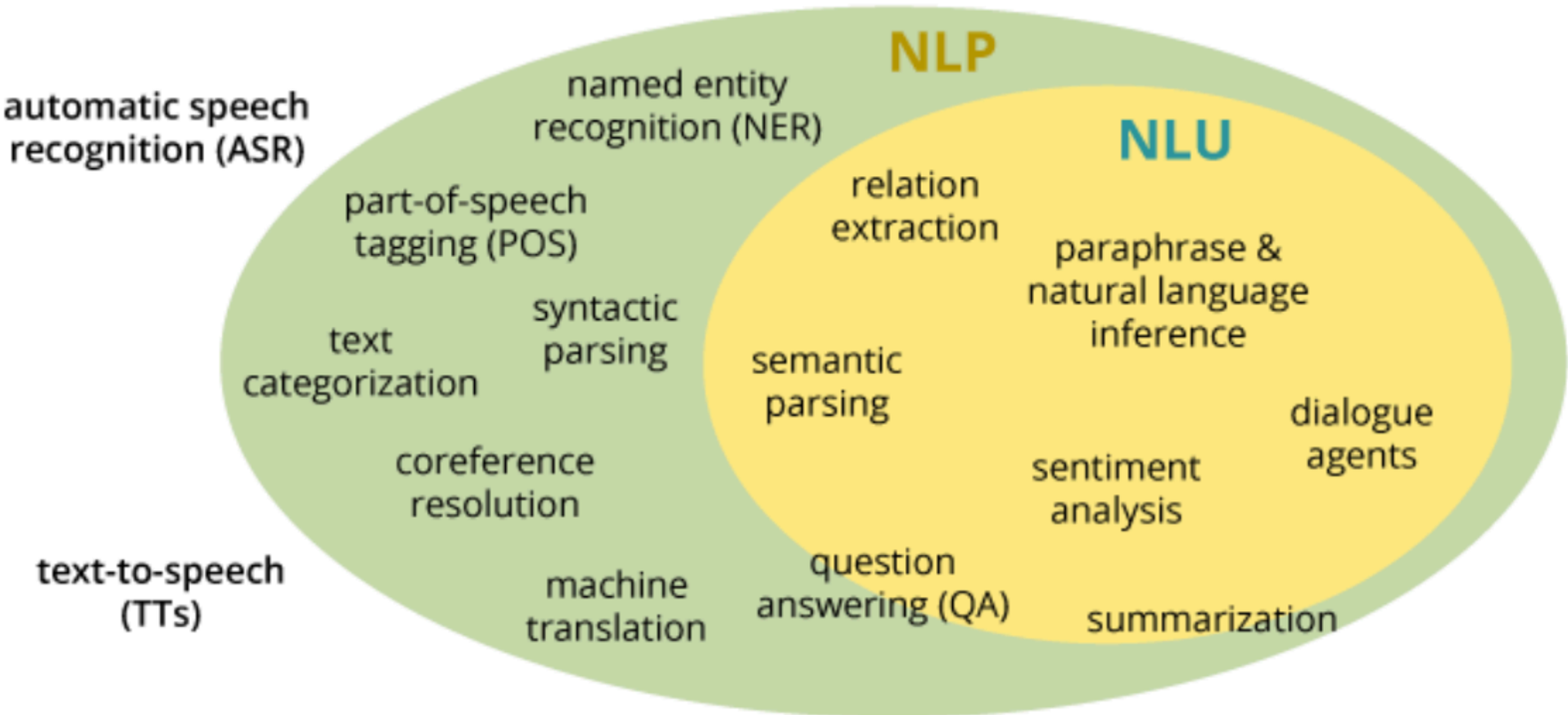
자연언어처리에서의 의미 (Meaning in Natural Language Processing)

인공지능과 포스트휴머니즘 워크샵
2019.9.27
신효필(서울대학교 언어학과)

Contents

- NLU vs. NLP
- Introduction to Meaning
    - What is Meaning?
    - Meaning Representation
- Lexical Meaning Representation
    - Word Embeddings
- Recent Trends in Word/Sentence Embeddings
- Triggering Debate on Meaning
- Meaning in NLP ?
- Reference

NLU vs. NLP



Source: https://nlp.stanford.edu/~wcmac/papers/20140716-UNLU.pdf

Introduction to Meaning:(by Emily M. Bender)

Understanding semantics and pragmatics is beneficial for building scalable NLP systems

How do strings of words relate to speaker intent?

• General purpose NLU requires modeling that which is general across language

• Knowledge of semantics and pragmatics can inform the design of NLP systems for understanding and generation

• Mapping between surface forms and intent representations

What is Meaning?

- Meaning in Natural Language is a Multi-facetted concept with semantic, pragmatic, cognitive and social aspect

- In part: Validity or truth conditional

$$\frac{\begin{array}{l}\text{All cats are mortal}\\ \text{Euler is a cat}\end{array}}{\text{Euler is mortal}} \qquad \frac{\begin{array}{l}\forall x(cat(x) \rightarrow mortal(x))\\ cat(e)\end{array}}{mortal(e)}$$

- Interpretation of logical formulae (LF) is compositional as is (generally) the translation of natural language expressions to LFs

- Validity is relevant for inference and thus RTE, QA, IE

What is Meaning?

- Not just validity or truth conditions
  - Construction meaning representations also requires commonsense reasoning:

John can open Bill's safe.

… He knows the combination.

… He should change the comb

Meaning Representation (by J. Einstein, 2018)

- Semantic analysis involves converting natural language into a meaning representation.

- To be useful, a meaning representation must meet several criteria:

    - c1: it should be **unambiguous**: unlike natural language, there should be exactly one meaning per statement;

    - c2: it should provide a way to **link** language to external knowledge, observations, and actions;

    - c3: it should support computational **inference**, so that meanings can be combined to derive additional knowledge;

    - c4: it should be **expressive** enough to cover full range of things that people talk about in natural language

Meaning Representations from J. Einstein(2018)

- Logical Semantics
    - Meaning and denotation
    - Logical representations of meaning
        - Propositional logic
        - First-order logic
- Semantic parsing and the lambda calculus
    - The lambda calculus
    - Quantification
- Learning semantic parsers
    - Learning from derivations
    - Learning from logical forms
    - Learning from denotations

Meaning Representations from J. Einstein(2018)

- Predicate-argument semantics
    - Semantic roles
        - VerbNet
        - Proto-roles and PropBank
        - FrameNet
    - Semantic role labeling
        - Semantic role labeling as classification
        - Semantic role labeling as constraint optimization
        - Neural semantic role labeling
- Abstract Meaning Representation
    - AMR Parsing

Meaning Representations from J. Einstein(2018)

- Distributional and distributed semantics
    - The distributional hypothesis
    - Latent semantic analysis
    - Brown Clusters
    - Neural word embeddings
        - Continuous bag-of-words(CBOW)
        - Skipgrams
        - Computational complexity
        - Word embeddings as matrix factorization
- Distributed representations beyond distributional statistics
    - Word-Internal structure
    - Lexical semantic resources

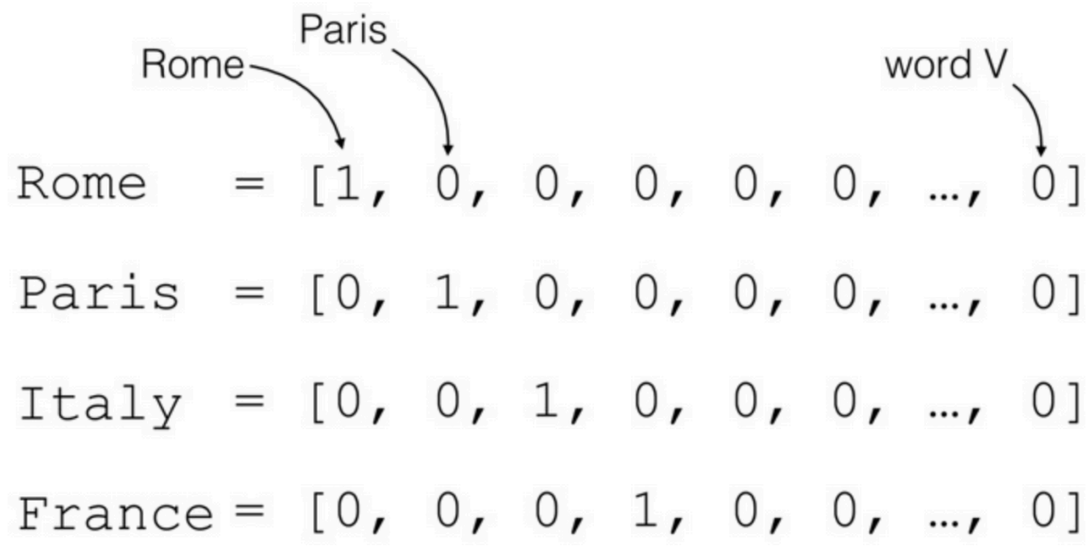
Meaning Representations from J. Einstein(2018)

- Distributed representations of multiword units
  - Purely distributional methods
  - Distributional-compositional methods
  - Supervised compositional methods
  - Hybrid distributed-symbolic representations

Lexical Meaning Representation: Word Embeddings

- A word embedding is a real number, vector representation of a word.

- One-Hot Encoding (Count Vectorizing)

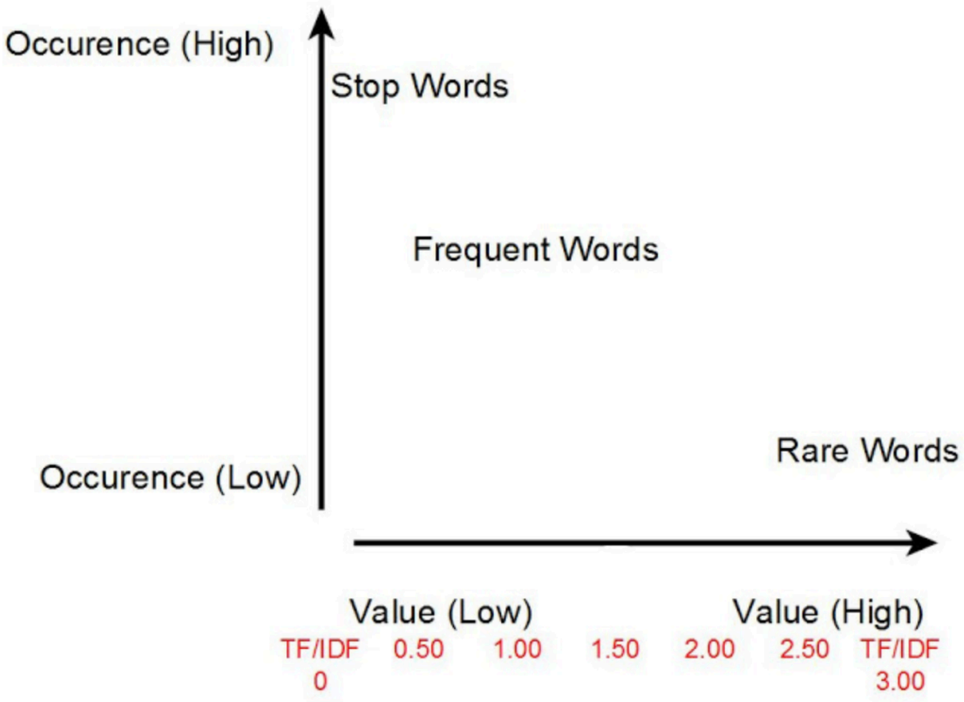


one-hot encoding

Word Embeddings
- TF-IDF Transform-

- Words are represented by their term frequency multiplied by their inverse document frequency.

- Words that occur a lot but everywhere should be given very little weighting or significance



TF-IDF

Word Embeddings
- Co-Ocurrence Matrix-

$$X = \begin{array}{c} \\ I \\ like \\ enjoy \\ deep \\ learning \\ NLP \\ flying \\ . \end{array} \begin{array}{c} \begin{array}{cccccccc} I & like & enjoy & deep & learning & NLP & flying & . \end{array} \\ \begin{bmatrix} 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

co-occurrence Matrix

Word Embeddings
-Neural Probabilistic Model-

- Word2Vec
- Glove
- FastText
- Poincare Embeddings (Hierarchal Representations)
    - They use hyperbolic geometry to capture hierarchal properties of words.
    - By placing their embedding into a hyperbolic space, they can use properties of hyperbolic space to use distance to encode similarity and the norm of vectors to encode hierarchal relationships.

ELMO(Embeddings from Language Models)

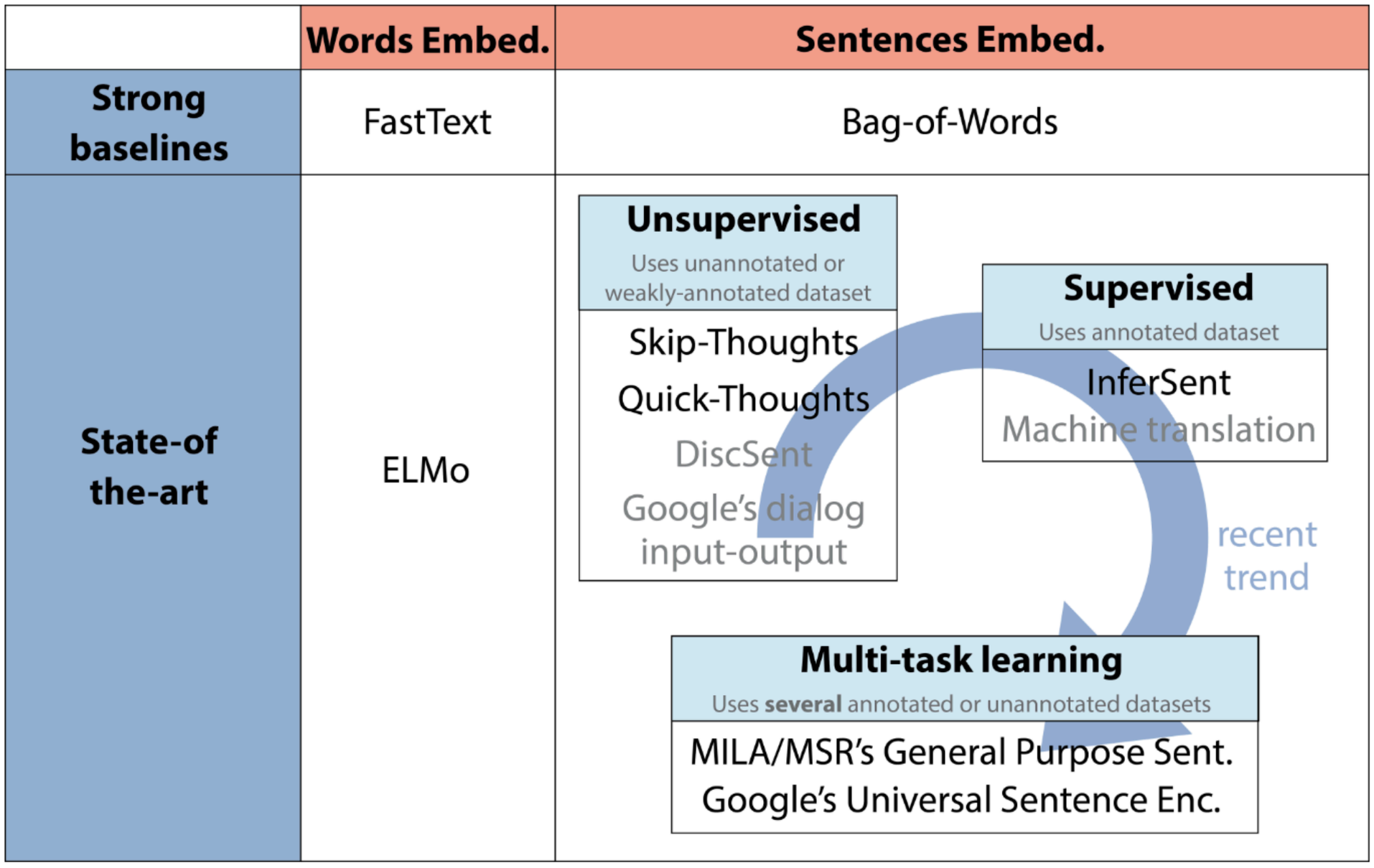
- ELMO (Peters et. al., 2018)

  - Instead of using a fixed embedding for each word, ELMo looks at the entire sentence before assigning each word in it an embedding. It uses a bi-directional LSTM trained on a specific task to be able to create those embeddings.

  - ELMo's inputs are **characters** rather than words

  - **ELMo are concatenations of the activations on several layers of the biLMs.** Different layers of a language model encode different kind of information on a word (e.g. Part-Of-Speech tagging is well predicted by the lower level layers of a biLSTM while word-sense disambiguation is better encoded in higher-levels)

Recent Trends in Word/Sentence Embeddings (from https://medium.com/huggingface/universal-word-sentence-embeddings-ce48ddc8fc3a)



Recent Trends in Universal Word / Sentence Embeddings

Universal Sentence Embeddings -Skip-thoughts Vectors

- **Skip-thoughts vectors**

  - Proposed by Jamie Kiros and co-workers in 2015

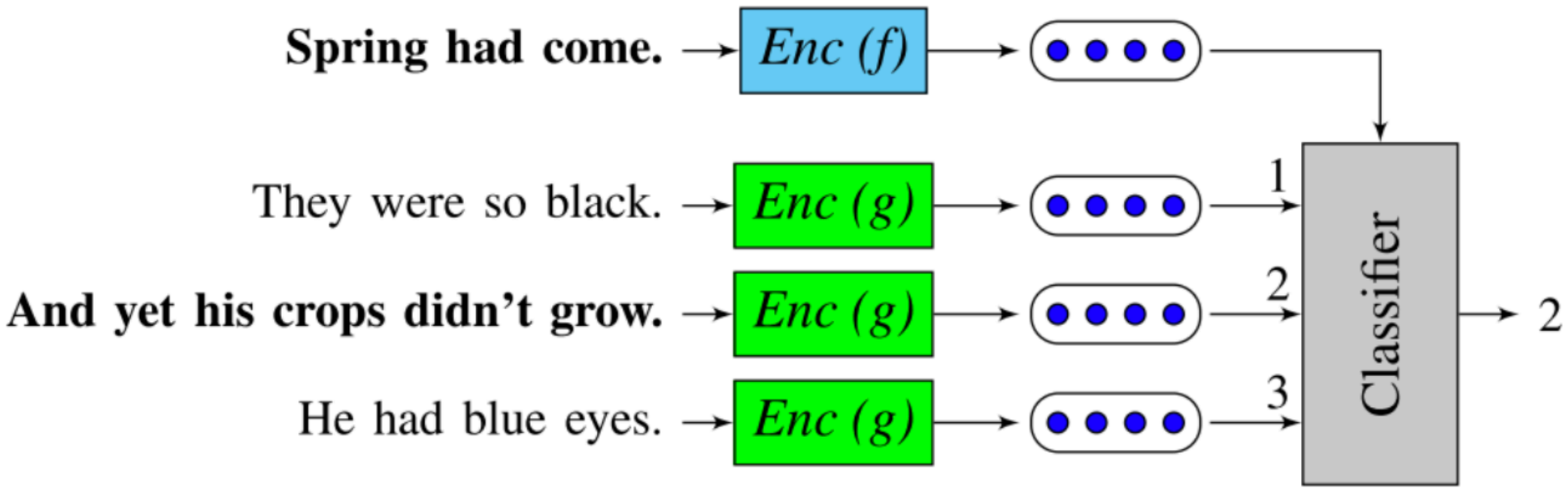
  - The archetypical example of learning unsupervised sentence embeddings.

  - It can be though as the equivalent for sentences of the skip-gram model developed for word embeddings: *rather than predicting the words surrounding a word, we try to predict the surroundings sentences of a given sentence*.

Universal Sentence Embeddings -Quick-thoughts Vectors

- A recent development of the Skip-thoughts vectors
- In this work, the task of predicting the next sentence given the previous one is reformulated as a **classification** task: *the decoder is replaced by a classifier which has to choose the next sentence among a set of candidates*.
- It can be interpreted as a discriminative approximation to the generation problem.

Universal Sentence Embeddings -Quick-thoughts Vectors-



Quick-thoughts

Universal Sentence Embeddings -InferSent-

- **InferSent** uses the *Stanford Natural Language Inference (SNLI) Corpus* (a set of of 570k pairs of sentences labelled with 3 categories: neutral, contradiction and entailment) to train a classifier on top of a sentence encoder.

- Both sentences are encoded using the same encoder while the classifier is trained on a pair representation constructed from the two sentence embeddings.

Universal Sentence Embeddings -Google's Sentence Encoder-

- **Google's Universal Sentence Encoder** uses a transformer-network that is trained on a variety of data sources and a variety of tasks with the aim of dynamically accommodating a wide variety of natural language understanding tasks. A pre-trained version has been made available for TensorFlow.

BERT (from McCormick (2019))

- 2018 was a breakthrough year in NLP.
    - Transfer learning, particularly models like Allen AI's ELMO, OpenAI's Open-GPT, and Google's BERT allowed researchers to smash multiple benchmarks with minimal task-specific fine-tuning and provided the rest of the NLP community with pretrained models that could easily (with less data and less compute time) be fine-tuned and implemented to produce state of the art results.
- BERT (Bidirectional Encoder Representations from Transformers), released in late 2018
- BERT is a method of pretraining language representations that was used to create models that NLP practicioners can then download and use for free.
- You can either use these models to extract high quality language features from your text data, or you can fine-tune these models on a specific task (classification, entity recognition, question answering, etc.) with your own data to produce state of the art predictions.

BERT (from McCormick (2019))

- Why BERT Embeddings?
    - We use BERT to extract features, namely word and sentence embedding vectors, from text data.
    - What can we do with these word and sentence embedding vectors? First, these embeddings are useful for keyword/search expansion, semantic search and information retrieval.
    - More importantly, these vectors are used as high-quality feature inputs to downstream models.
    - BERT offers an advantage over models like Word2Vec, because while each word has a fixed representation under Word2Vec regardless of the context within which the word appears, BERT produces word representations that are dynamically informed by the words around them. For example, given two sentences: “The man was accused of robbing a bank.” “The man went fishing by the bank of the river.”
    - Word2Vec would produce the same word embedding for the word “bank” in both sentences, while under BERT the word embedding for “bank” would be different for each sentence. Aside from capturing obvious differences like polysemy, the context-informed word embeddings capture other forms of information that result in more accurate feature representations, which in turn results in better model performance.

ULM-FiT (Universal Language Model Fine-tuning for Text Classification)

- By fast.ai founder Jeremy Howard and Sebastian Ruder
- ULM-FiT introduced methods to effectively utilize a lot of what the model learns during pre-training – more than just embeddings, and more than contextualized embeddings. ULM-FiT introduced a language model and a process to effectively fine-tune that language model for various tasks.

The Transformer

- The release of the Transformer paper and code, and the results it achieved on tasks such as machine translation started to make some in the field think of them as a replacement to LSTMs.

- This was compounded by the fact that Transformers deal with long-term dependancies better than LSTMs.

OpenAI Transformer: Pre-training a Transformer Decoder for Language Modeling

- It turns out we don't need an entire Transformer to adopt transfer learning and a fine-tunable language model for NLP tasks.

- We can do with just the decoder of the transformer. The decoder is a good choice because it's a natural choice for language modeling (predicting the next word) since it's built to mask future tokens – a valuable feature when it's generating a translation word by word.

- The model stacked twelve decoder layers. Since there is no encoder in this set up, these decoder layers would not have the encoder-decoder attention sublayer that vanilla transformer decoder layers have.

- With this structure, we can proceed to train the model on the same language modeling task: predict the next word using massive (unlabeled) datasets. Just, throw the text of 7,000 books at it and have it learn! Books are great for this sort of task since it allows the model to learn to associate related information even if they're separated by a lot of text – something you don't get for example, when you're training with tweets, or articles.

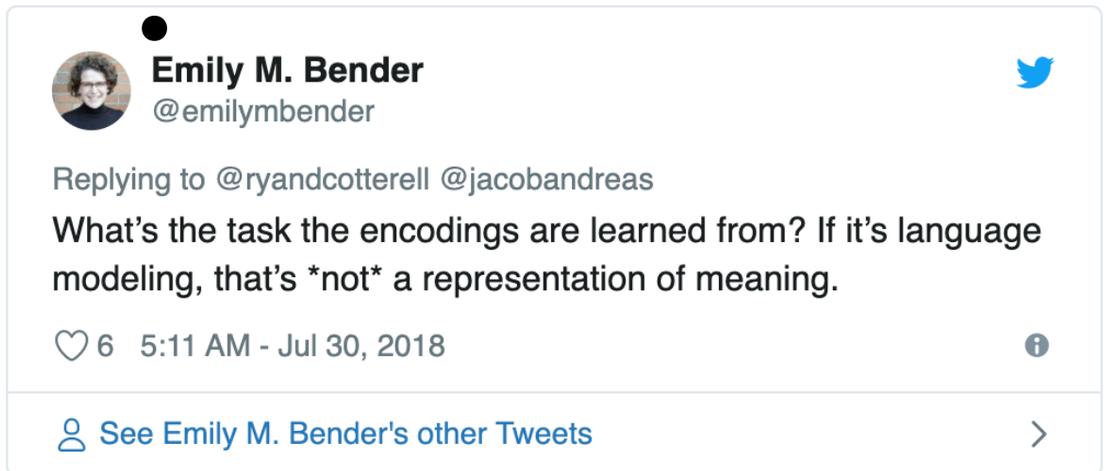
Triggering a debate on Meaning (from Thomas Wolf, https://medium.com/huggingface/learning-meaning-in-natural-language-processing-the-semantics-mega-thread-9c0332dfe28e)

- Triggering a debate on Meaning

- This argument was the main trigger for the mega-discussion that followed.

- Further in the thread, Emily M. Bender reformulates her argument as: *If all the learner gets is text the learner cannot learn meaning.*





Can a model trained only on raw text learn meaning? (from Thomas Wolf, https://medium.com/huggingface/learning-meaning-in-natural-language-processing-the-semantics-mega-thread-9c0332dfe28e)

- What aspect of Meaning can a model learn?
  - Can it learn *meaning* or just learn *similarity in meaning* (i.e. learn that some expressions are similar without knowing what they mean. It is still very useful for transfer learning)?
  - Can it learn *grounded meaning* (learn the meaning of each expression as a state of the world state) or learn *lexical meaning* (e.g. learn how the meaning of sub-expressions compose together, as in our logical forms)?

- How can the model Learn?
  - If the model cannot learn meaning from raw text alone, what would be the *minimal amount of additional supervision* needed? Should we add supervision from Logical Forms, Textual Entailment…?
  - Could we encode some *inductive bias* in the model so that it can learn aspects of meaning from raw text?

Meaning in NLP?

- Especially difficult for current natural language processing technologies are multi-word expressions that are not compositional (e.g., “Going Dutch” doesn’t mean the composition of the semantics of “going” and being “Dutch”).

- Even more difficult is metaphor, humor, satire, and sarcasm, where context, style and personality of the writer colors the semantics of what is written.

- “Crash blossoms” : *Violinist Linked to JAL Crash Blossoms*

    - What is the meaning of this sentence?

I

References

- Thomas Wolf (2018), Learning Meaning in Natural Language Processing - The Semantics Mega-Thread (https://medium.com/huggingface/learning-meaning-in-natural-language-processing-the-semantics-mega-thread-9c0332dfe28e)
- Thomas Wolf (2018), The Current Best of Universal Word Embeddings and Sentence Embeddings(https://medium.com/huggingface/universal-word-sentence-embeddings-ce48ddc8fc3a)
- Ryan Kiros et al. (2015), Skip-Thought Vectors (https://arxiv.org/abs/1506.06726)
- Lajanugen Logeswaran, and Honglak Lee (2018). An Efficient Framework for Learning Sentence Representations (https://openreview.net/forum?id=rJvJXZb0W)
- Yacine Jernite, et al. (2017), Discourse-Based Objectives for Fast Unsupervised Sentence Representation Learning (https://arxiv.org/abs/1705.00557)
- Alexis Conneau et al. (2017) Supervised Learning of Universal Sentence Representations from Natural Language Inference Data (https://arxiv.org/abs/1705.02364)
- Daniel Cer et al. (2018), Universal Sentence Encoder (https://arxiv.org/abs/1803.11175)
- Matthew E. Peters et al. (2018), Deep Contextualized Word Representation (https://arxiv.org/abs/1802.05365)
- fastText, https://github.com/facebookresearch/fastText
- Glove (Gloval Vectors for Word Representation), https://nlp.stanford.edu/projects/glove/

References

- Word2Vec, https://github.com/dav/word2vec/

- Jay Alammar (2018), The illustrated BERT, ELMO, and co. (How NLP cracked transfer learning), http://jalammar.github.io/illustrated-bert/

- Jacob Einstein (2018), Natural Language Processing(https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes-10-15-2018.pdf)

- Emily M. Bender (2018), 100 Things You Always Wanted to Know about Semantics & Pragmatics But Were Afraid to Ask (http://faculty.washington.edu/ebender/papers/Bender-ACL2018-tutorial.pdf)

- Hunter Heidenreich (2018), Introduction to Word Embeddings (http://hunterheidenreich.com/blog/intro-to-word-embeddings/)

- McCormick Chris (2019), BERT Word Embeddings Tutorial (http://mccormickml.com/2019/05/14/BERT-word-embeddings-tutorial/)



기계번역 기술의 이해와 발전

NAVER 파파고 신중휘 책임리더



목차

1. 기계번역이란
2. 기계 번역 기술 발전의 흐름
3. 오번역의 원인과 대응
4. 향후 연구 방향
5. Q & A



papago

1. 기계 번역이란



1. 기계 번역이란

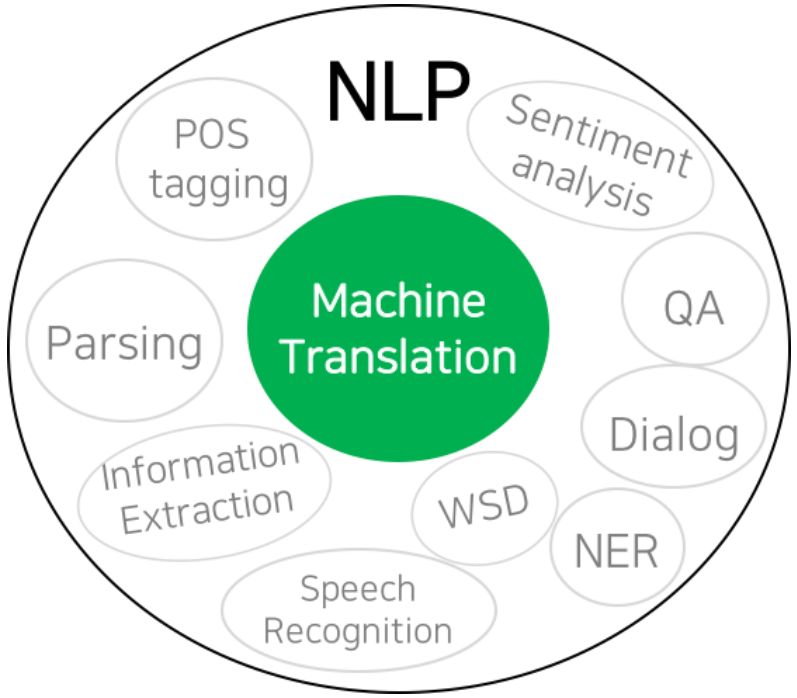
papago



기계 번역(機械飜譯, Machine translation)은 인간이 사용하는 자연 언어를 컴퓨터를 사용하여 다른 언어로 번역하는 일을 말한다. '자동 번역'이라고도 한다.

From wikipedia

1. 기계 번역이란



* 기계번역 또한 NLP의 한 분야

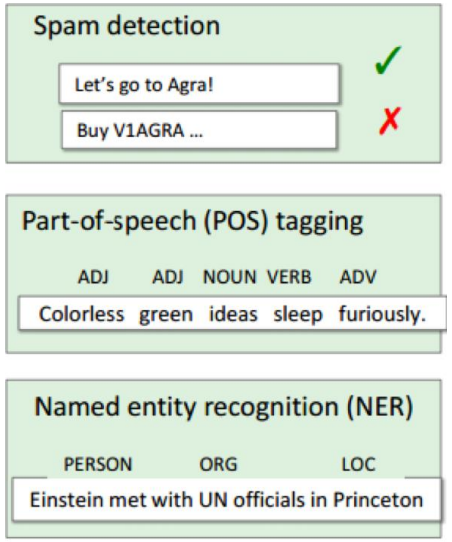
NLP : 자연어처리
(Natural Language Processing)

프로그래밍 언어처럼 기계가 이해하는 언어가 아닌
한국어, 영어 처럼 사람이 사용하는 언어를
기계가 분석하고 이해해서 사람과 커뮤니케이션 할 수
있도록 해 주는 연구분야

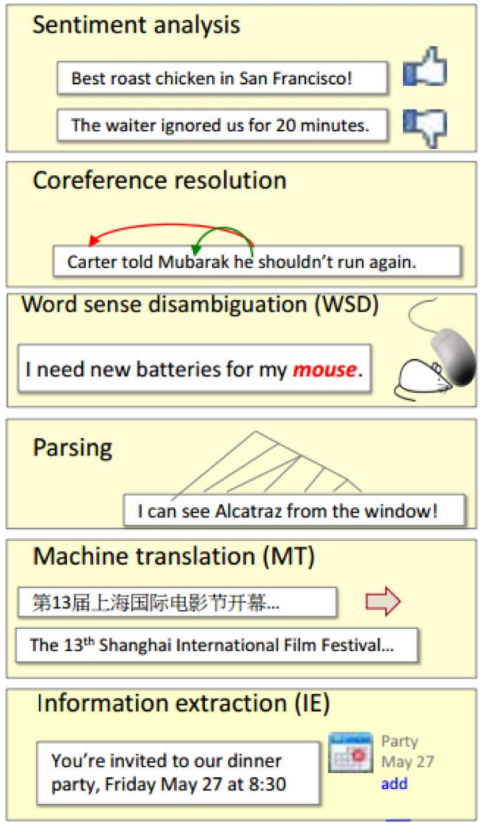
1. 기계 번역이란

papago

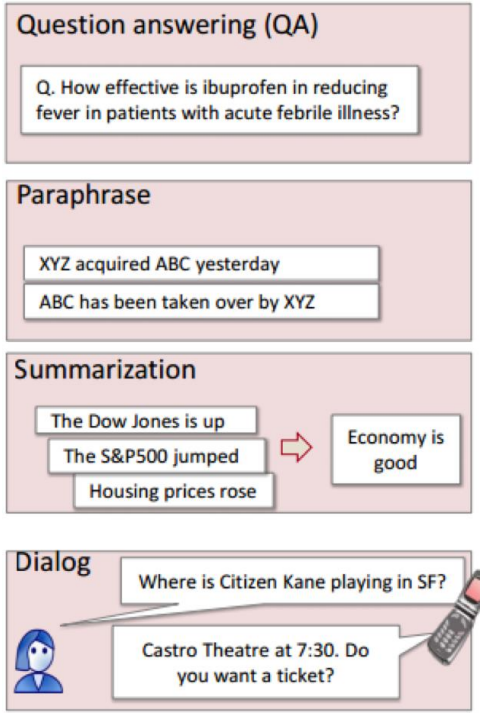
Mostly solved



Making good Progress



Still really hard



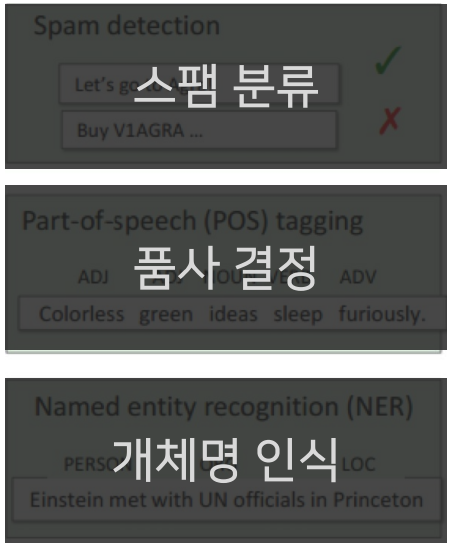
NAVER

1. 기계 번역이란

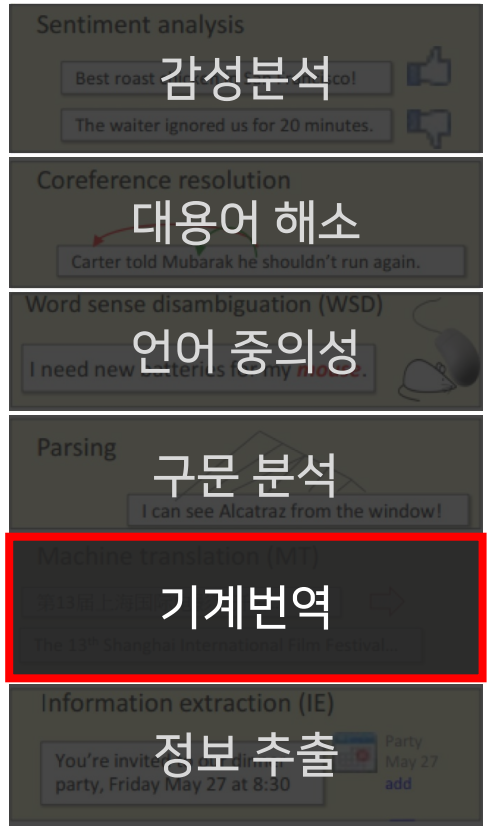
papago



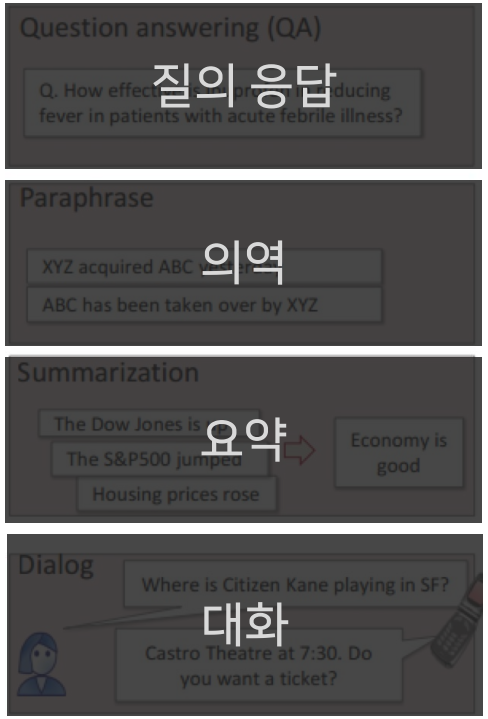
Mostly solved



Making good Progress



Still really hard



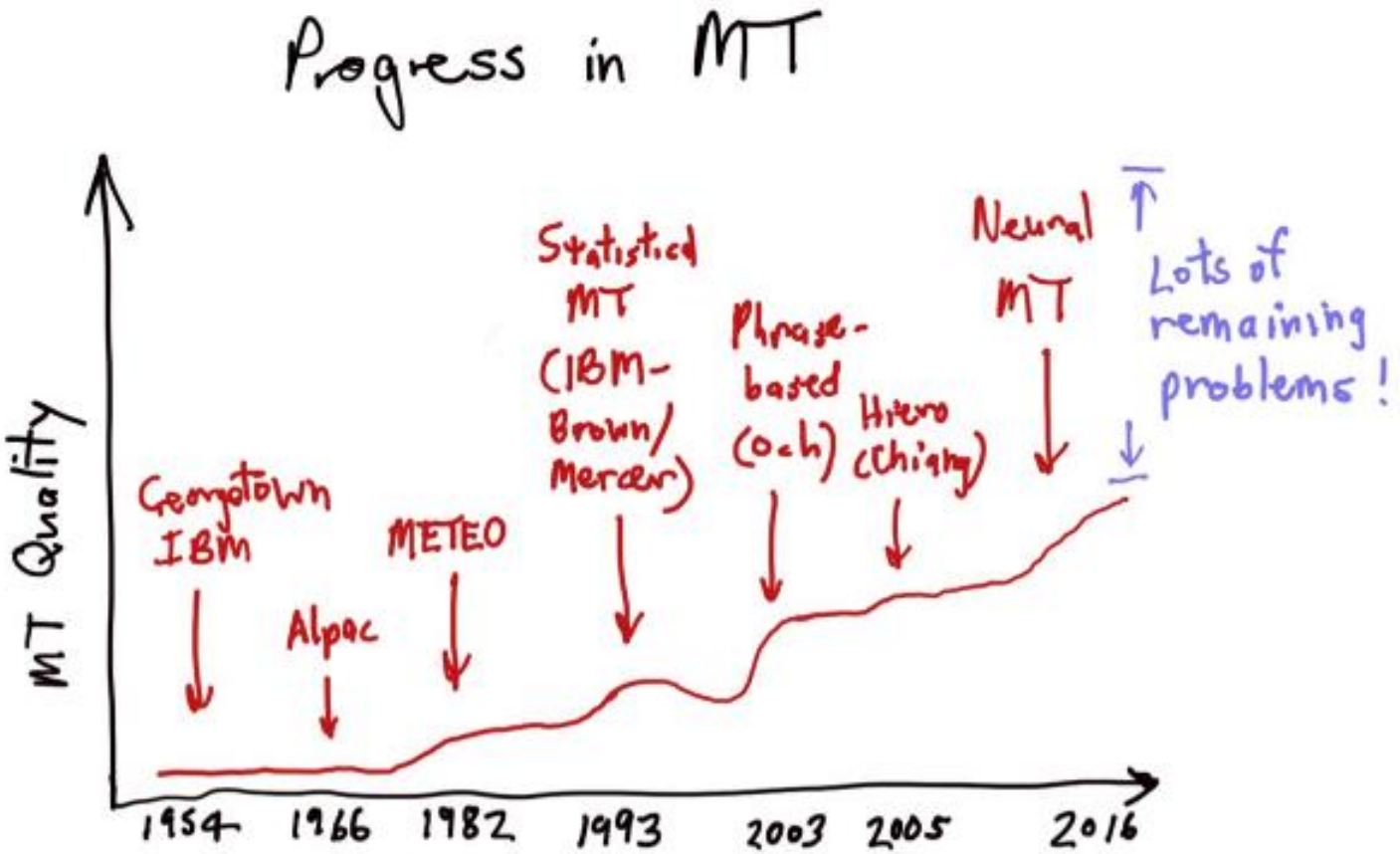


Image source: Luong, Cho, Manning's ACL 2016 Tutorial Slides



papago

2. 기계 번역 **기술 발전의 흐름**



2. 기계 번역 **기술 발전의 흐름**

papago

약 70년

1949 Rule based Machine Translation
규칙 기반의 기계번역

1954 Georgetown – IBM

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2006 Google PBSMT 적용
2011 NAVER 기계번역팀 구성
2012 NAVER PBSMT 적용

2014 Neural Machine Translation
인공신경망 기계번역

2016 NAVER Papago NMT 적용
Google NMT 적용

2017 NMT Transformer
인공신경망 기계번역의 발전형

2. 기계 번역 **기술 발전의 흐름**

papago

1949 **Rule based Machine Translation**
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형

워렌 위버(warren weaver)가 최초 제안

예를 들면..

규칙1. '사과'는 'apple'이라고 번역한다.
규칙2. '나'는 'I'라고 번역한다.
규칙3. 주어는 제일 앞에 둔다.

.

.

규칙 493567. 형용사가 앞에 있고 부사가 뒤에 오면…

2. 기계 번역 **기술 발전의 흐름**

papago

1949 **Rule based Machine Translation**
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형

워렌 위버(warren weaver)가 최초 제안

▲ 규칙에 적합한 경우 매우 정확

▽ 규칙에 없는 경우 번역이 안됨

▽ 전문가의 손길이 필요함

▽ 개발이 너무 오래 걸림

▽ 유지 보수가 어려움

2. 기계 번역 **기술 발전의 흐름**

papago

1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

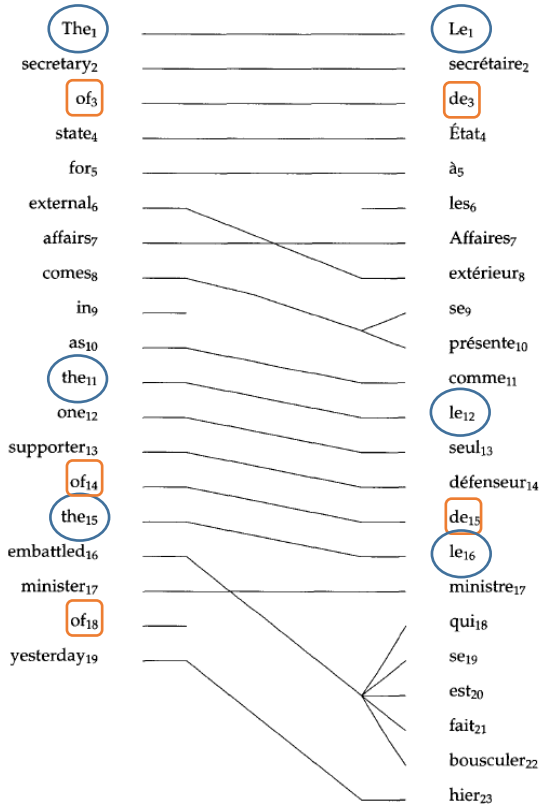
2017 NMT Transformer
인공신경망 기계번역의 발전형

단어 대 단어 대역 확률
[Translation Model]

nodding

f	$t(f \mid e)$	ϕ	$n(\phi \mid e)$
signe	0.164	4	0.342
la	0.123	3	0.293
tête	0.097	2	0.167
oui	0.086	1	0.163
fait	0.073	0	0.023
que	0.073		
hoche	0.054		
hocher	0.048		
faire	0.030		
me	0.024		
approuve	0.019		
qui	0.019		
un	0.012		
faites	0.011		

Alignment 확률
[Alignment Model]



NAVER

2. 기계 번역 **기술 발전의 흐름**

papago

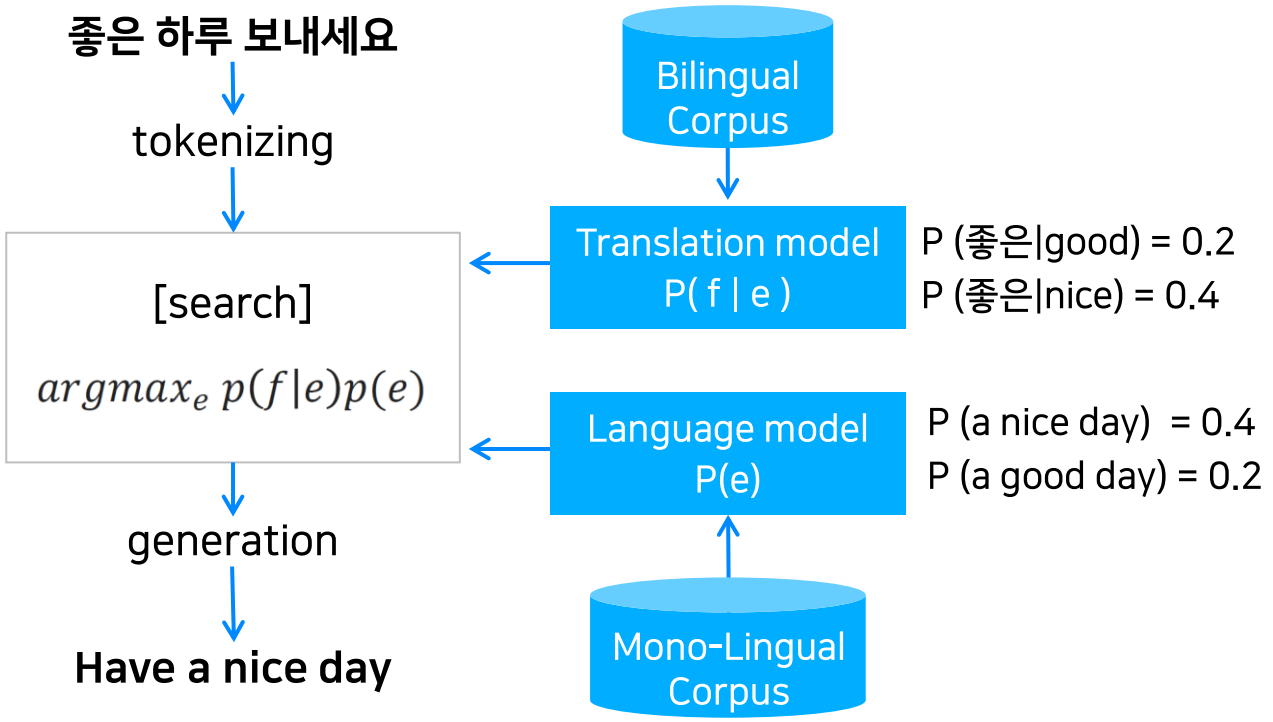
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



2. 기계 번역 **기술 발전의 흐름**

papago

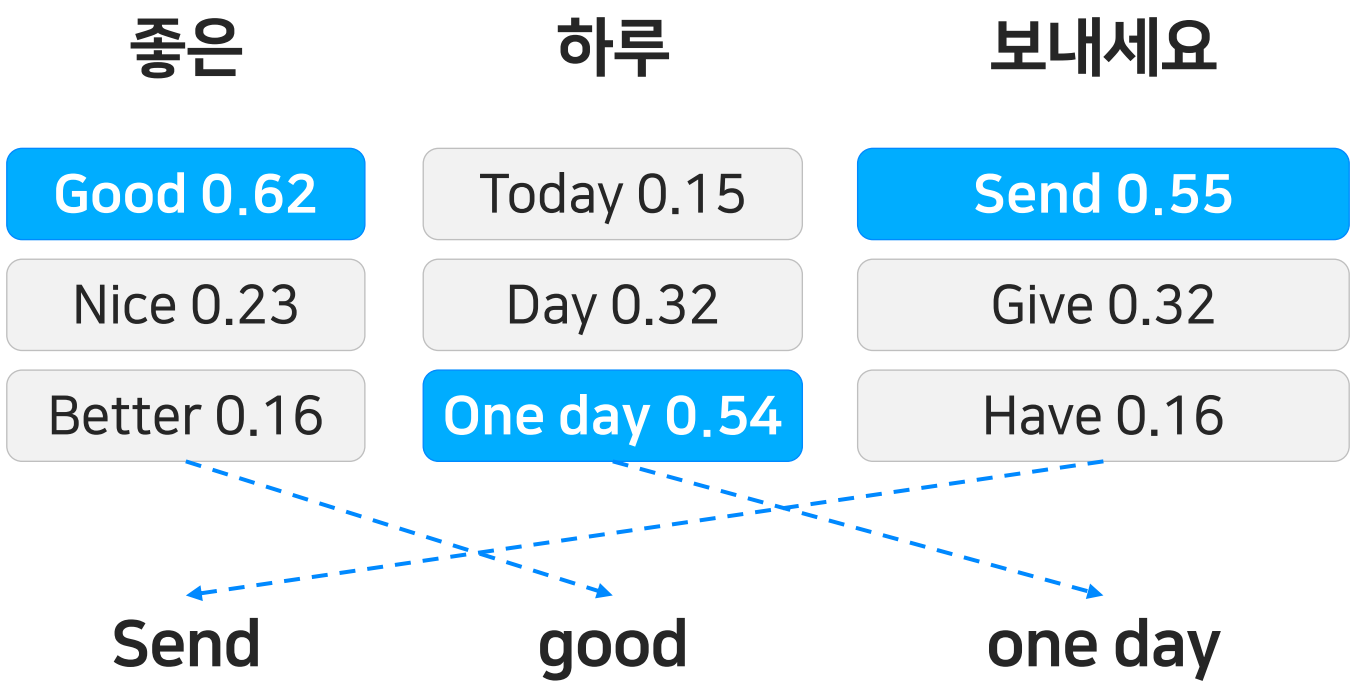
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



2. 기계 번역 **기술 발전의 흐름**

papago

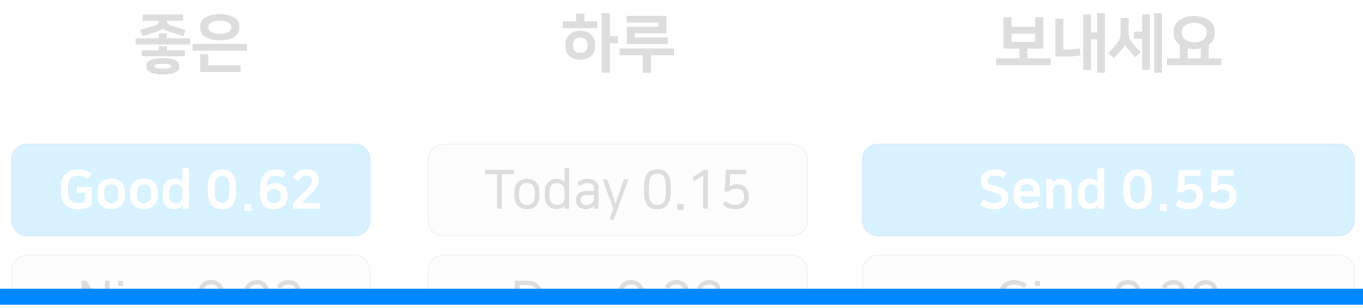
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



▲ **최초의 통계기반 기계학습 모델**

▲ **단어 단위 번역 확률 모델**

▽ **모호성 해소 실패**

▽ **규칙 기반 기계번역 극복 못함**

2. 기계 번역 **기술 발전의 흐름**

papago

1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

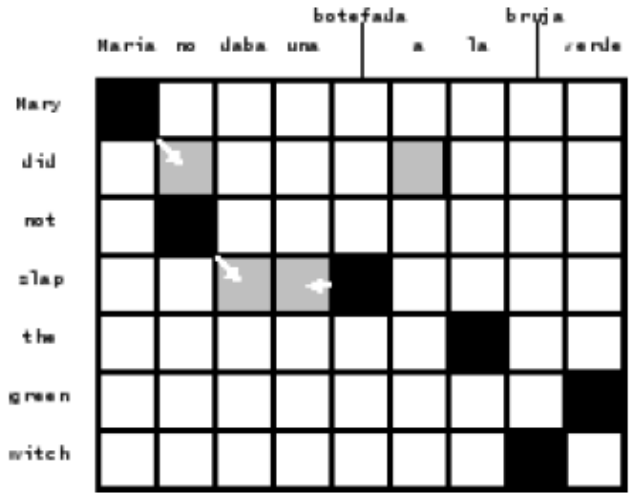
2017 NMT Transformer
인공신경망 기계번역의 발전형

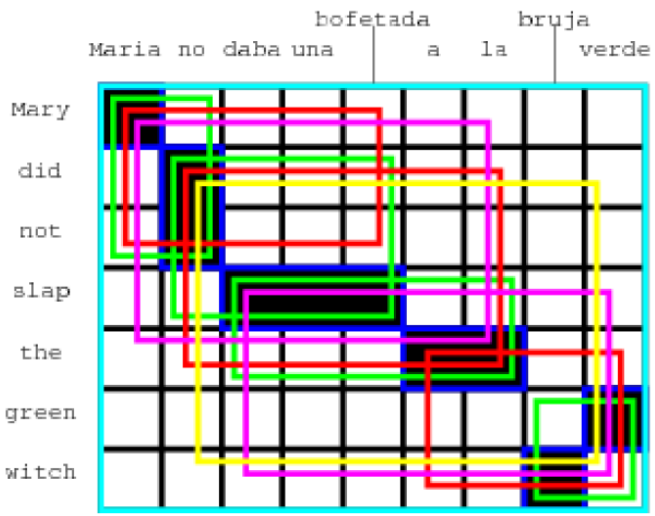
구 단위 번역 확률
[Phrase based Translation Model]

▲ 단어 단위 번역 확률의 한계를 극복

▽ 모든 구 단위 확률의 계산 복잡도는 너무 높음

Alignment를 이용한 효과적인 Phrase Extraction





NAVER

2. 기계 번역 **기술 발전의 흐름**

papago

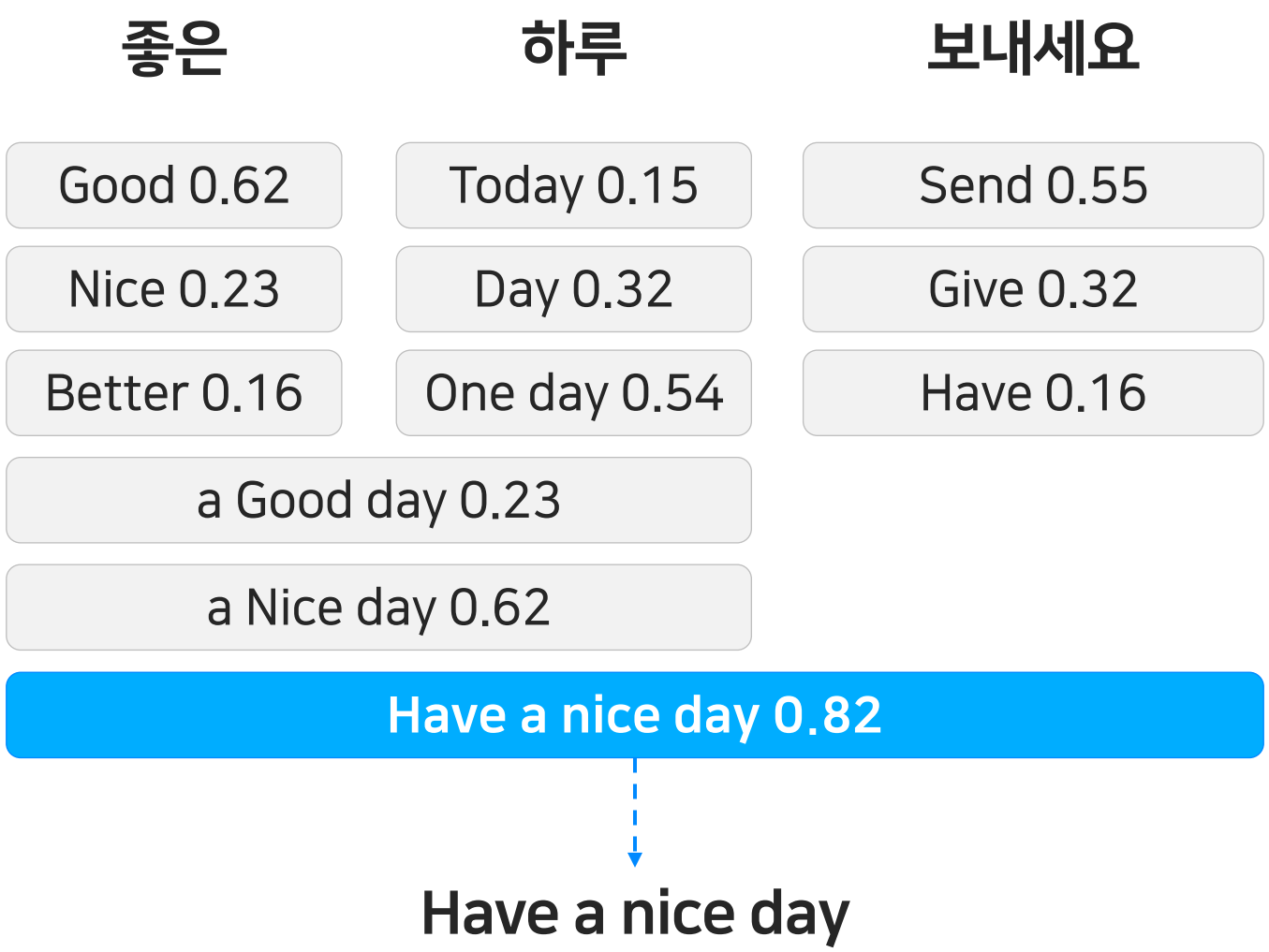
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



2. 기계 번역 **기술 발전의 흐름**

papago

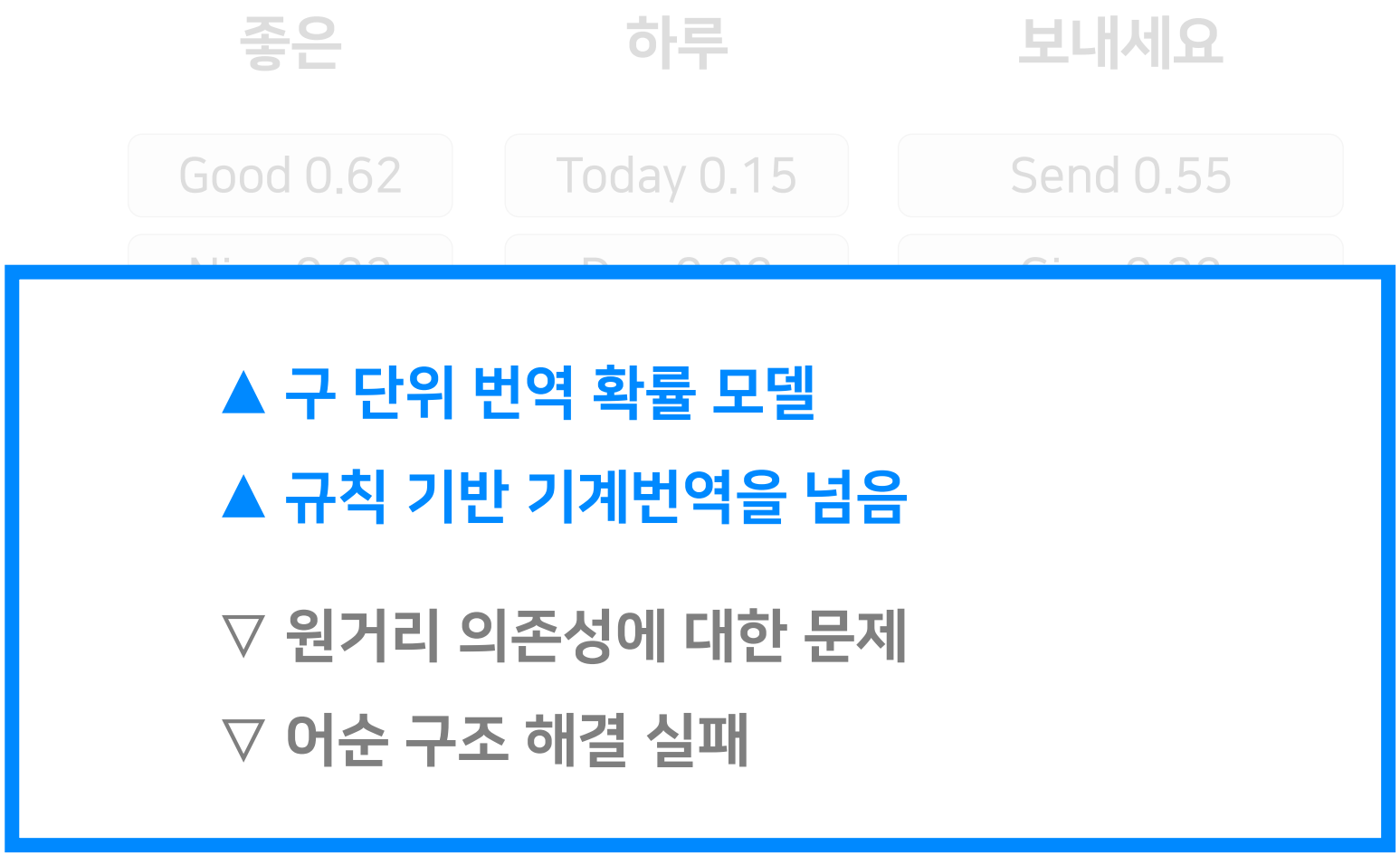
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



▲ 구 단위 번역 확률 모델

▲ 규칙 기반 기계번역을 넘음

▽ 원거리 의존성에 대한 문제

▽ 어순 구조 해결 실패

2. 기계 번역 **기술 발전의 흐름**

papago

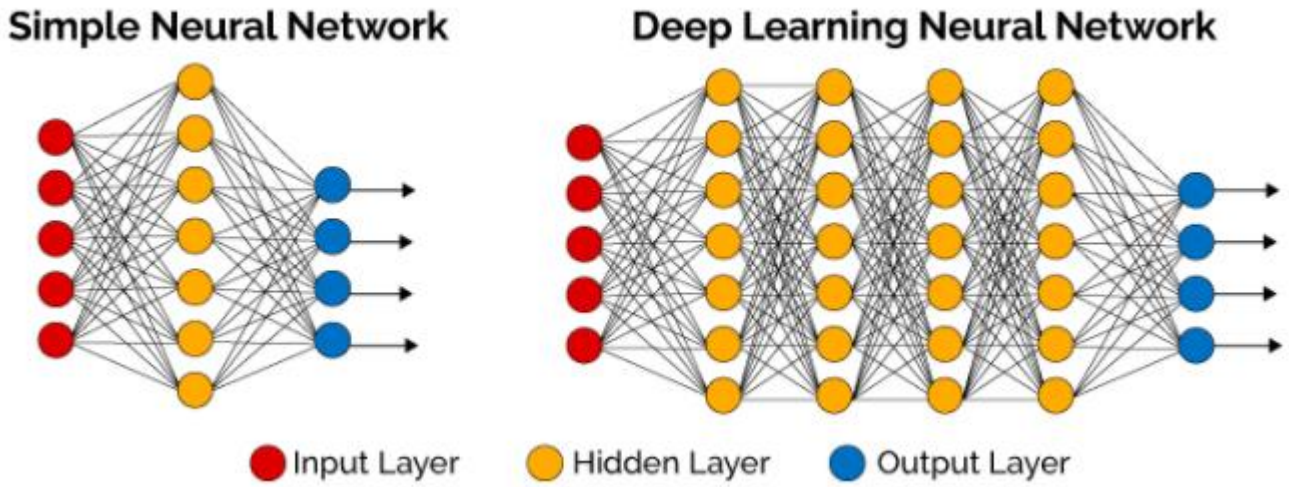
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



2. 기계 번역 **기술 발전의 흐름**

papago

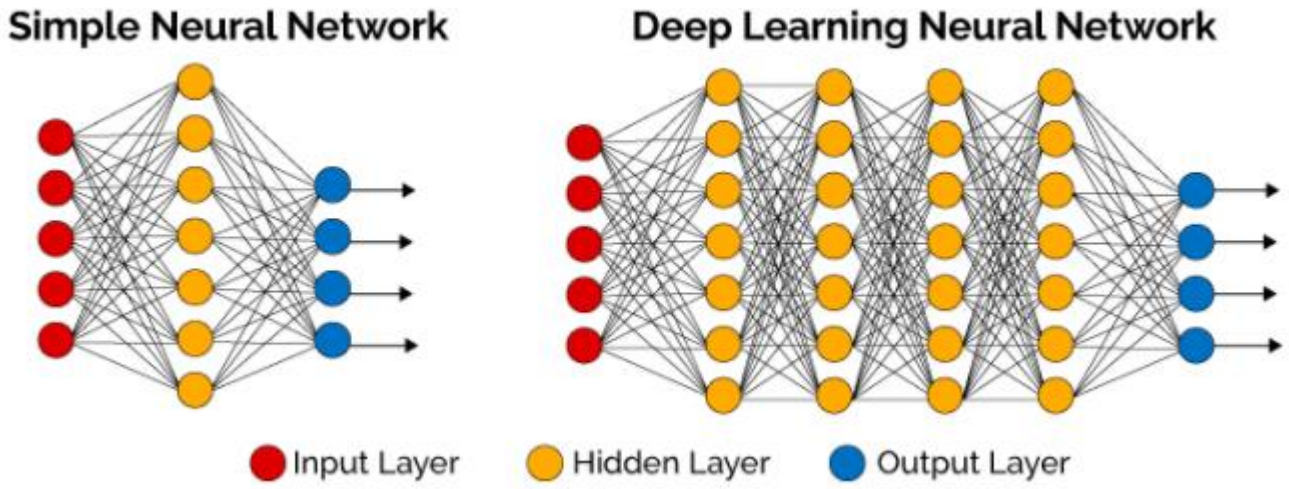
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



AlphaGo

2. 기계 번역 **기술 발전의 흐름**

papago

1949 Rule based Machine Translation
규칙 기반의 기계번역

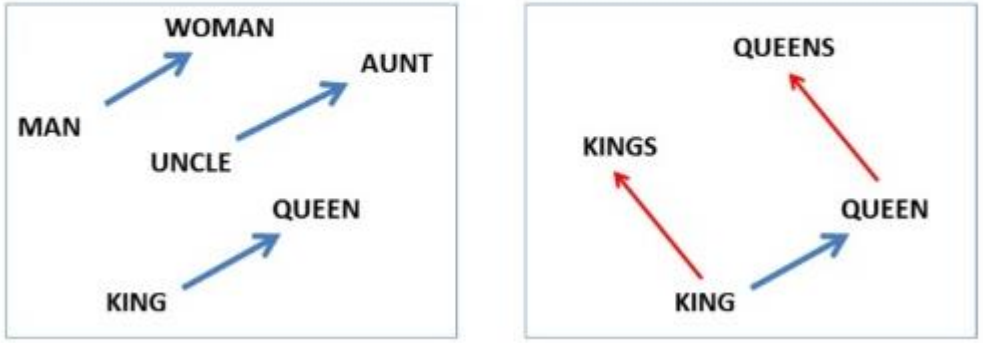
1993 Word based Statistical MT
단어 기반의 통계적 기계번역

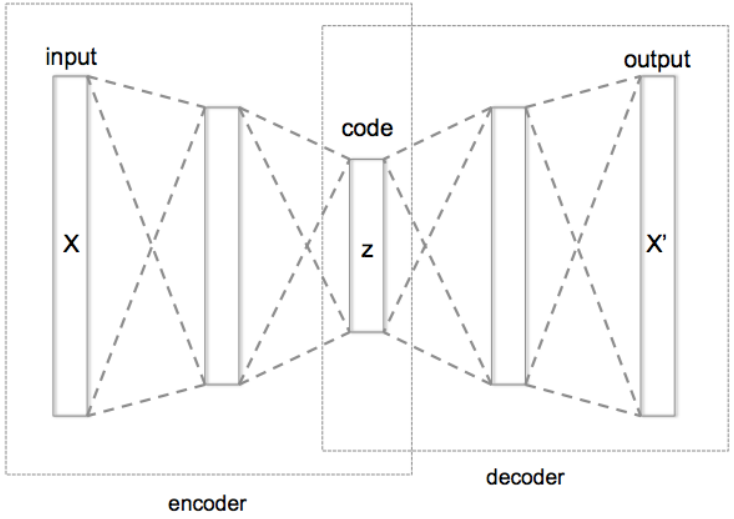
2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형

vec("man") – vec("king") + vec("woman") = vec("queen")





NAVER

2. 기계 번역 **기술 발전의 흐름**

papago

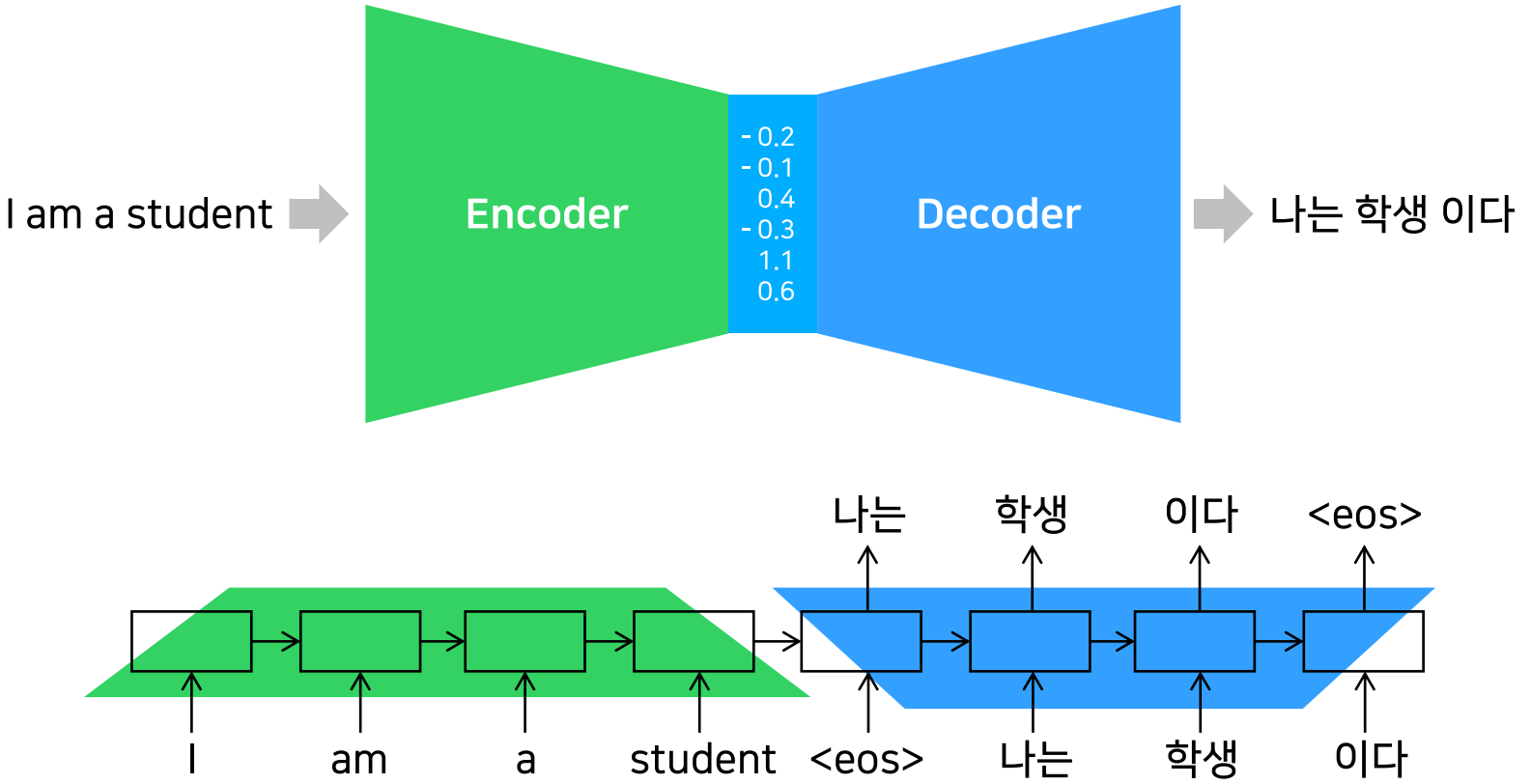
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



2. 기계 번역 **기술 발전의 흐름**

papago

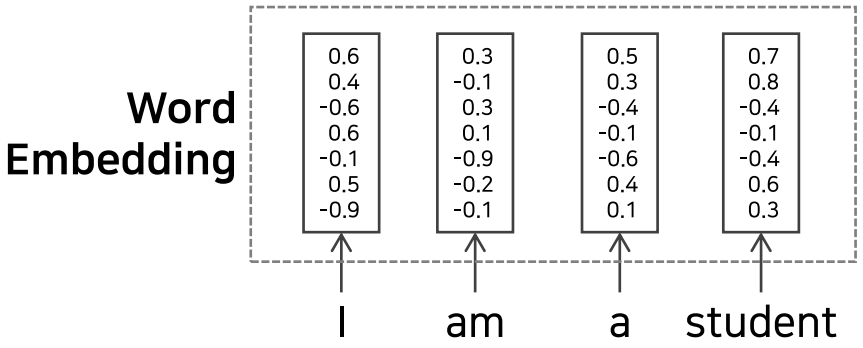
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



2. 기계 번역 **기술 발전의 흐름**

papago

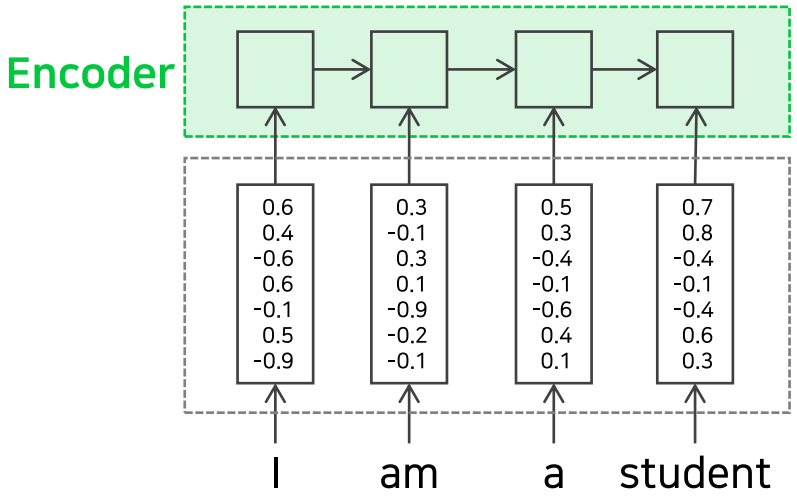
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



2. 기계 번역 **기술 발전의 흐름**

papago

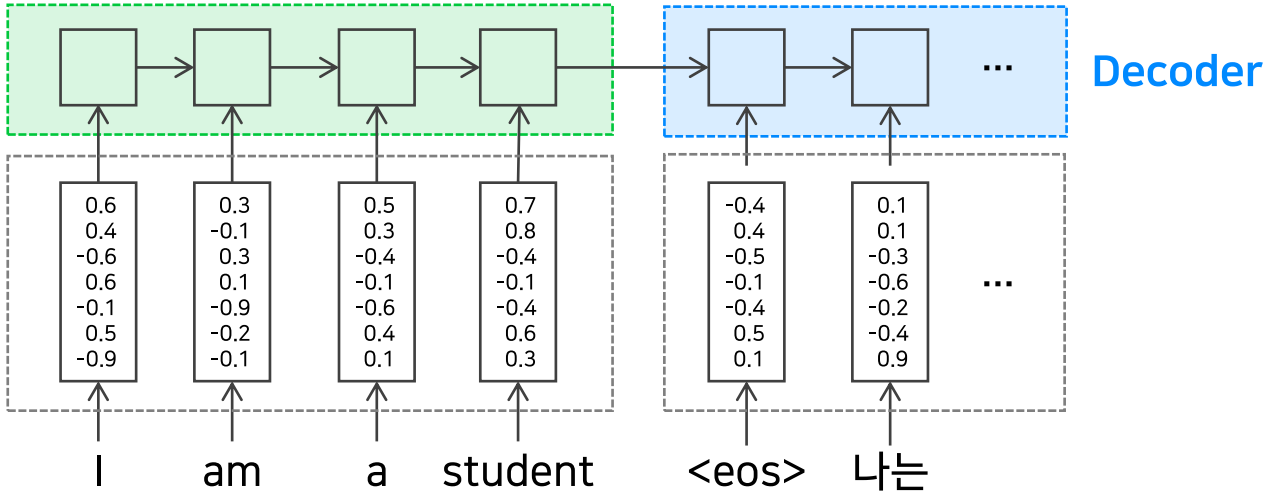
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



2. 기계 번역 **기술 발전의 흐름**

papago

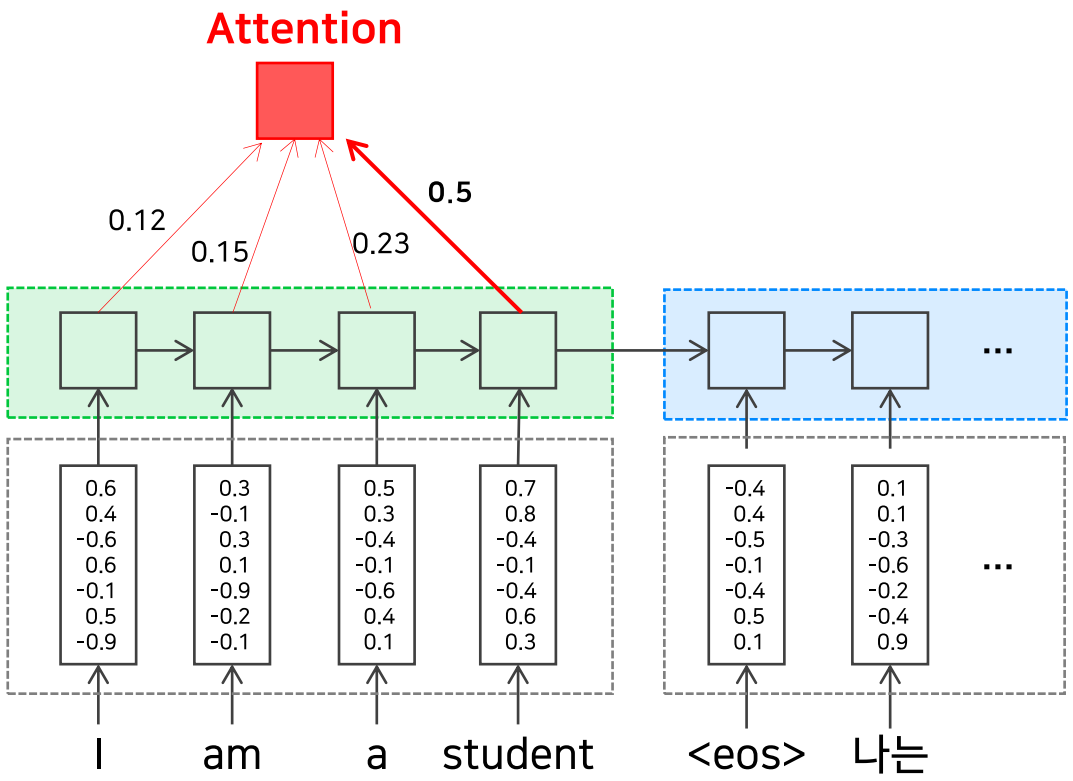
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



NAVER

2. 기계 번역 **기술 발전의 흐름**

papago

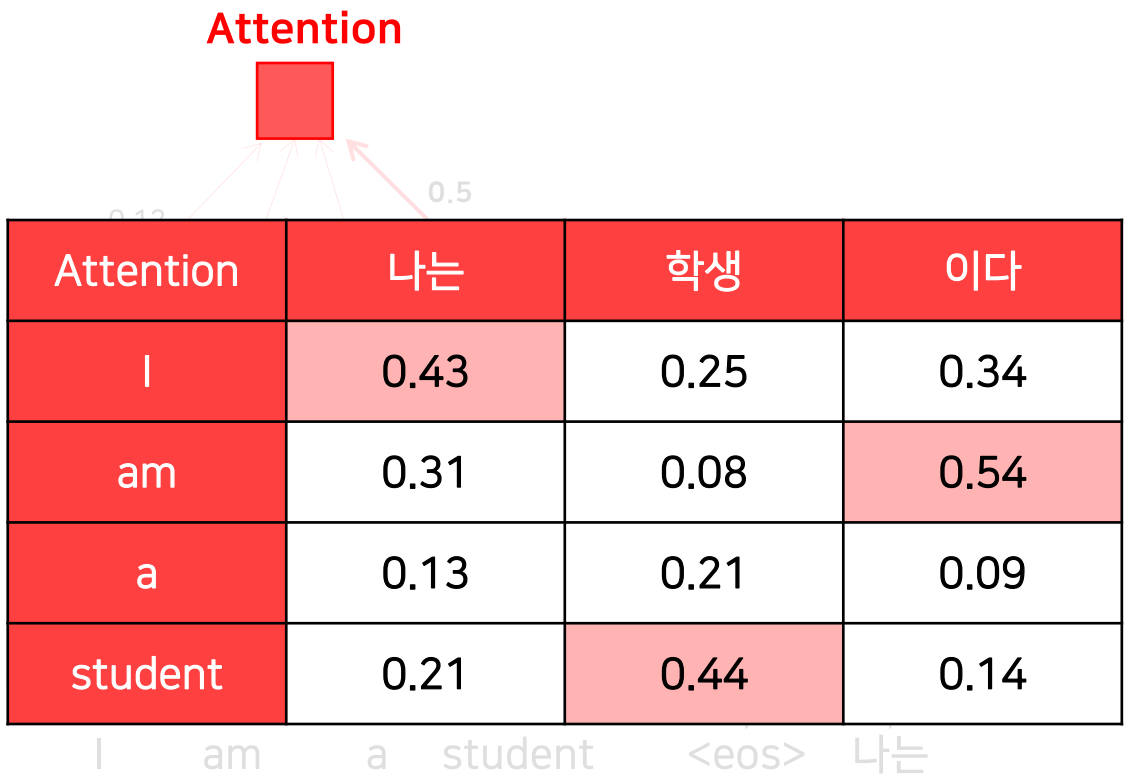
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



Attention	나는	학생	이다
I	0.43	0.25	0.34
am	0.31	0.08	0.54
a	0.13	0.21	0.09
student	0.21	0.44	0.14

2. 기계 번역 **기술 발전의 흐름**

papago

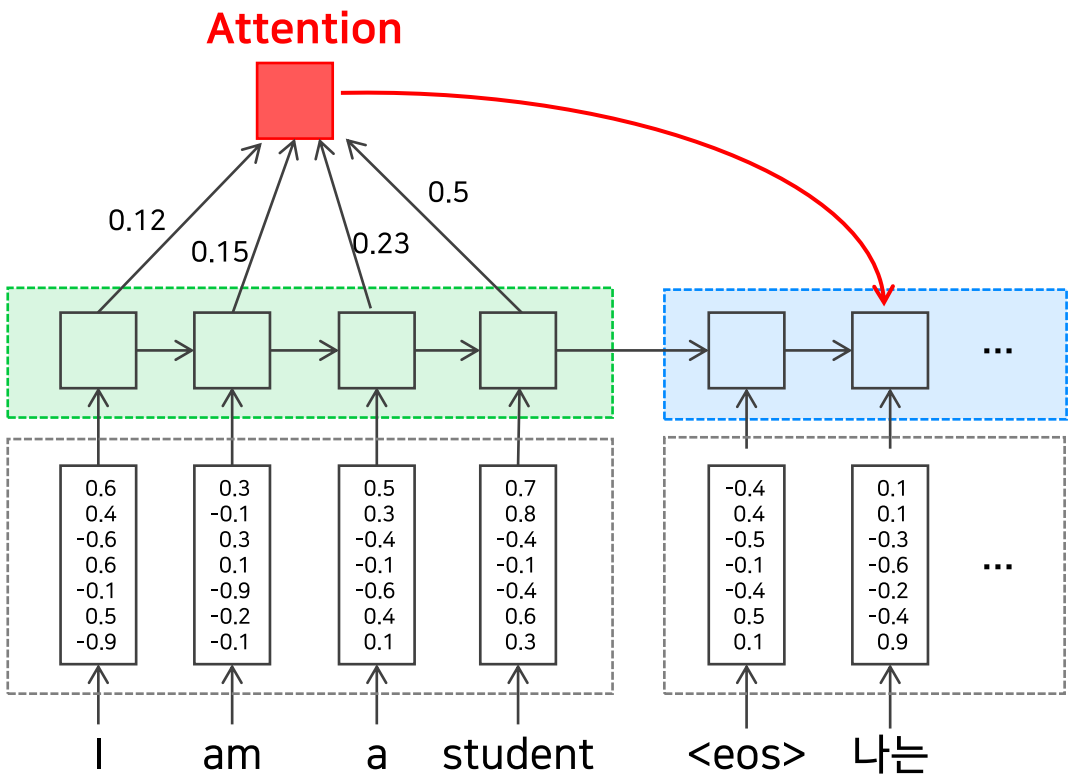
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



NAVER

2. 기계 번역 **기술 발전의 흐름**

papago

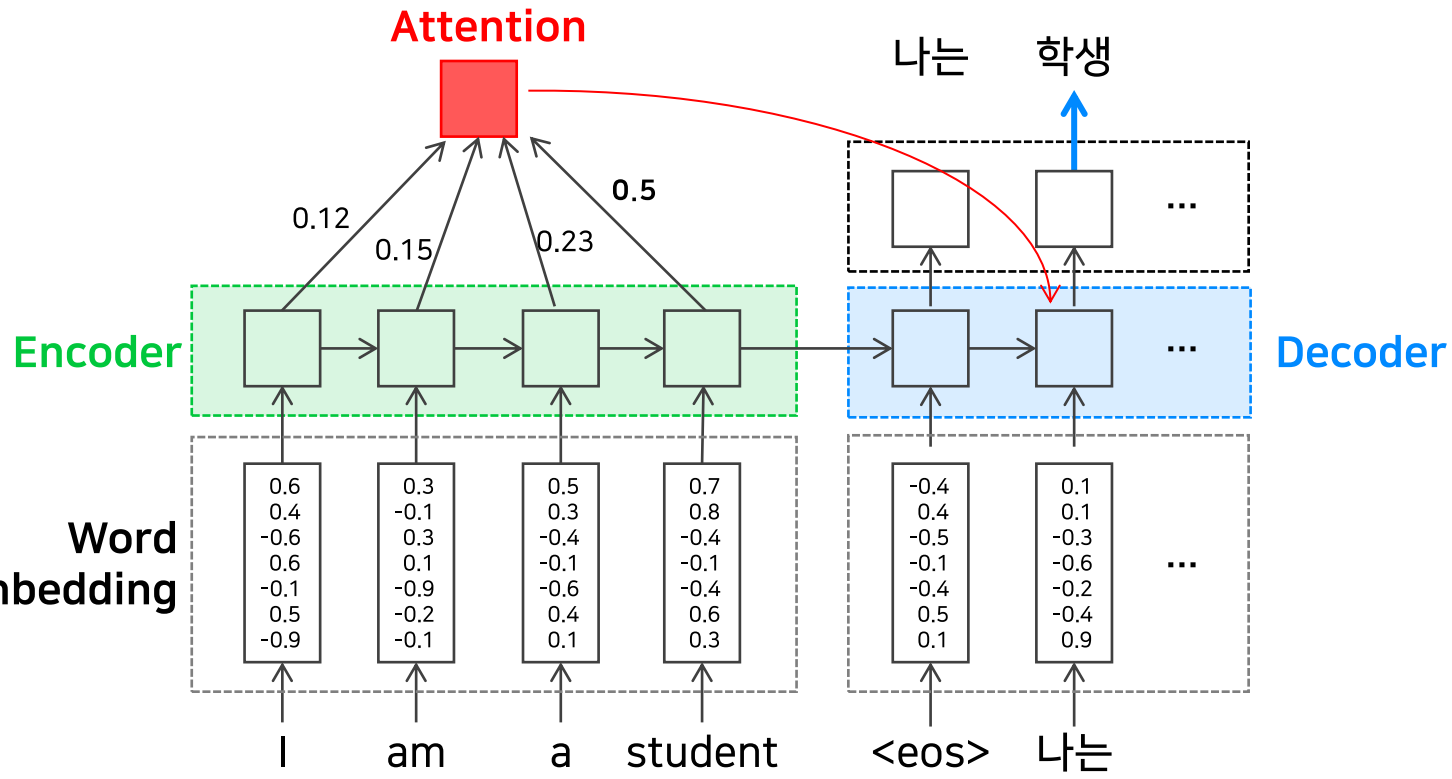
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



NAVER

2. 기계 번역 **기술 발전의 흐름**

papago

1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형

▲ 문장 단위 번역기 개발

▲ 문장의 어순, 의미를 잘 표현함

▲ SMT 대비 2배 이상의 품질 향상

▽ 번역 생략 문제 발생

▽ 표현 가능한 단어 수 제약

▽ 고비용 고사양 장비가 필요함

2. 기계 번역 **기술 발전의 흐름**

papago

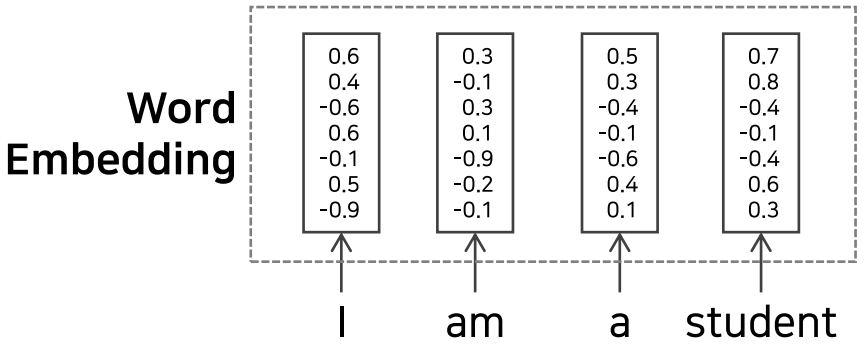
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



2. 기계 번역 **기술 발전의 흐름**

papago

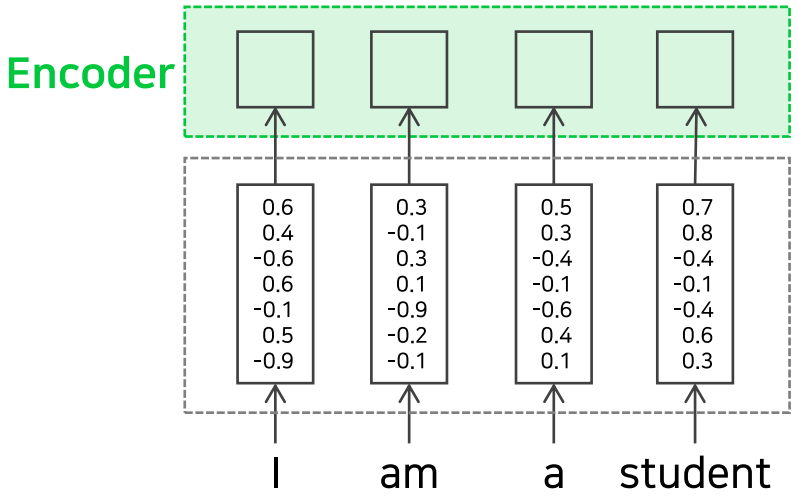
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



2. 기계 번역 **기술 발전의 흐름**

papago

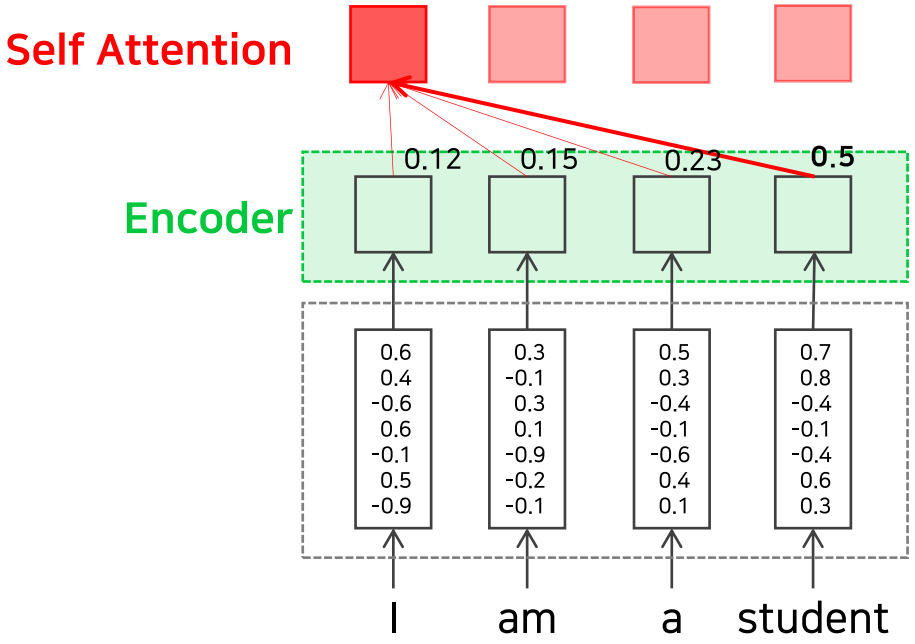
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



2. 기계 번역 **기술 발전의 흐름**

papago

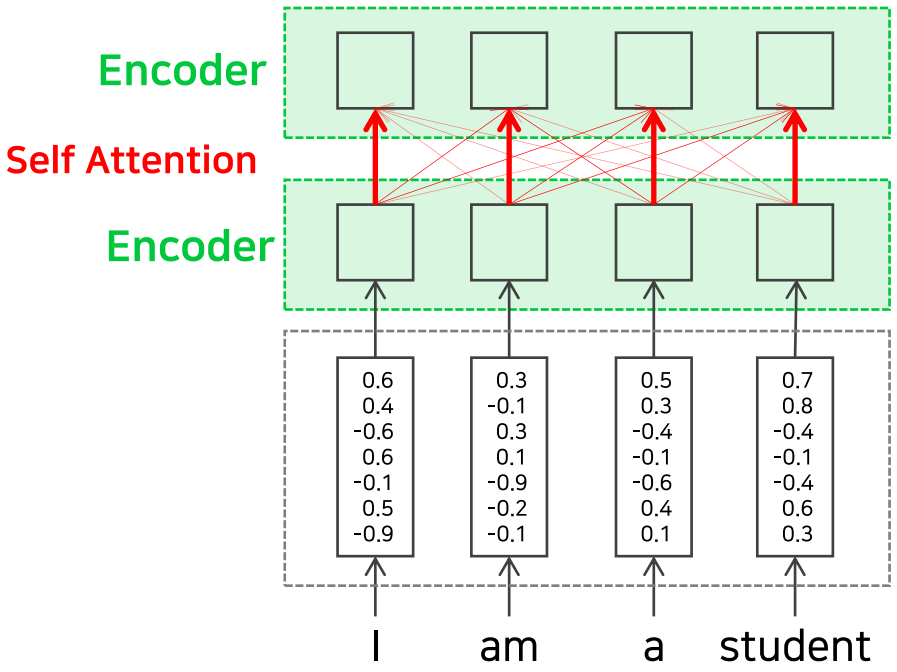
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



2. 기계 번역 **기술 발전의 흐름**

papago

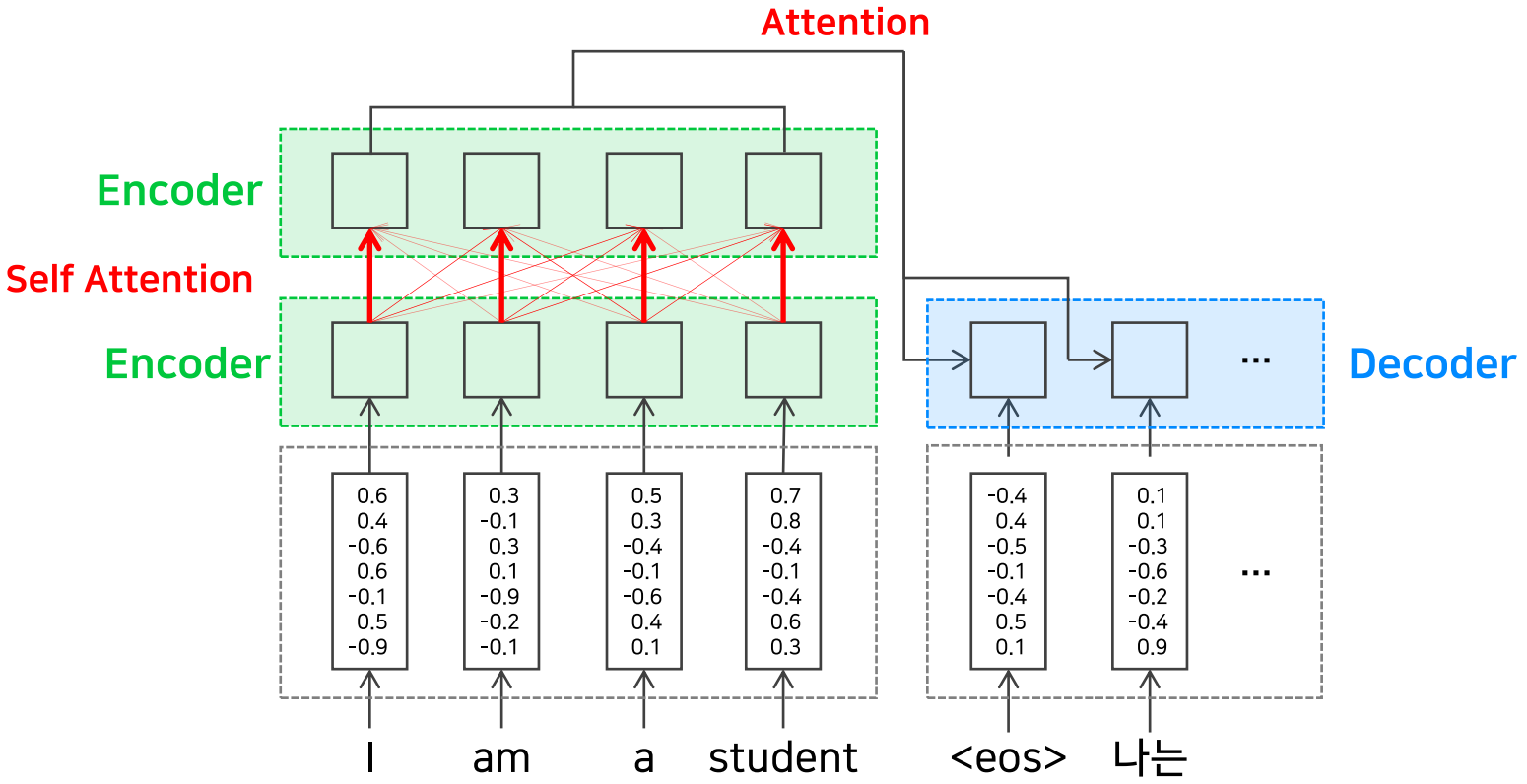
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



2. 기계 번역 **기술 발전의 흐름**

papago

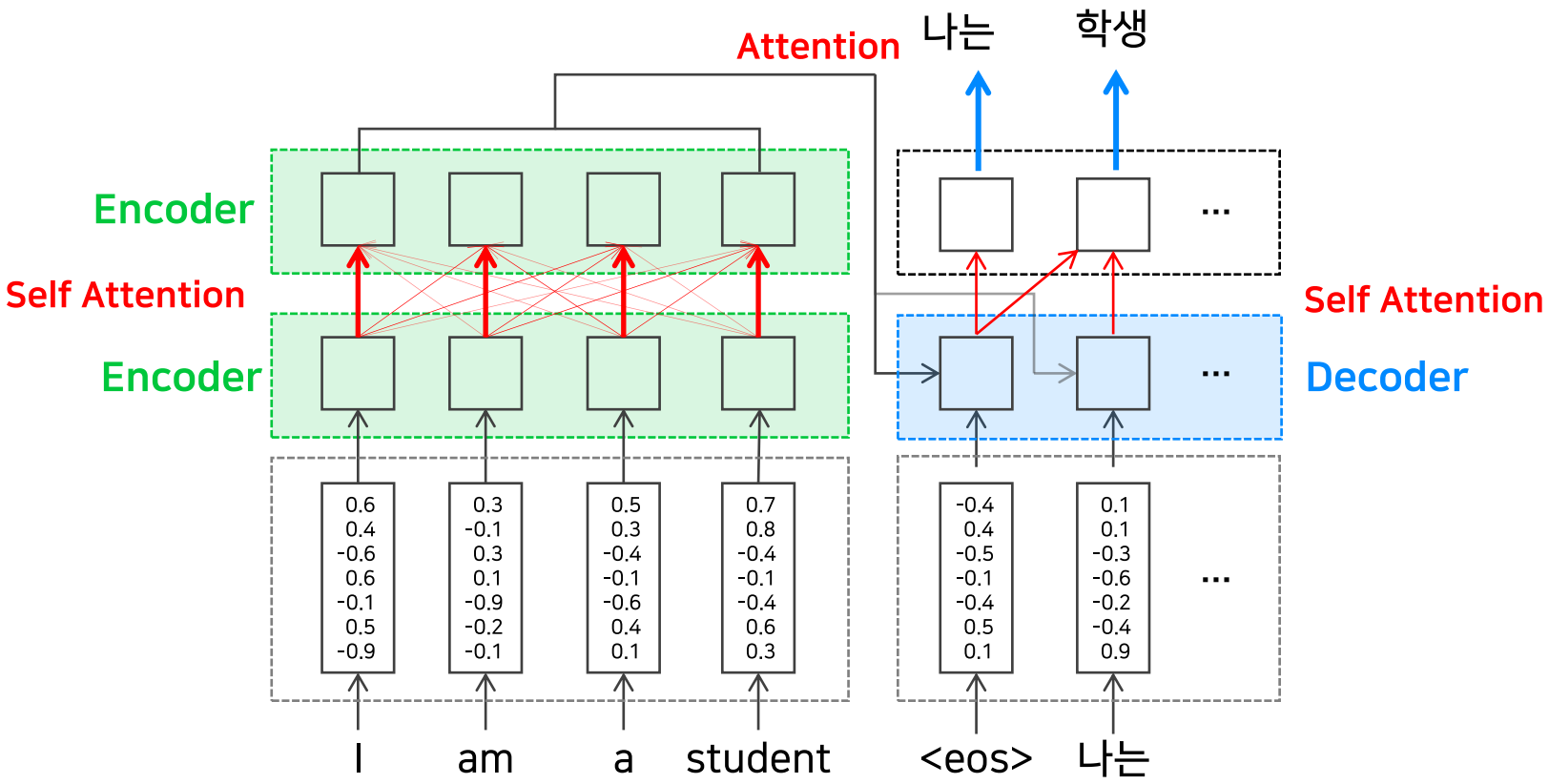
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



2. 기계 번역 **기술 발전의 흐름**

papago

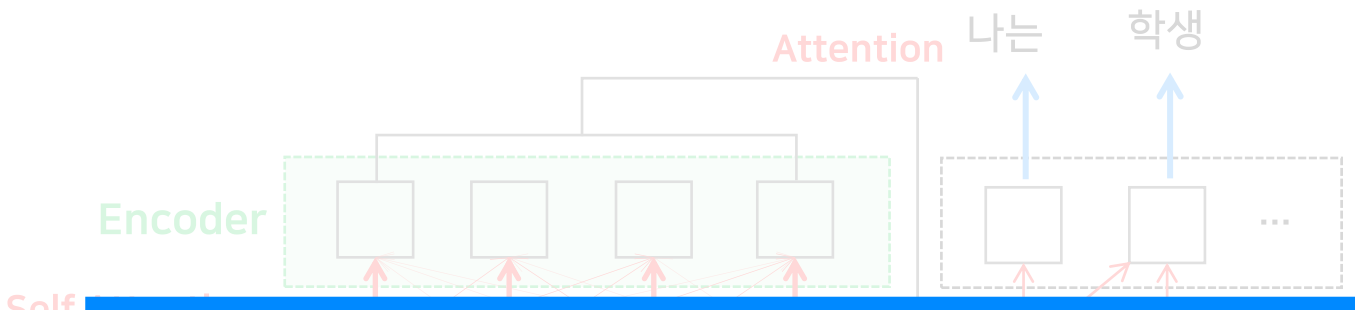
1949 Rule based Machine Translation
규칙 기반의 기계번역

1993 Word based Statistical MT
단어 기반의 통계적 기계번역

2003 Phrase based Statistical MT
구 기반의 통계적 기계번역

2014 Neural Machine Translation
인공신경망 기계번역

2017 NMT Transformer
인공신경망 기계번역의 발전형



▲ 학습 속도 개선, 학습 안정화

▲ 문장의 어순, 의미에 대한 표현력 개선

▽ 기존 NMT와 동일한 이슈 : 생략, 고유명사

▽ 고비용 고사양 장비가 필요함



papago

3. 오번역의 **원인과 대응**



오번역 사례 1)

日本は先進国だ。

> 우리나라는 선진국이다.

원인: 자동 수집된 의역 데이터

오번역 사례 2)

Mr. Trump…

> 트럼프 회장

원인: 시간에 따라 변화하는 호칭

원인은 학습 데이터

- 대용량 데이터에는 다양한 번역문이 존재함
- 의역과 시의성은 문제를 양산함
  - 자동수집 의역 데이터 : 日本は先進国だ > 우리나라는 선진국이다.
  - 시간에 따른 호칭변화 : Mr. Trump 회장, 방송인, 대통령, 전 대통령
- 고유명사는 처리가 어려움
  - 고유명사의 대명사화 : Park scored., 박지성 선수가 득점했다.

대응

- 오류 이슈는 미리 예측하기 어려운 영역
- 인공신경망 번역기는 방법론 자체로는 임의의 번역 수정이 매우 어려움
- 일반적 품질 저하 문제는 모델 품질 향상으로 대응
- 최신 데이터 확보를 위해 노력

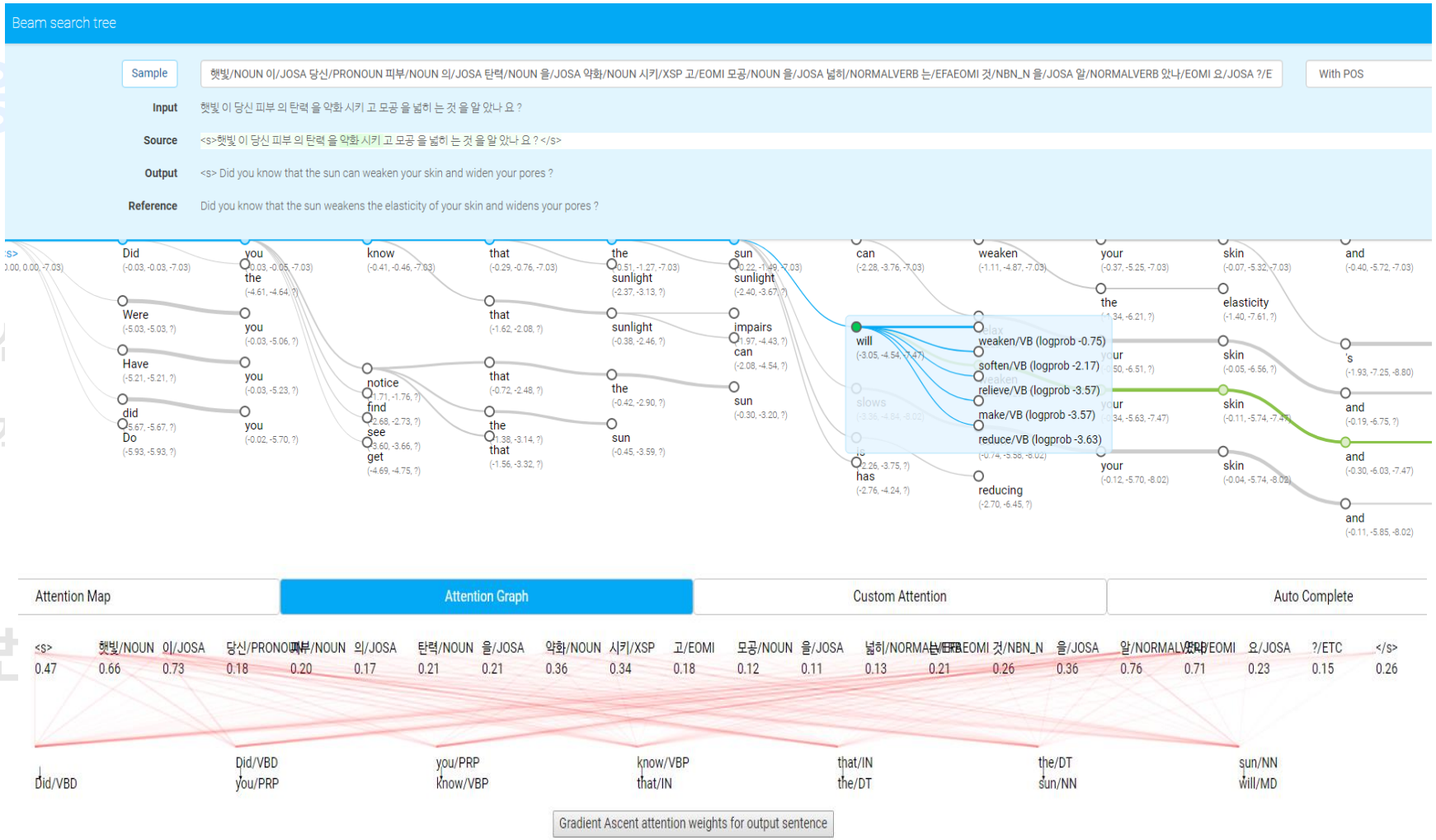
대응

- 오류 이슈는 미리 예측하기 어려운 영역
- 인공신경망 번역기는 방법론 자체로는 임의의 번역 수정이 매우 어려움
- 일반적 품질 저하 문제는 모델 품질 향상으로 대응
- 최신 데이터 확보를 위해 노력

- 오번역 원인 파악을 위한 연구

3. 오번역의 **원인과 대응**

papago



대응
- 오류
- 인공
- 일반
- 최신
- 오번



4. 향후 연구 **방향**



문맥 기반 번역 / 문서 단위 번역

- 현재, **문서 단위 번역은 문장 단위 번역 결과의 조합**
- 문장을 넘나드는 문맥 번역이 어려움
  - [원문] 진아는 여행을 좋아한다. 그녀는 발이 넓다.
  - [번역] Jina loves traveling. She has a wide feet. (She knows a lot of people)
- 일관된 고유명사 처리가 어려움
  - [원문] When Park was here, Park scored.
  - [번역] 공원이 여기에 있을때에, 박지성 선수가 득점했다.

다국어 번역 모델 (Multilingual)

- **번역 데이터 부족은 기계번역의 고질적인 문제**
- 상대적으로 풍부한 언어쌍의 번역 데이터(Bilingual)를 활용하는 방법
- [예시] 한국어 -> 베트남어 번역기 학습을 위한 다국어 번역 모델
  - 한국어 -> 영어 데이터 (한국어 데이터 보충)
  - 영어 -> 베트남어 데이터 (베트남어 데이터 보충)

실시간 번역

- 현재 기계번역은 문장 전체를 입력 받은 후에, 번역을 수행하는 구조

- 입력 문장의 일부를 기반으로 번역을 시작하는 모델

  - [원문] We would like to introduce you to ..

  - [번역] 여러분에게 소개시켜드리고 싶은 것은 ..

- 다국어 컨퍼런스, 외국인과의 실시간 대화, 강의 발표 등에 적용 가능한 기술



papago

Q & A





papago

thanks

