	my favorite character in that book series is hermione granger	my favorite character in that movie is **real-life mr potato head**
slinky dog is a toy dachschund with a metal slinky for a body	my favorite character in that book series is hermione granger	cool! my favorite character is the **slinky dog**
slinky dog is a toy dachschund with a metal slinky for a body	**i really like** the character hermione granger	cool! **i really like** slinky dog
slinky dog is a toy dachschund with a metal slinky for a body	**my favorite character** of all time **has to be** hermione granger	i love that movie, **my favorite character has to be** slinky dog the dachshund
slinky dog is a toy dachschund with a metal slinky for a body	i agree with you! that's **my favorite** character as well	i think so too! **my favorite** is slinky

Extending Human-Machine Dialog with External Retrieval



Generating from AMR Graphs into Multiple Languages

AMR Graph



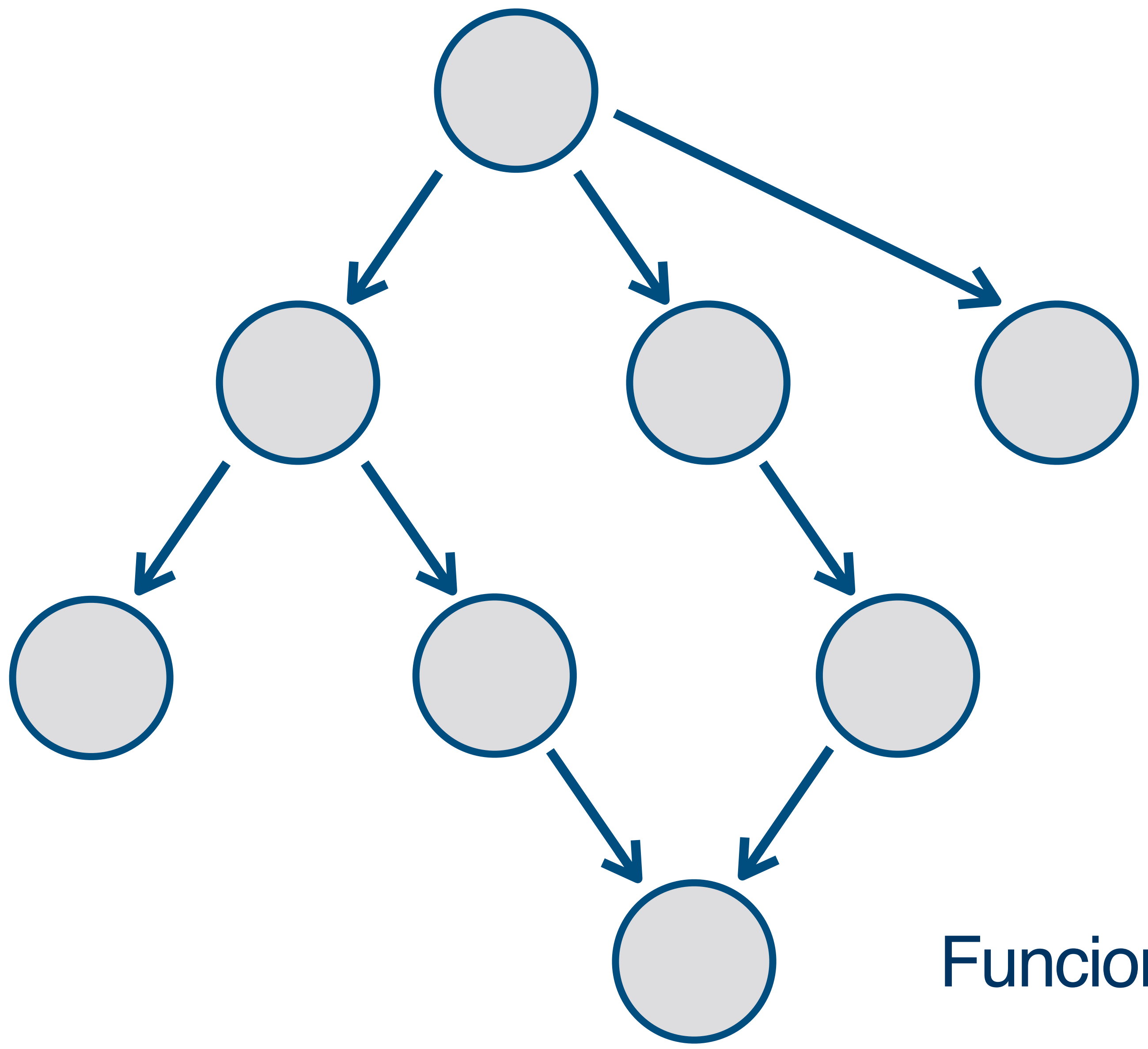
I have known a planet that was inhabited by a lazy man

- Rooted Directed Acyclic Graph
- Nodes: concepts (nouns, verbs, NE, etc.)
- Edges: Semantic Roles

Graphs are frequent data structures

- Knowledge Graphs
- RDF stores
- Tabular data
- Meaning Representations

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 Yor

Romance, Germanic, Slavic, Uralic

Challenges

- Structured Input has a different surface form

Challenges

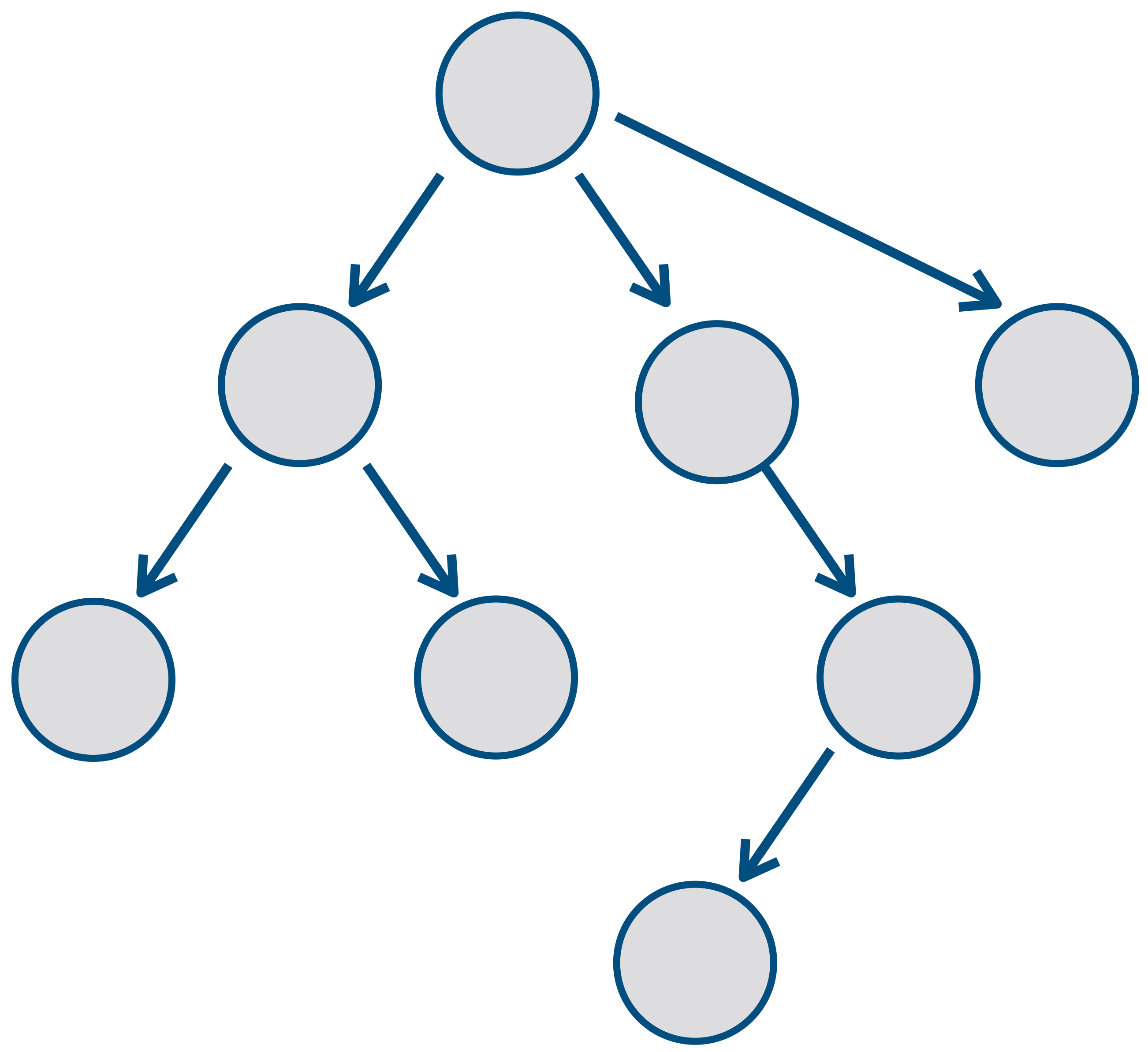
- Structured Input has a different surface form
- Structured Input is underspecified

Challenges

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

Encoder-Decoder MODEL

Graph Encoding



hold

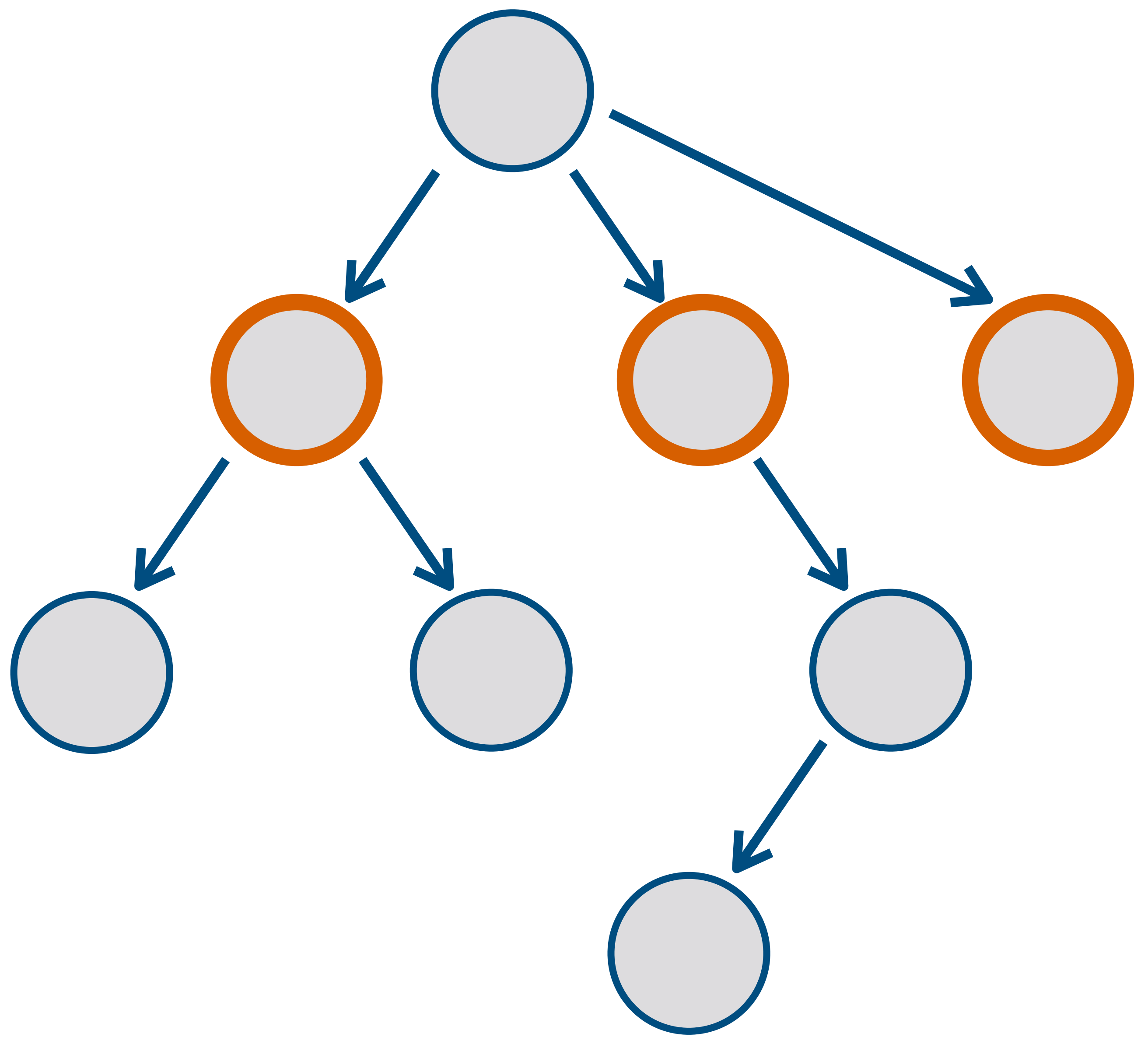
:ARG0 person : ARG0-of have-org-role :ARG1 :op1 United :op2 States :ARG2 official

:ARG1 meet :ARG0 person :ARG1-of expert :ARG2-of group

:time date-entity :year 2002 :month 1

:location city :op1 New :op2 York

Graph Encoding



hold

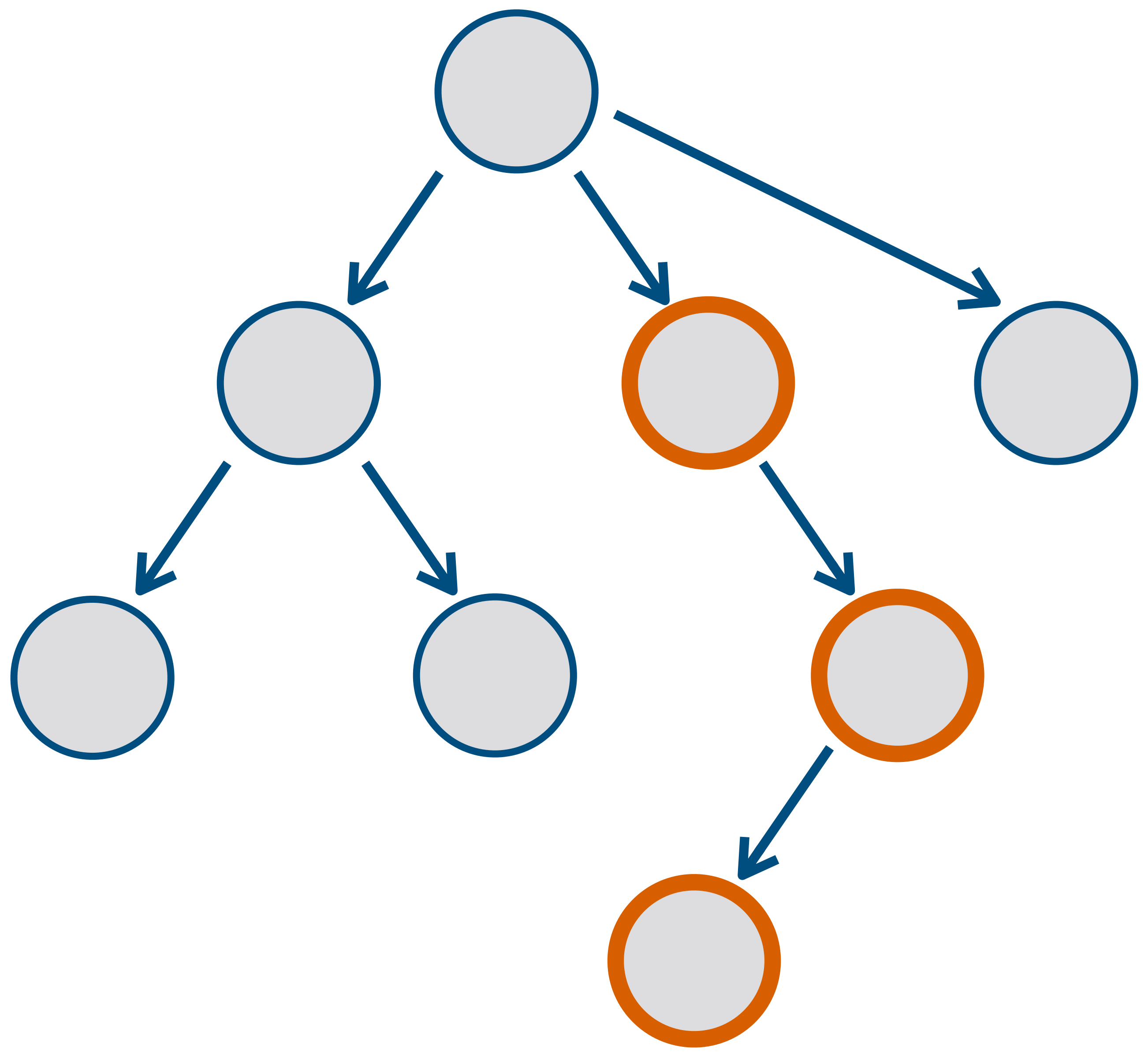
:ARG0 person : ARG0-of have-org-role :ARG1 :op1 United :op2 States :ARG2 official

:ARG1 meet :ARG0 person :ARG1-of expert :ARG2-of group

:time date-entity :year 2002 :month 1

:location city :op1 New :op2 York

Graph Encoding



hold

:ARG0 person : ARG0-of have-org-role :ARG1 :op1 United :op2 States :ARG2 official

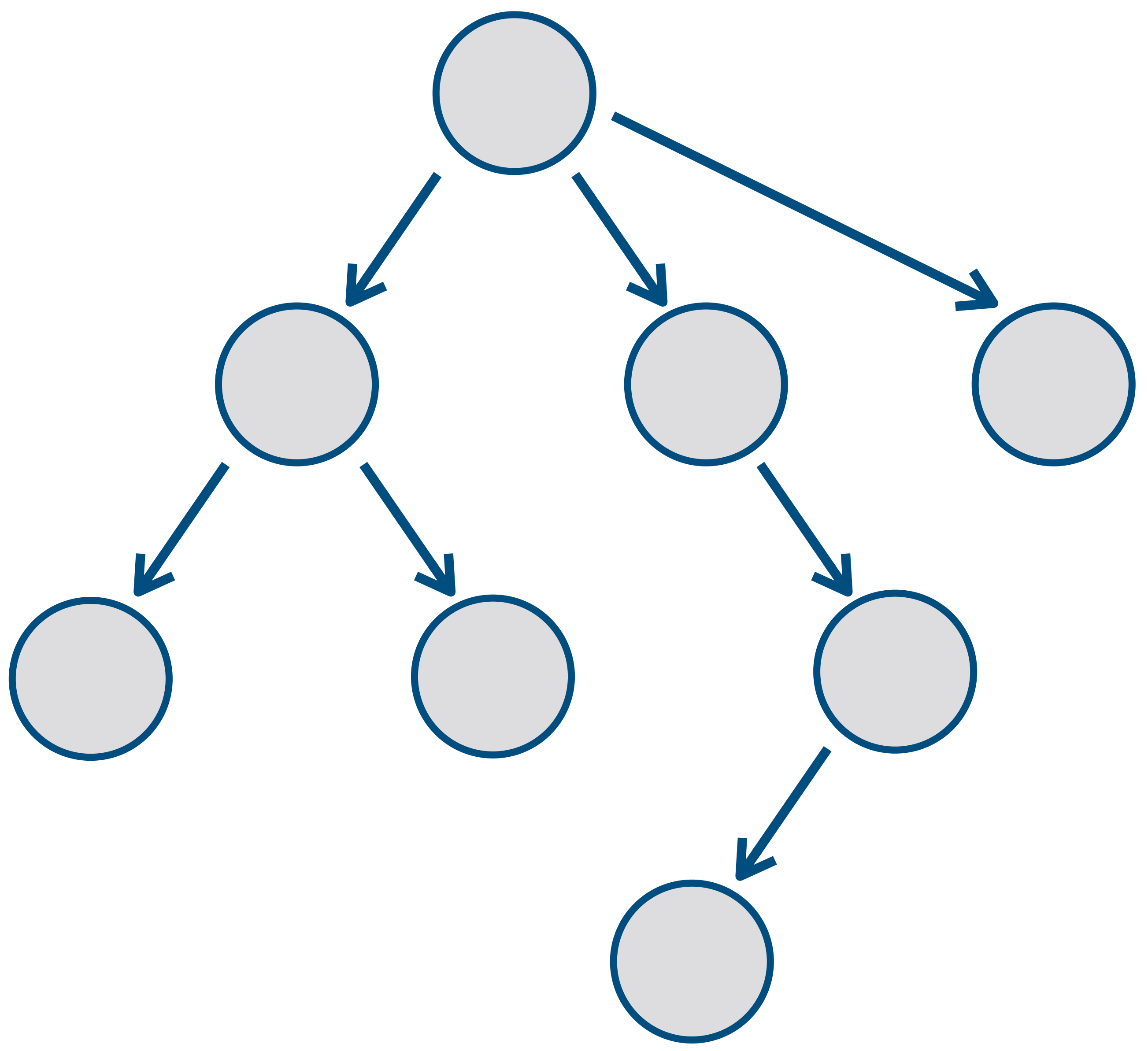
:ARG1 meet :ARG0 person :ARG1-of expert :ARG2-of group

:time date-entity :year 2002 :month 1

:location city :op1 New :op2 York

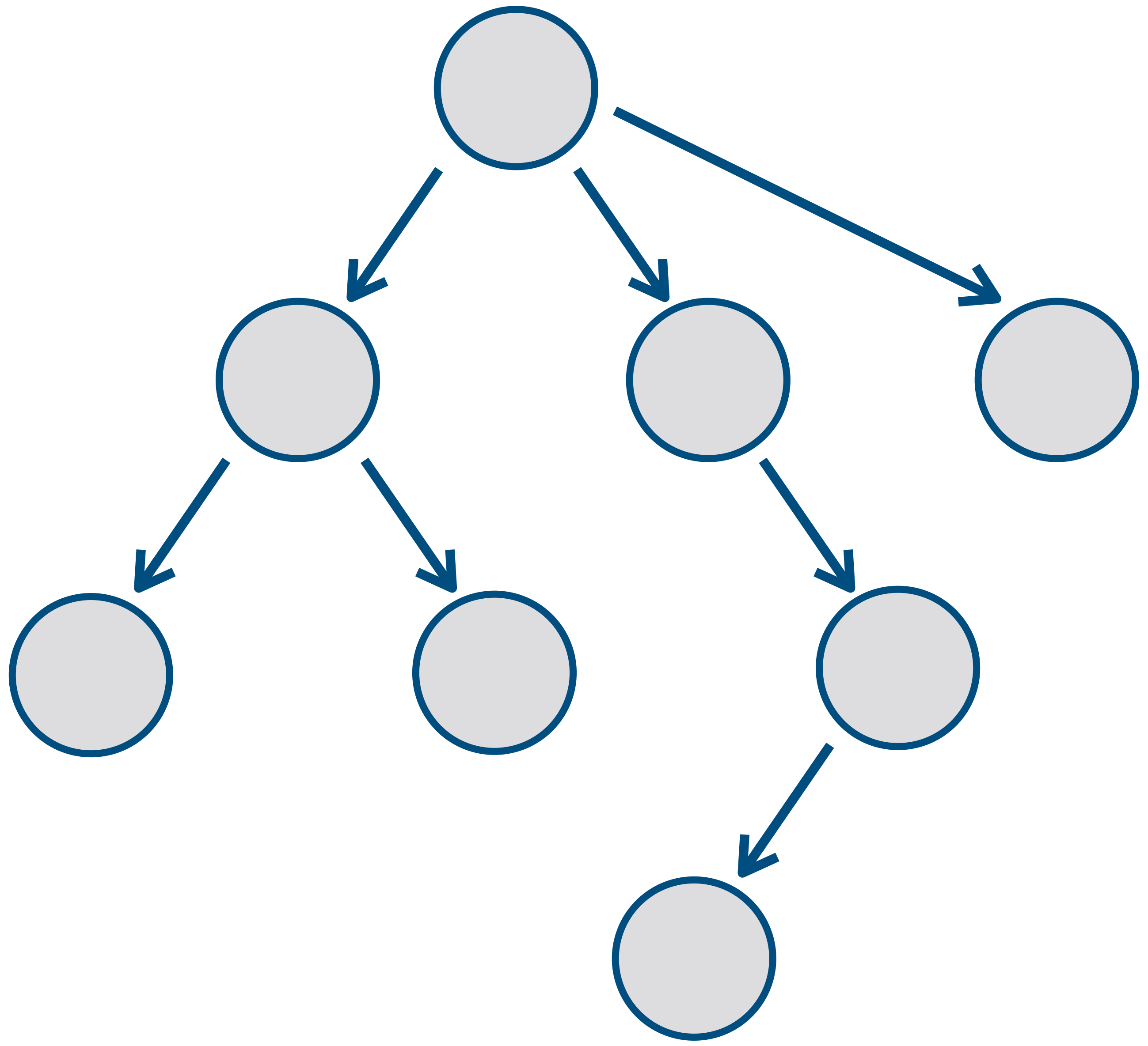
Preprocessing

- Remove variable names and instance-of relation
- No anonymisation
- Sentence piece model with 32K operations



Pretraining

- Pretraining on silver AMRs
  - 30M sentences from CCNET
  - Using JAMR



Decoding into multiple Languages

- XLM cross-lingual embeddings and vocabulary (32K sentence piece subwords)
- Language Model pretraining on 30M sentences
- Multilingual Encoder-Decoder

French

Des responsables américains

….

Spanish

Funcionarios estadounidenses

….

Slovak

Americkí predstavitelia

….

Bulgarian

Американските служители

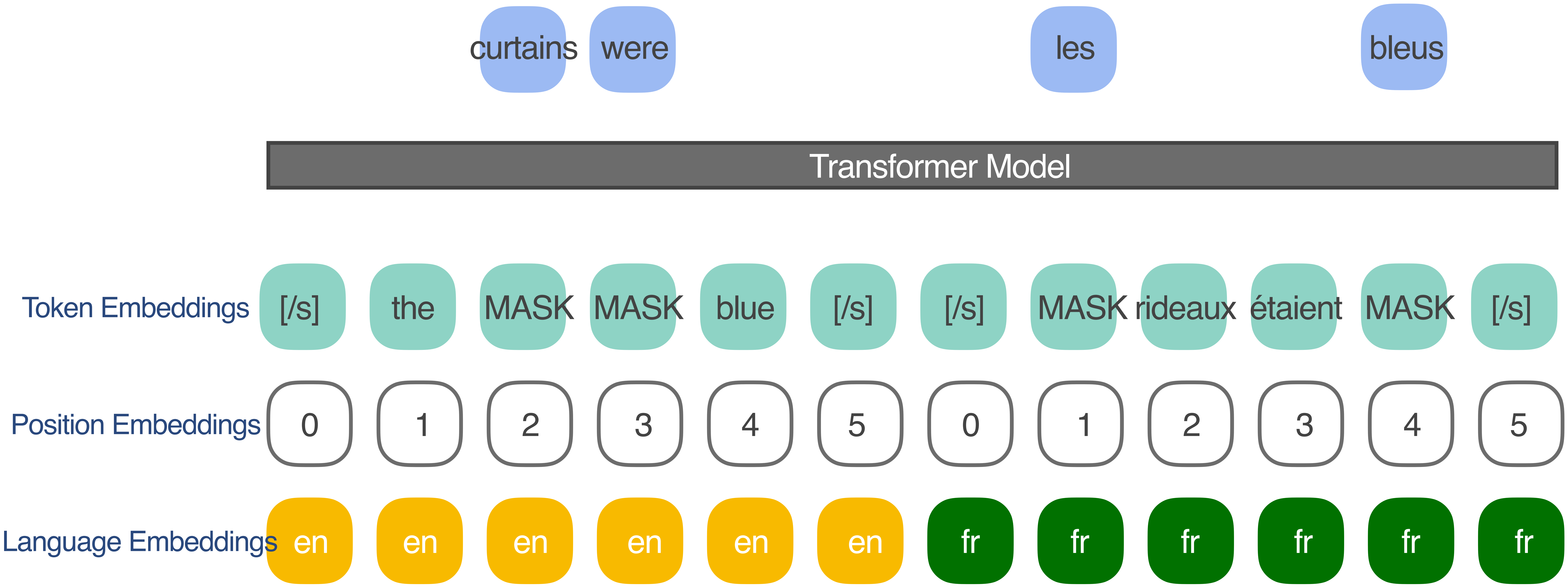
….

Swedish

Amerikanska tjänstemän

…

XLM Cross-lingual embeddings



Cross-lingual Language Model Pretraining
Guillaume Lample, Alexis Conneau

Multilingual Encoder-Decoder

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.

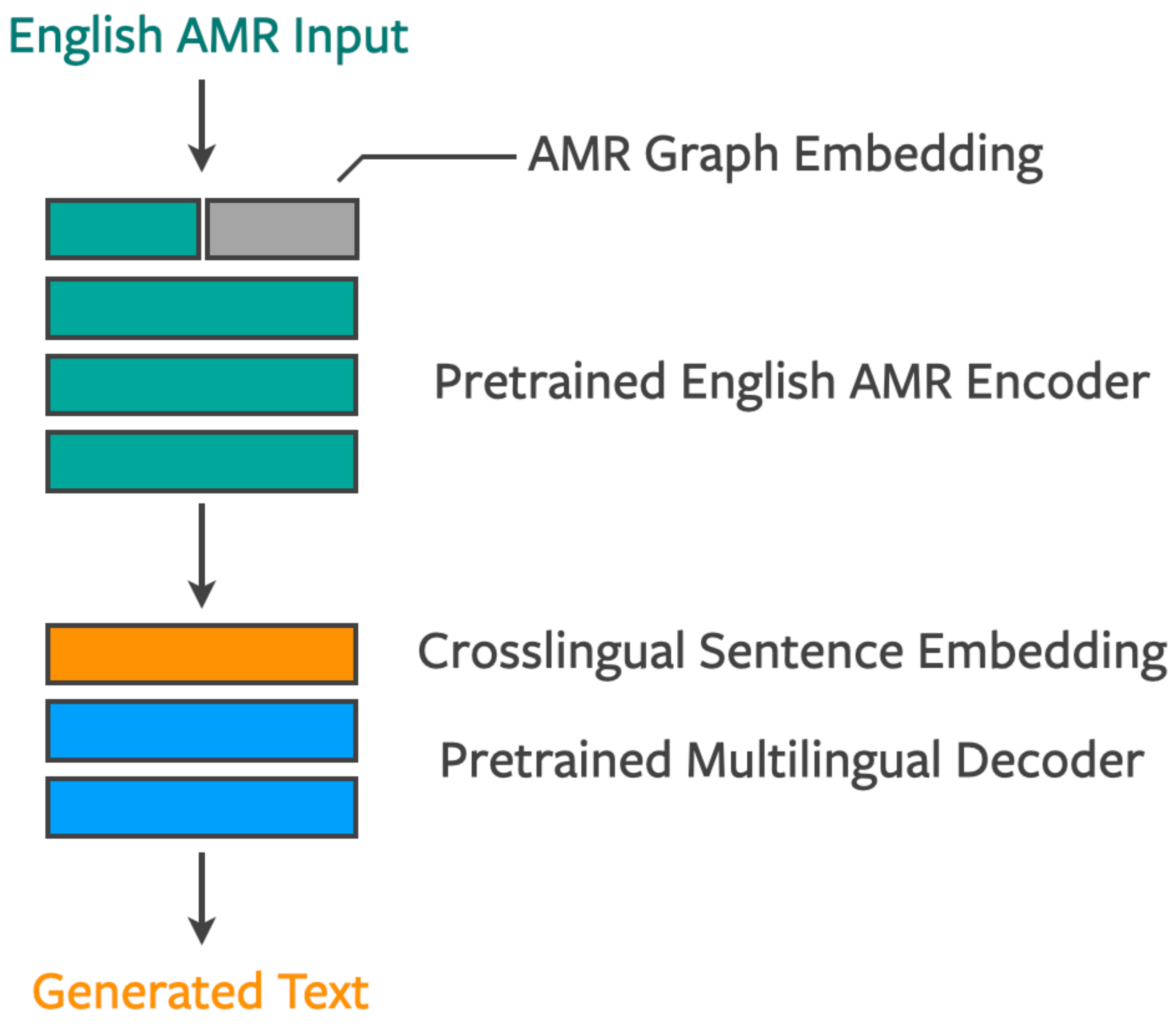
Decoding into French

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.

Multilingual AMR-to-NL Model



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

DATA

Training Data

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

https://github.com/jflanigan/jamr

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.

EVALUATION

Evaluation

Automatic (BLEU)

- Ablation
- Comparison with two strong baselines
- Impact of training data (which languages ?)
- Correlation I/O (sub)word overlap and BLEU

Human-Based

- Word-Oder, Morphology, Semantic adequacy, Paraphrasing

Ablation Study

Base Model (English)	32.5
+ Graph embeddings	32.9
+ Crosslingual embeddings.	33.0
+ Encoder pretraining	33.4
+ Decoder pretraining	33.8

https://github.com/jflanigan/jamr

Comparison: Monolingual v. Multilingual

Monolingual 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: Monolingual v. Multilingual

Monolingual 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

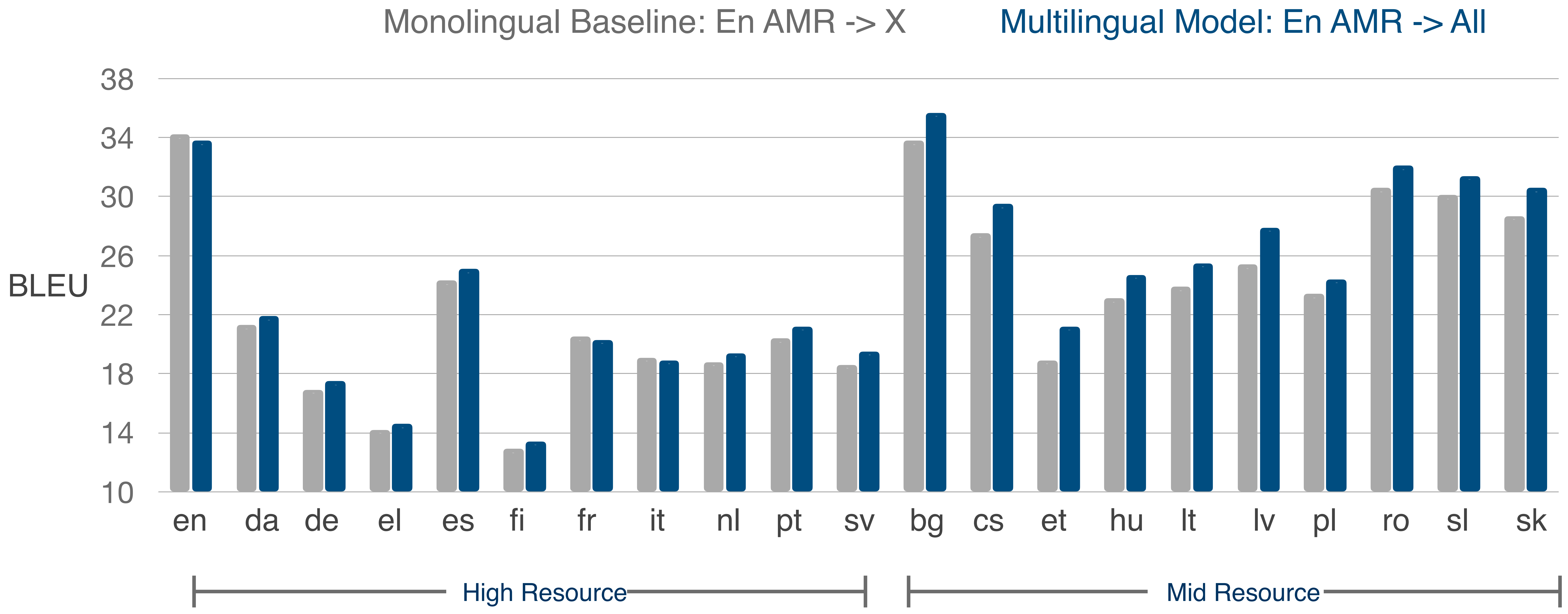
fr

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

↓

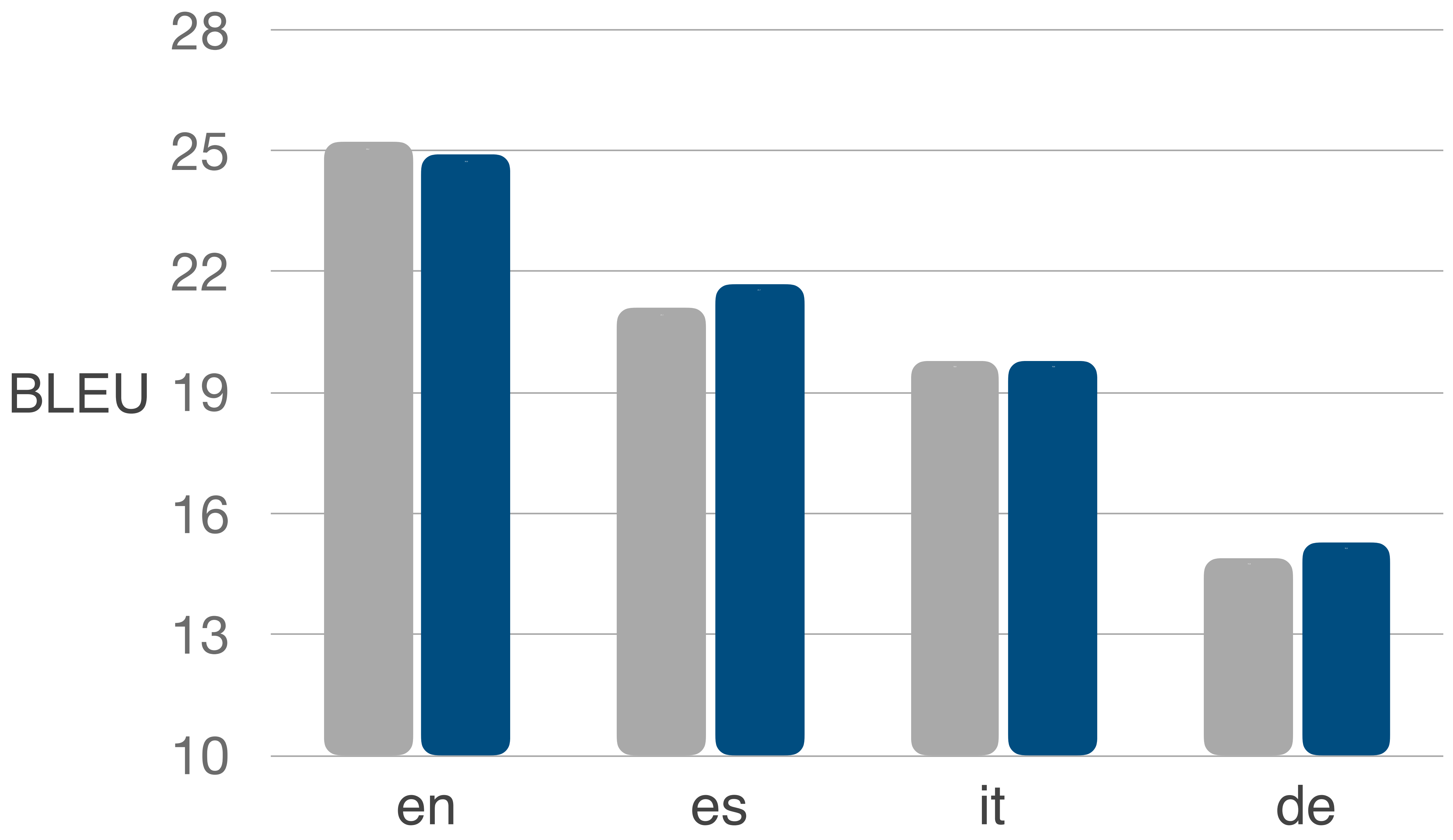
Des responsables américains ont tenu une réunion d'un groupe d'experts en janvier 2002 à New York.

Results: Europarl



Results: Gold AMR

Bilingual Baseline: En AMR -> X   Multilingual Model: En AMR -> All



Comparison: Hybrid Translation v. 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 Translation v. 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.

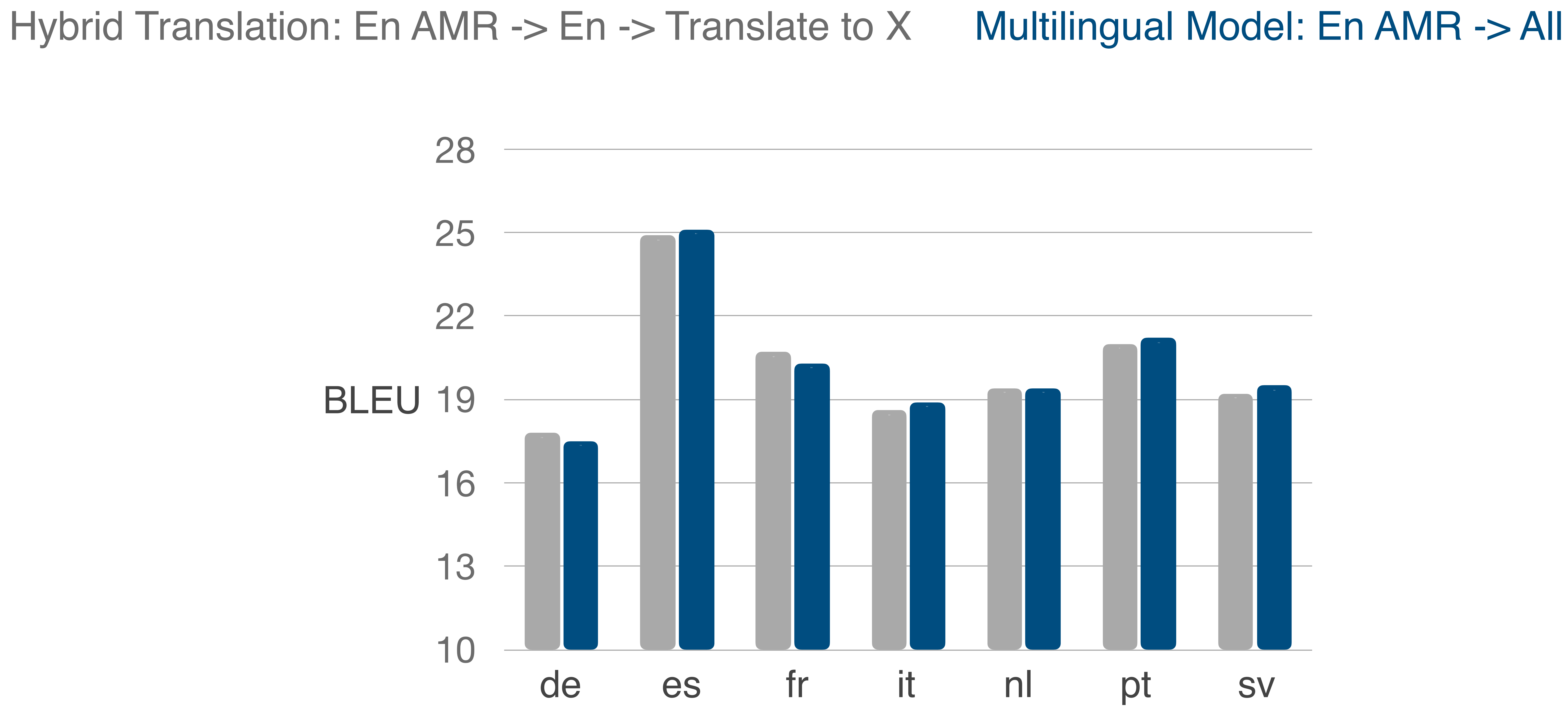
Multilingual Model

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 to Hybrid Translation Baseline

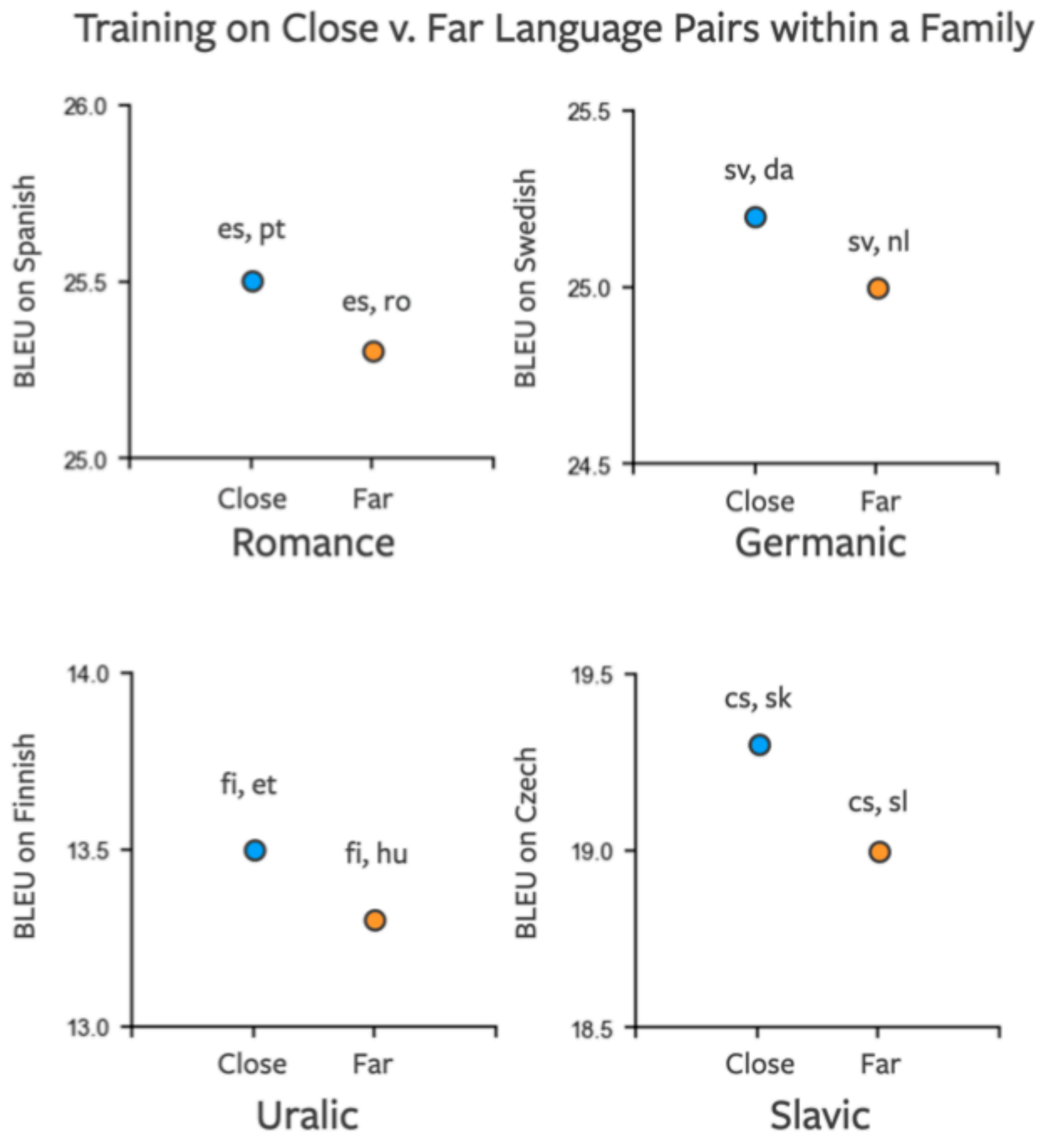


Training on languages from the same family

	Da	De	Nl	Sv
One Language	21.3	17.0	18.5	18.7
Germanic Family	21.8	21.9	19.6	19.3
All Languages	21.9	17.5	19.4	19.5

Training on the closest language

- Multilingual models trained on language pairs
- Within a family, the most closely related pairs get best results
  - Romance: Spanish/Portuguese
  - Germanic: Swedish/Danish
  - Uralic: Finnish/Estonian
  - Slavic: Czech/Slovak



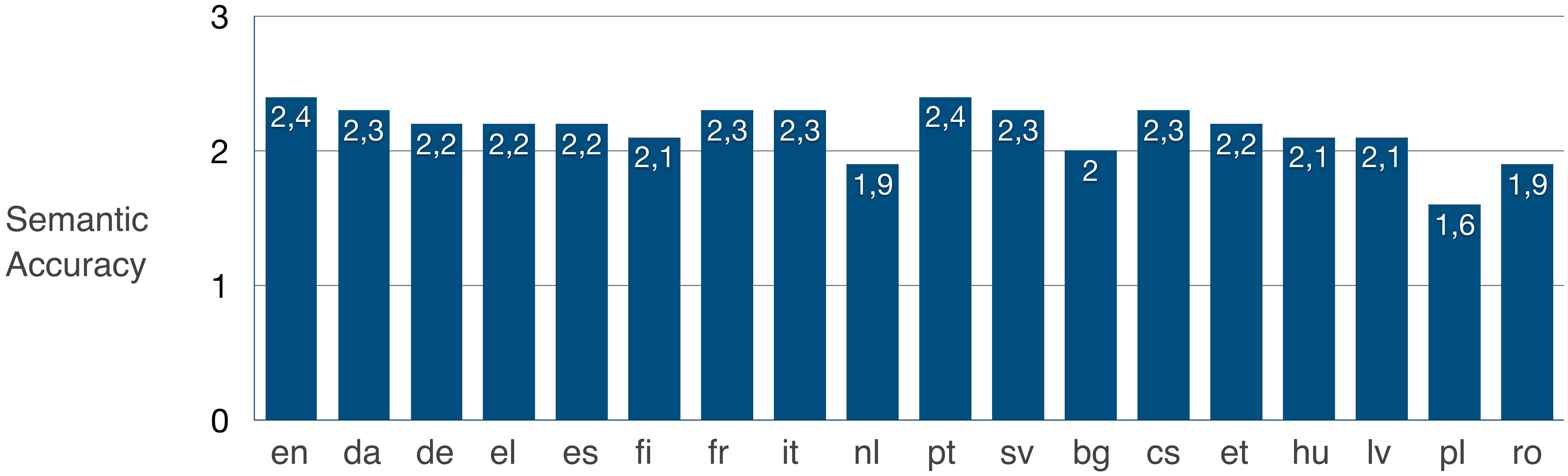
Human Evaluation

- Semantic Accuracy:
  Does the hypothesis 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



Human Evaluation

The scores are uniformly high across languages for both Morphology and Word Order

A Multilingual model generalises well across languages

Example Paraphrases

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

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

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

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

Human evaluation demonstrates multilingual techniques generalize across languages

Human evaluation demonstrates multilingual techniques generalize across languages

Multilingual benefits from increased training data and performs better than monilingual

Human evaluation demonstrates multilingual techniques generalize across languages

Multilingual benefits from increased training data and performs better than monilingual

Using English-Centric AMR, we can decode into many different target-side languages

Retrieval-Based Generation of Long Form Text

Generating Woman Biographies

Generating Wikipedia Biographies from Web Retrieval

PERSON NAME

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]

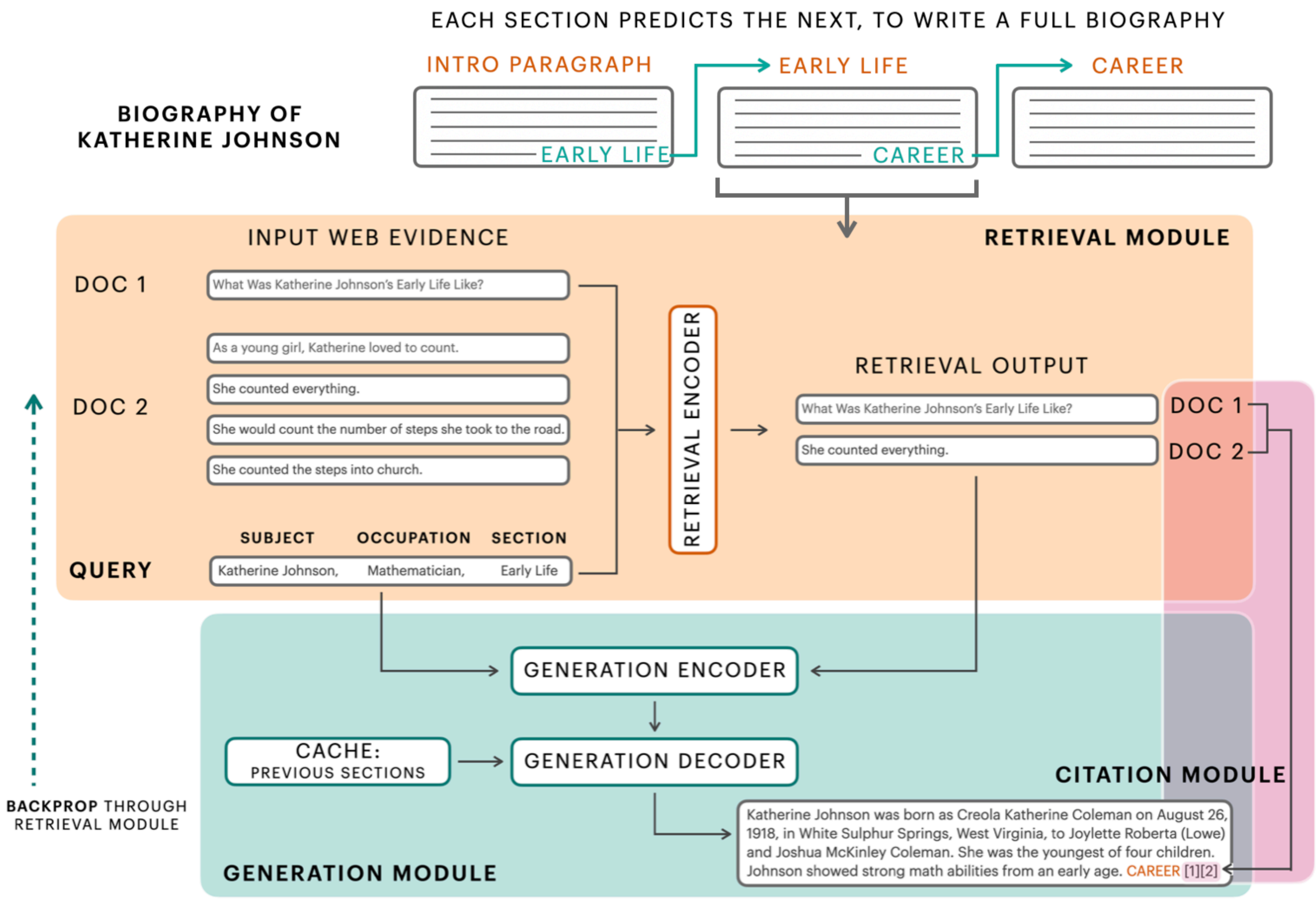
Paton was active in the RAOU, as well as in the South Australian Ornithological Association (SAOA), of which she was elected Vice-President 1974–1979, and President 1979–1982. She was one of the first women to become a member of the exclusive Adelaide Ornithologists Club, of which she was elected president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

Legacy and honours

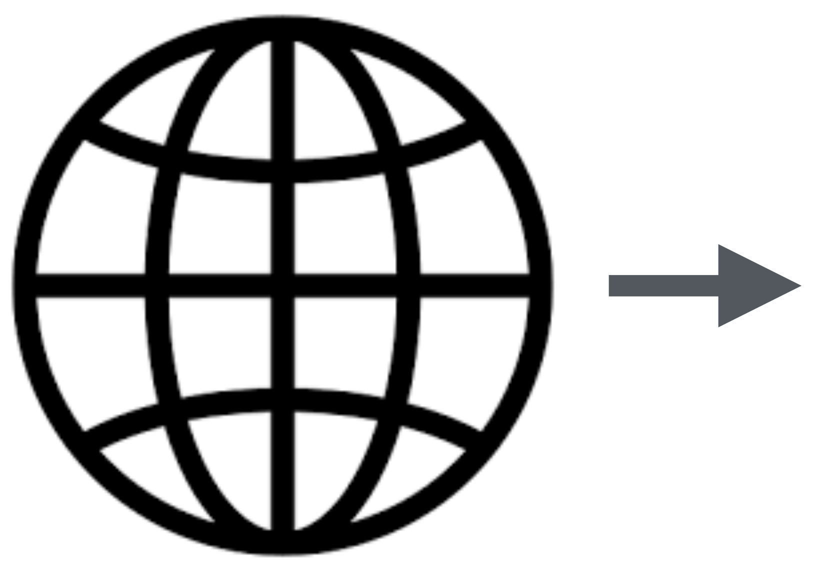
- 1990, she was made an Honorary Member of the SAOA.
- 1996, she was made an Honorary Member of the Adelaide Ornithologists Club.

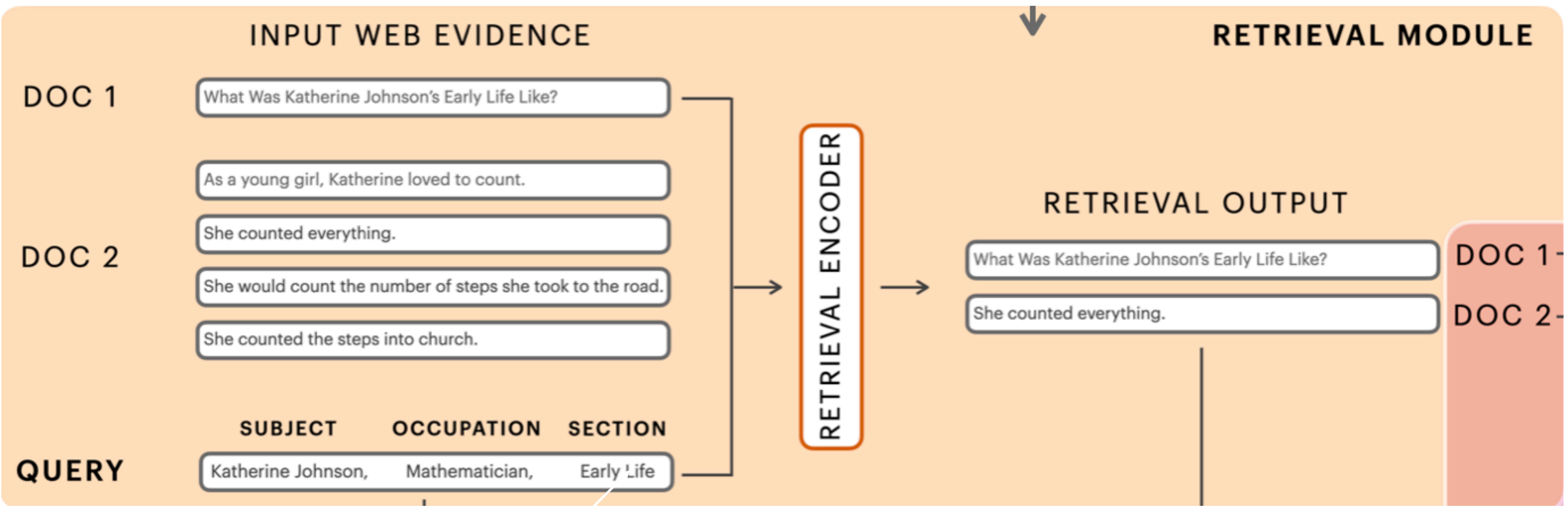
Challenges

- Gather relevant evidence
- Generate a structured text
- Ensure factuality


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

Retrieval





QUERY

Katherine Johnson

Mathematician

Early Life

SEARCH OUTPUT

Top 20 search results segmented into sentences

OUTPUT

40 sentences most similar with the query

(1,000 words)

Generation

QUERY

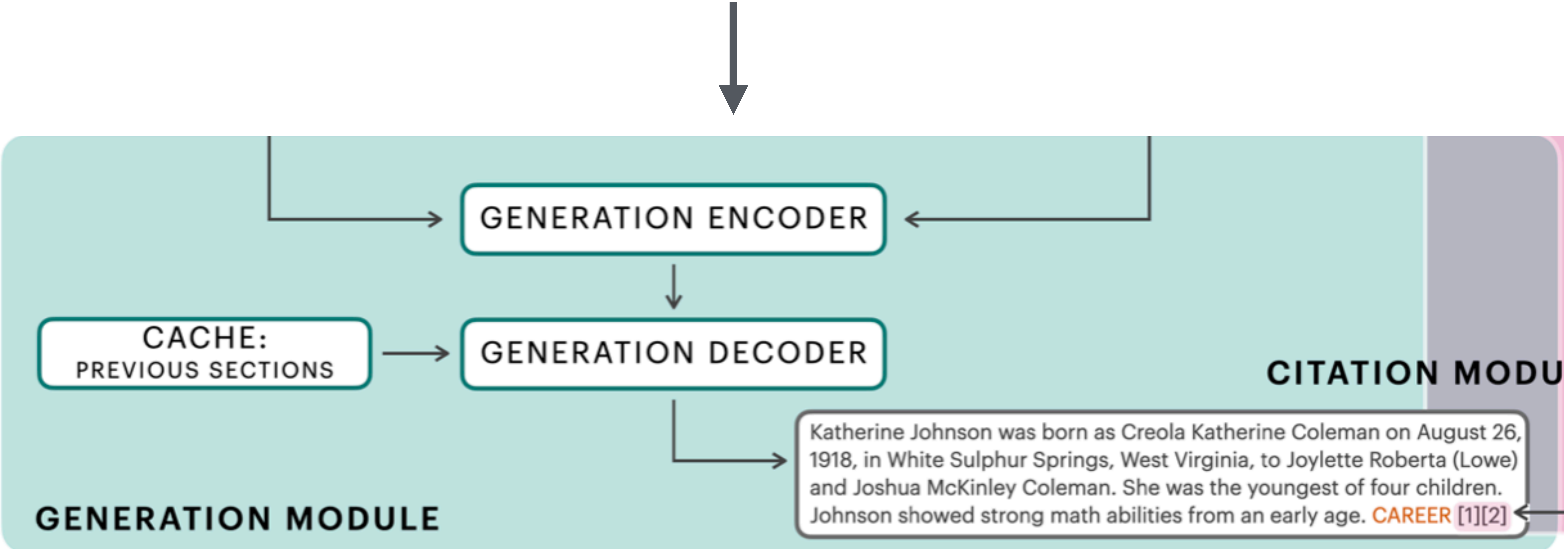
Katherine Johnson

Mathematician

Early Life

RETRIEVED EVIDENCE

1,000 words



Transformer-XL Cache Mechanism

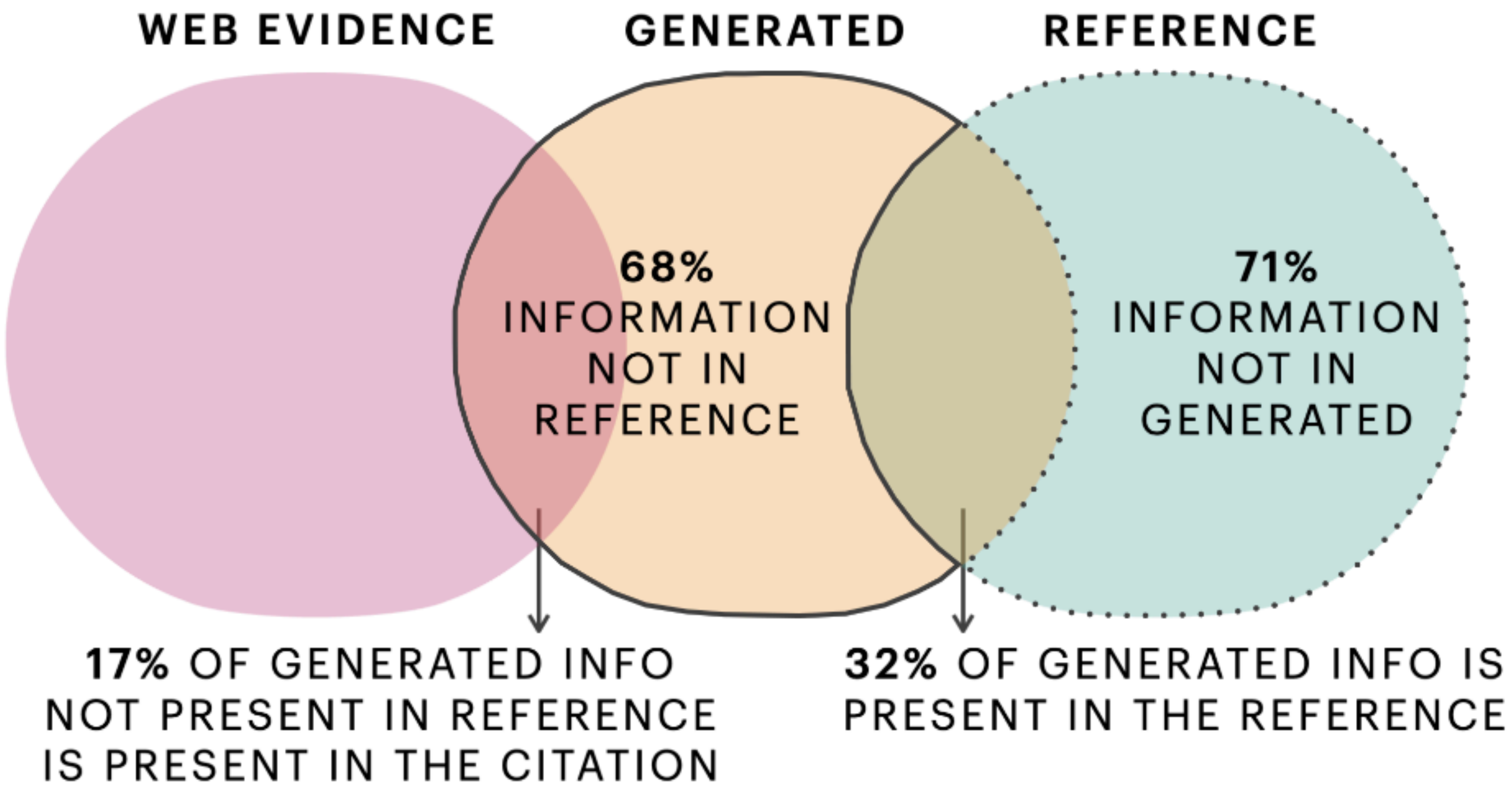
- Caches the previous section’s hidden states at every later
- Usd as a memory to generate the current section

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

The retrieval and the cache module statistically significantly improve results

Human Evaluation of Factuality



The Evidence Gap

Data

(person name, web evidence, Wikipedia biography)

- Wikisum: Wikipedia biographies
- Our dataset: Women 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%

Less Web Evidence, Less Good Texts

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

Conclusion

Open Challenges

- Factuality
  - Evaluation
  - Improvement

- Other Generation Tasks
  - Document level Simplification
  - Multi-document, multi-format, summarisation
  - Multilingual KB verbalisation, simplication, summarisation

Thank you !