NOVEMBER 26, 2019
ANTONIS ANASTASOPOULOS
MATERIALS LARGELY BORROWED FROM JUNJIE HU AND AUSTIN MATTHEWS

# MACHINE TRANSLATION OVERVIEW

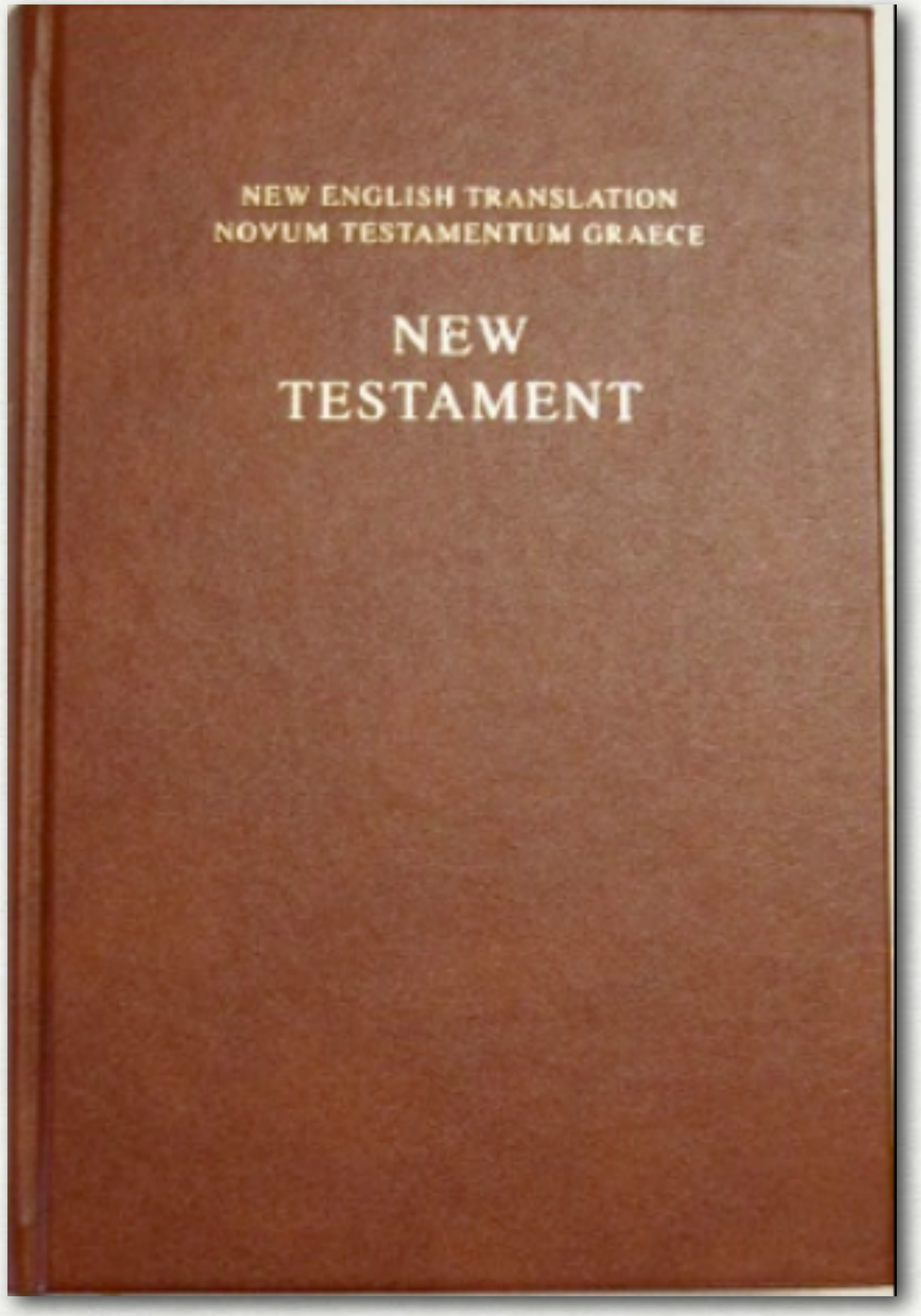
- One naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: ***‘This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.’***

Warren Weaver to Norbert Wiener, March, 1947

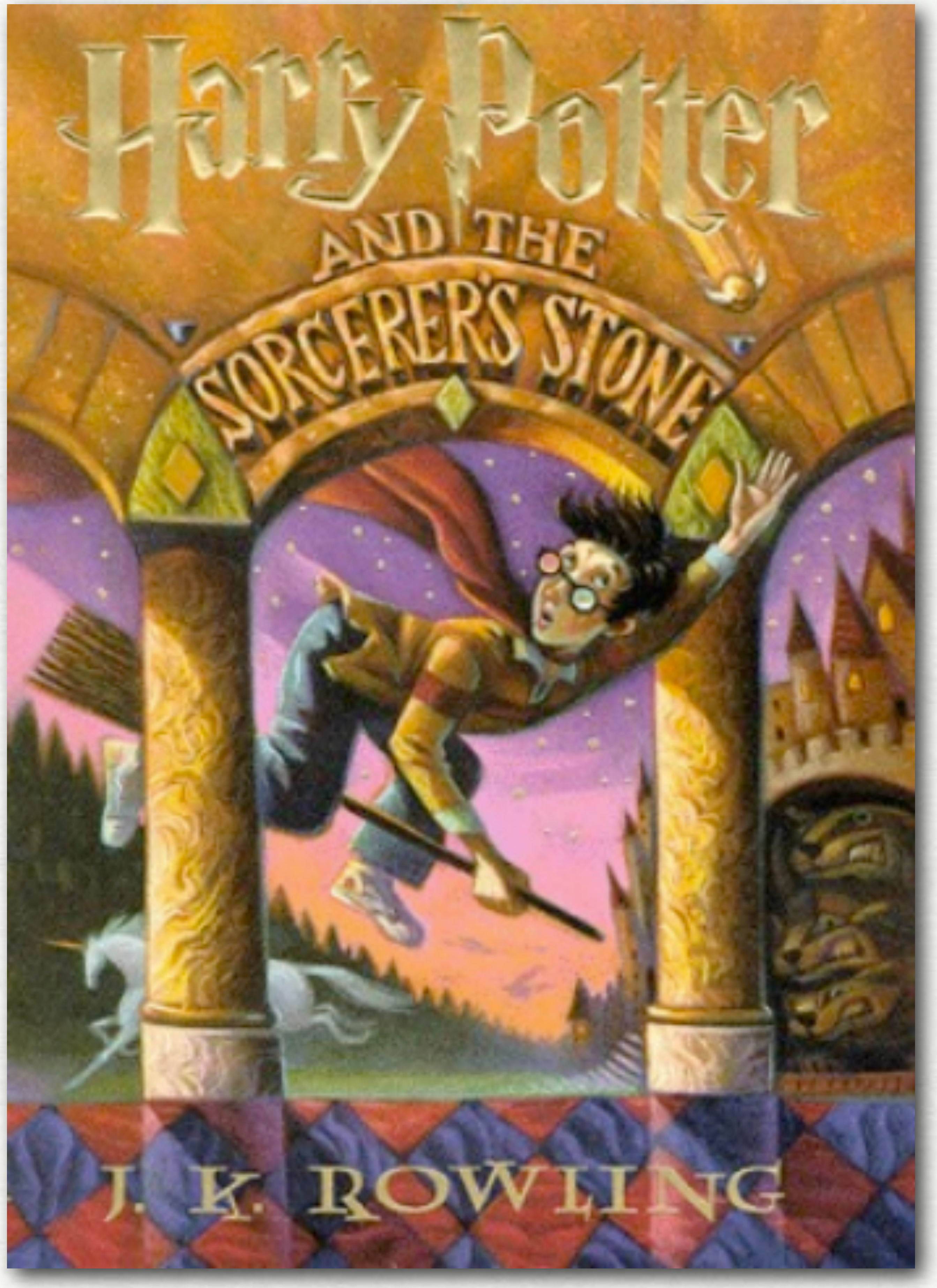

ORDER YOUR
KAWHE/COFFEE
IN MĀORI
He mōwai māku I'll have a flat white
He pango poto māku I'll have a short black
He pango roa māku I'll have a long black
He rate pīni māku I'll have a soy latte
He kaputino māku I'll have a cappuccino
He rate māku I'll have a latte
He tiakarete wera māku I'll have a hot chocolate
Rahi Size
Kei te pēhea koe?
How's it going?
Anei taku kapu mau tonu
Here is my reusable cup
(S) Paku
(M) Waenga
(L) Nui
Hei kawe atu
To take away
Ki konei
To have here
McCafé

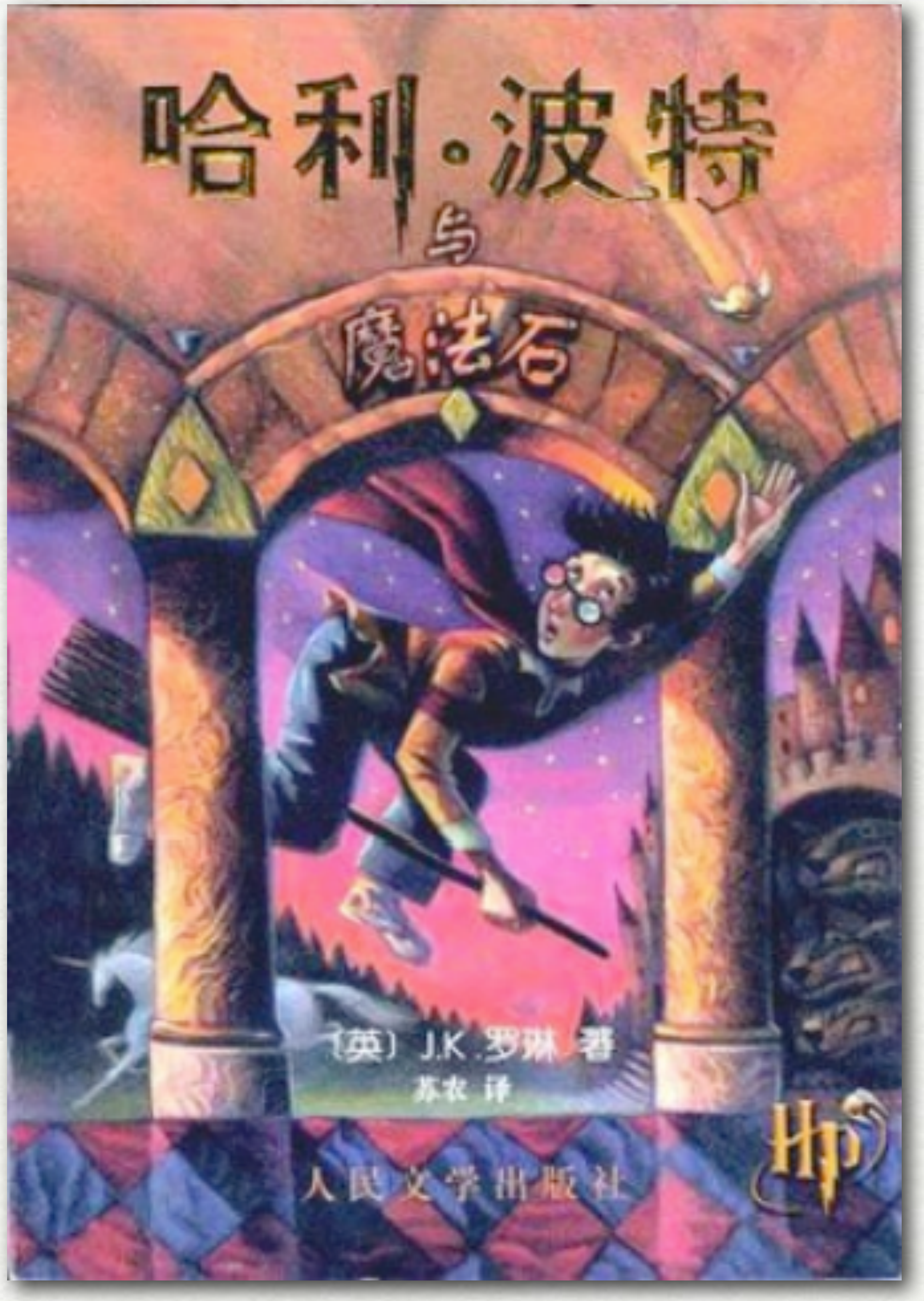

www.un.org
http://www.un.org/english/
Google
We the peoples
Daily Briefing | Radio, TV, Photo | Documents, Maps | Publications, Stamps, Databases | UN Works | Search
Peace & Security | Economic & Social Development | Human Rights | Humanitarian Affairs | International Law
Welcome to the
United Nations
UN Millennium Development Goals
United Nations News Centre
About the United Nations
Main Bodies
Conferences & Events
Member States
General Assembly President
8 September 2005 >>
Secretary-General
Situation in Iraq
Mideast Roadmap
Renewing the UN
UN Action against Terrorism
Issues on the UN Agenda
Civil Society / Business
UN Webcast
CyberSchoolBus
Home | Recent Additions | Employment | UN Procurement | Comments | Q & A | UN System Sites | Index
عربي | 中文 | English | Français | Русский | Español
Copyright, United Nations, 2000-2005 | Use of UN60 Logo | Terms of Use | Privacy Notice | Help
[ Text version ]
Live and On-Demand Webcasts, 24 Hours a Day: Click on UN Webcas


联合国主页
http://www.un.org/chinese/
Google
我们人民
每日简报 | 多媒体 | 文件与地图 | 出版物 邮票 数据库 | 服务全球 | 网址搜索
和平与安全 | 经济与社会发展 | 人权 | 人道主义事务 | 国际法
欢 迎 来 到
联 合 国
联合国千年发展目标
联合国新闻
联合国概况
联合国主要机关
会议与活动
联合国会员国
联合国大会主席
联大第60届会议一般性辩论
联合国秘书长
伊拉克局势
中东路线图
更新联合国
反恐怖主义
联合国日常议题
民间团体/商业
联合国网络直播
空中校车
新增内容 | 工作机会 | 联合国采购 | 建议 | 问题与解答 | 其他网址 | 网址索引
عربي | 中文 | English | Français | Русский | Español
联合国2000-2005年版权|联合国60周年徽标使用准则|使用条件|隐私通告|帮助
[纯文字版]


NEW ENGLISH TRANSLATION
NOVUM TESTAMENTUM GRAECE
NEW
TESTAMENT


聖經
導讀本


Harry Potter
AND THE
SORCERER'S STONE
J. K. ROWLING

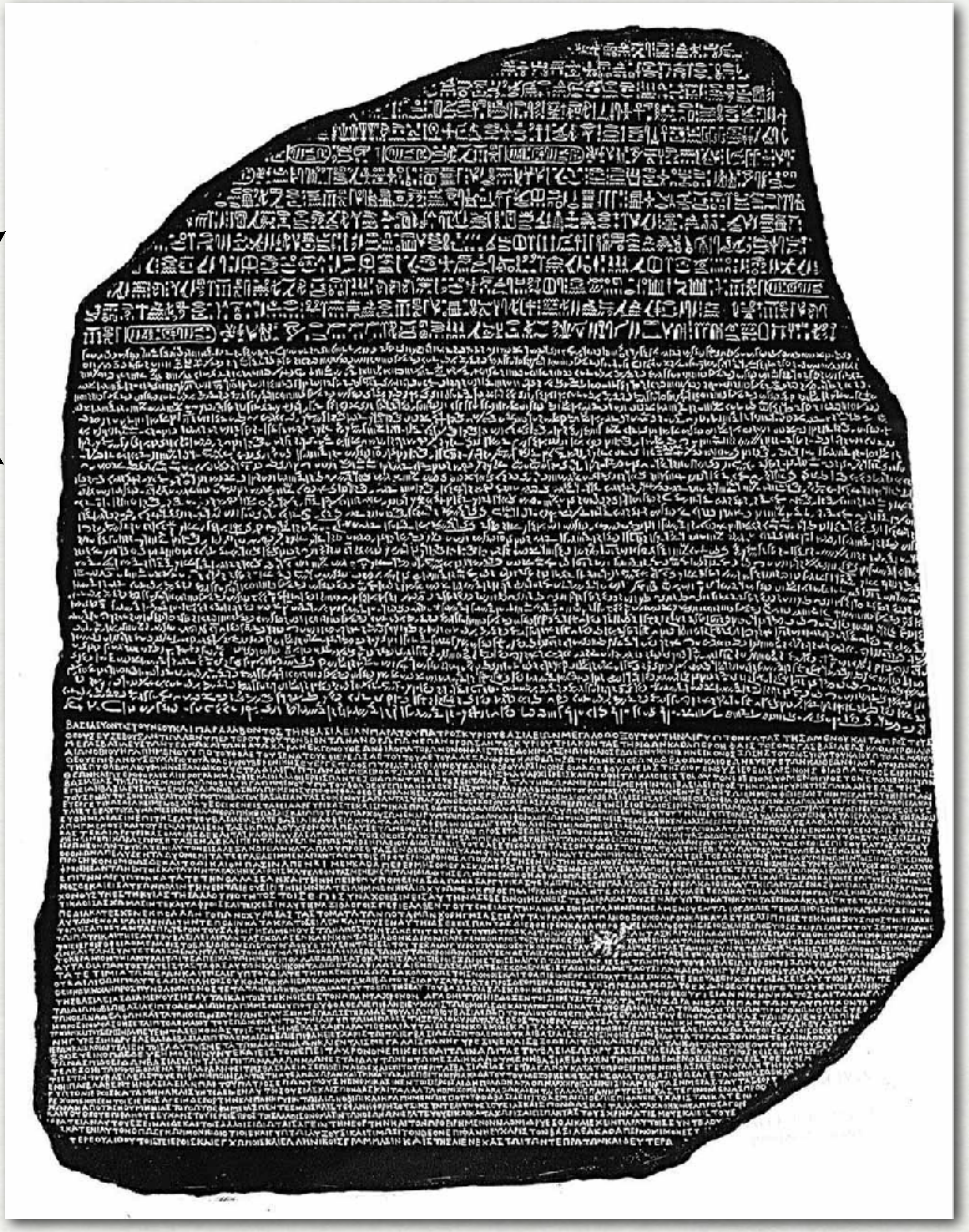

哈利·波特
与
魔法石
(英) J.K.罗琳 著
苏农 译
人民文学出版社

# CLASSIC SOUPS

| | No. | | Sm. | Lg. |
|---|---|---|---|---|
| 清燉雞湯 | 57. | House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot) | 1.50 | 2.75 |
| 雞飯湯 | 58. | Chicken Rice Soup | 1.85 | 3.25 |
| 雞麵湯 | 59. | Chicken Noodle Soup | 1.85 | 3.25 |
| 廣東雲吞 | 60. | Cantonese Wonton Soup | 1.50 | 2.75 |
| 蕃茄蛋湯 | 61. | Tomato Clear Egg Drop Soup | 1.65 | 2.95 |
| 雲吞湯 | 62. | Regular Wonton Soup | 1.10 | 2.10 |
| 酸辣湯 | 63. | Hot & Sour Soup | 1.10 | 2.10 |
| 蛋花湯 | 64. | Egg Drop Soup | 1.10 | 2.10 |
| 雲蛋湯 | 65. | Egg Drop Wonton Mix | 1.10 | 2.10 |
| 豆腐菜湯 | 66. | Tofu Vegetable Soup | NA | 3.50 |
| 雞玉米湯 | 67. | Chicken Corn Cream Soup | NA | 3.50 |
| 蟹肉玉米湯 | 68. | Crab Meat Corn Cream Soup | NA | 3.50 |
| 海鮮湯 | 69. | Seafood Soup | NA | 3.50 |

Egyptian

Greek

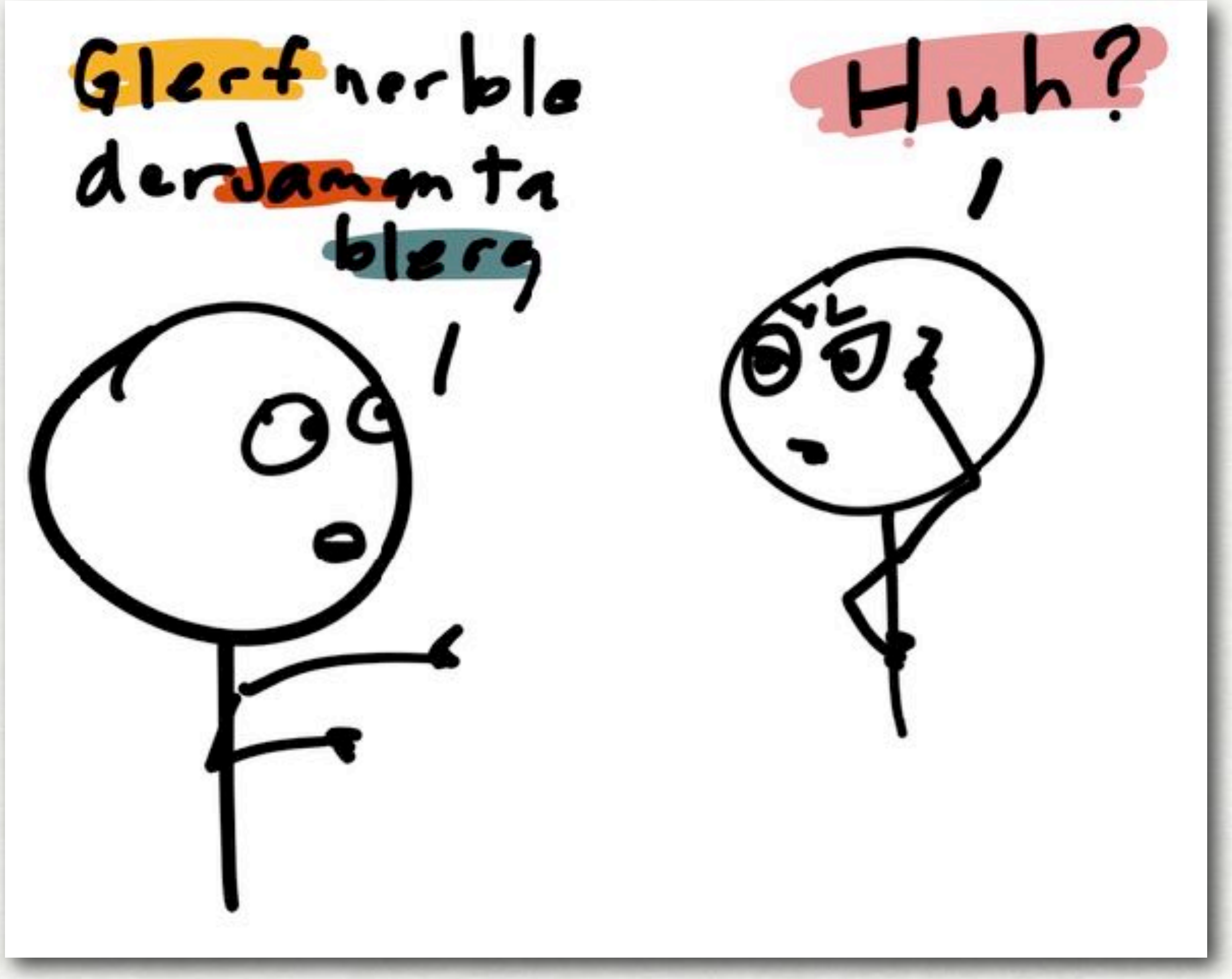


# NOISY CHANNEL MT

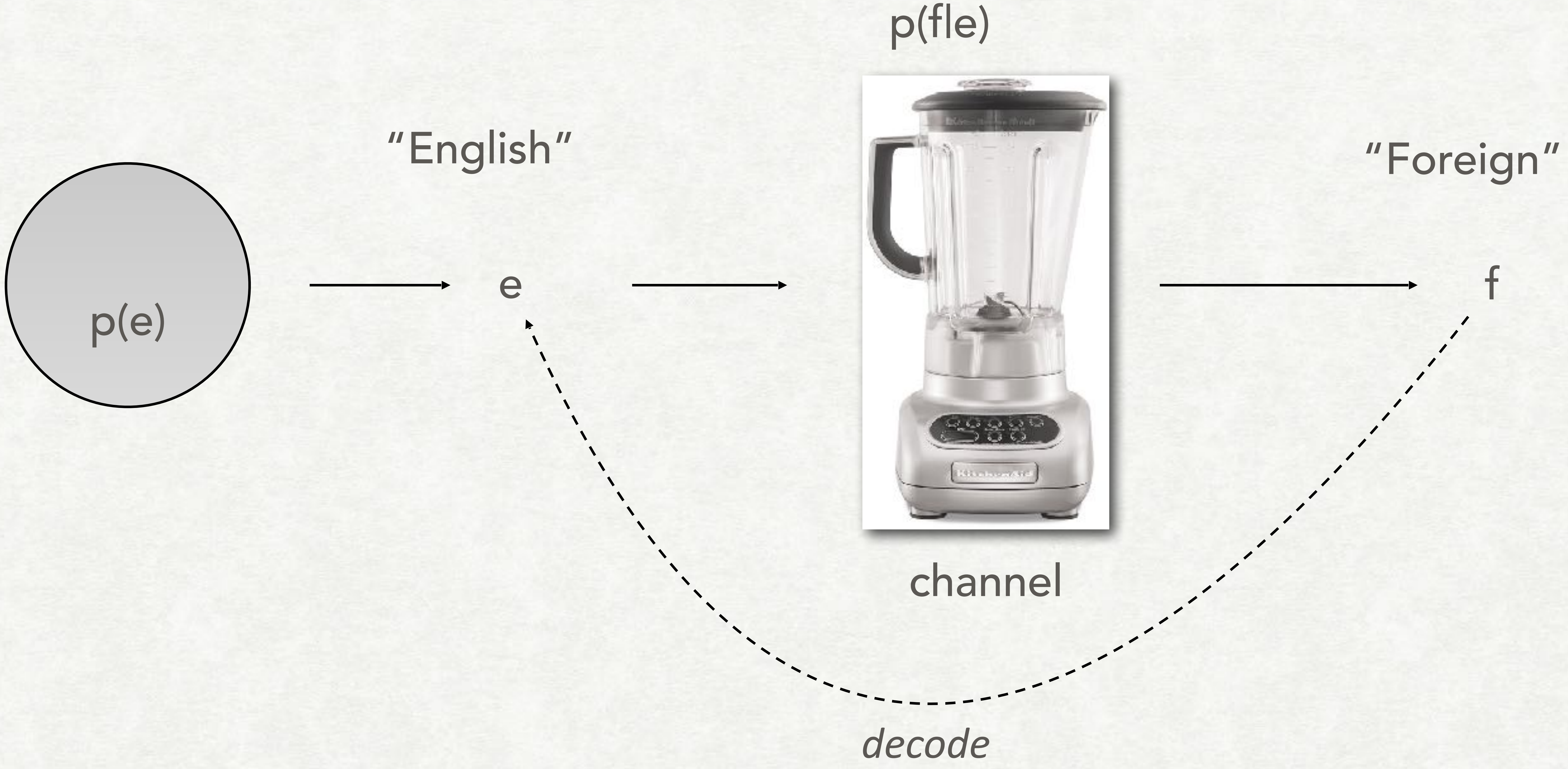
- We want a model of $p(\boxed{\mathbf{e}} \mid \boxed{\mathbf{f}})$

Confusing foreign sentence

Possible English translation



# NOISY CHANNEL MT



# NOISY CHANNEL MT

$$\hat{e} = \arg\max_e p(e|f)$$

$$= \arg\max_e \frac{p(e) \times p(f|e)}{p(f)}$$

$$= \arg\max_e \boxed{p(e)} \times \boxed{p(f|e)}$$

"Language Model" "Translation Model"

# NOISY CHANNEL DIVISION OF LABOR

- Language model – $p(\mathbf{e})$
  - is the translation fluent, grammatical, and idiomatic?
  - use any model of $p(\mathbf{e})$ – typically an $n$-gram model
- Translation model – $p(\mathbf{f}|\mathbf{e})$
  - "reverse" translation probability
  - ensures adequacy of translation

# LANGUAGE MODEL FAILURE


*My legal name is Alexander Perchov.*

# LANGUAGE MODEL FAILURE


*My legal name is Alexander Perchov. But all of my many friends dub me Alex, because that is a more flaccid-to-utter version of my legal name. Mother dubs me Alexi-stop-spleening-me!, because I am always spleening her.*

# LANGUAGE MODEL FAILURE


*My legal name is Alexander Perchov. But all of my many friends dub me Alex, because that is a more flaccid-to-utter version of my legal name. Mother dubs me Alexi-stop-spleening-me!, because I am always spleening her. If you want to know why I am always spleening her, it is because I am always elsewhere with friends, and disseminating so much currency, and performing so many things that can spleen a mother.*

# TRANSLATION MODEL

*p*(***f***|***e***) gives the channel probability – the probability of translating an English sentence into a foreign sentence

***f*** = je voudrais un peu de frommage

| | |
|---|---|
| | *p*(***f***\|***e***) |
| $\mathbf{e}_1$ = I would like some cheese | 0.4 |
| $\mathbf{e}_2$ = I would like a little of cheese | 0.5 |
| $\mathbf{e}_3$ = There is no train to Barcelona | >0.00001 |

# TRANSLATION MODEL

- How do we parameterize *p(f|e)*?

$$p(f|e) = \frac{count(f, e)}{count(e)} \text{ ?}$$

- There are a lot of sentences: this won't generalize to new inputs

# LEXICAL TRANSLATION

How do we translate a word? Look it up in a dictionary!

*Haus: house, home, shell, household*

Multiple translations

Different word senses, different registers, different inflections

*house, home* are common

*shell* is specialized (the Haus of a snail is its shell)

# HOW COMMON IS EACH?

| Translation | Count |
| --- | --- |
| house | 5000 |
| home | 2000 |
| shell | 100 |
| household | 80 |

# MLE

$$\hat{p}_{\mathrm{MLE}}(e \mid \texttt{Haus}) = \begin{cases} 0.696 & \text{if } e = \texttt{house} \\ 0.279 & \text{if } e = \texttt{home} \\ 0.014 & \text{if } e = \texttt{shell} \\ 0.011 & \text{if } e = \texttt{household} \\ 0 & \text{otherwise} \end{cases}$$

# LEXICAL TRANSLATION

- Goal: a model $p(\mathbf{e}|\mathbf{f},m)$

- where **e** and **f** are complete English and Foreign sentences

$$\mathbf{e} = \langle e_1, e_2, \ldots, e_m \rangle \quad \mathbf{f} = \langle f_1, f_2, \ldots, f_n \rangle$$

# LEXICAL TRANSLATION

Goal: a model $p(\mathbf{e}|\mathbf{f},m)$

where **e** and **f** are complete English and Foreign sentences

Lexical translation makes the following ***assumptions***:

1. Each word $\mathbf{e}_i$ in **e** is generated from exactly one word in **f**
2. Thus, we have a latent *alignment* $\mathbf{a}_i$ that indicates which word $\mathbf{e}_i$ "came from." Specifically it came from $\mathbf{f}_{\mathbf{a}_i}$.
3. Given the alignments **a**, translation decisions are conditionally independent of each other and depend *only* on the aligned source word $\mathbf{f}_{\mathbf{a}_i}$.

# LEXICAL TRANSLATION

- Putting our assumptions together, we have:

$$p(\mathbf{e} \mid \mathbf{f}, m) = \underbrace{\sum_{\mathbf{a} \in [0,n]^m} p(\mathbf{a} \mid \mathbf{f}, m)}_{\text{p(Alignment)}} \times \underbrace{\prod_{i=1}^{m} p(e_i \mid f_{a_i})}_{\text{p(Translation | Alignment)}}$$

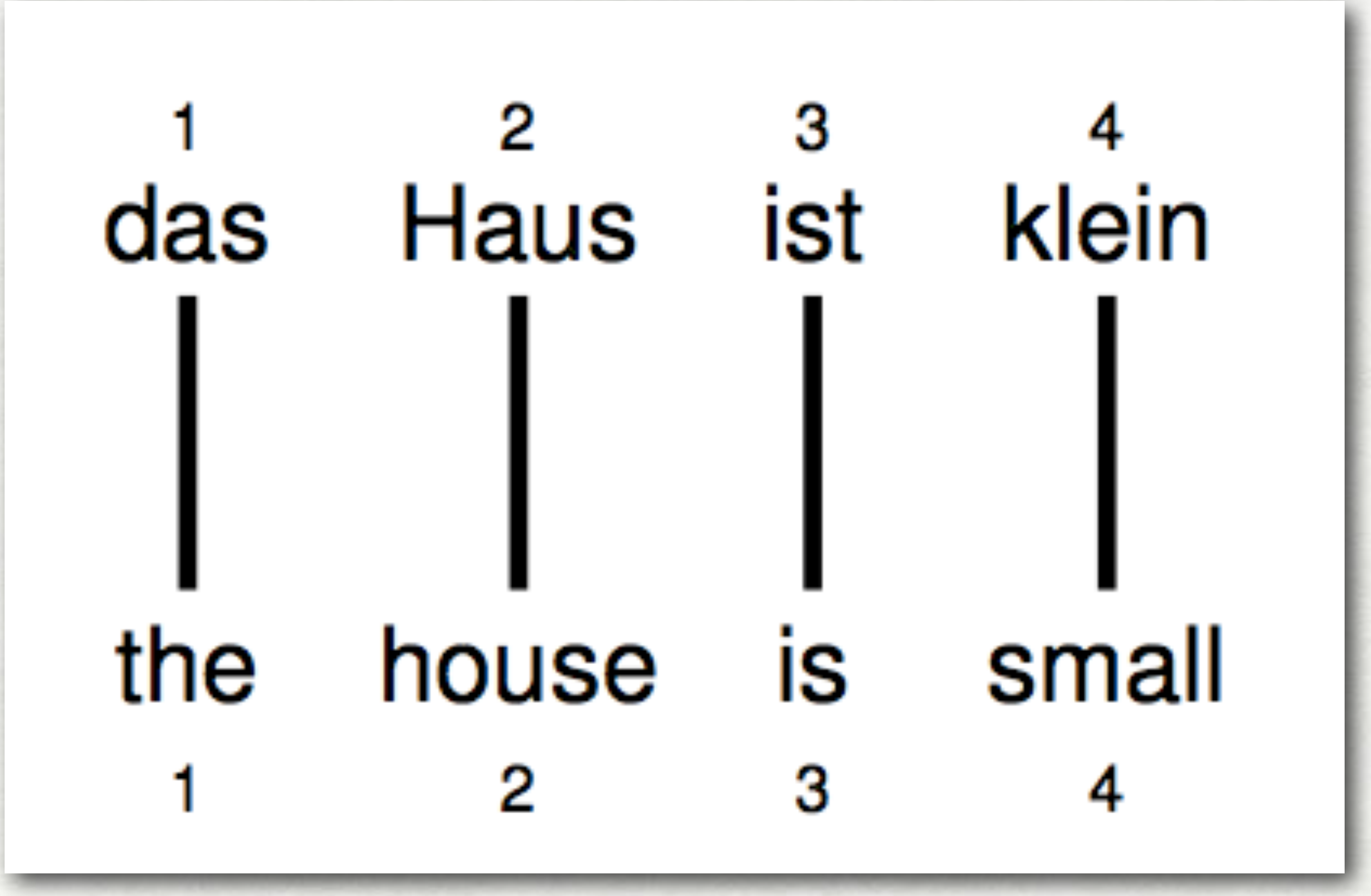
# ALIGNMENT

$$p(\mathbf{a} \mid \mathbf{f}, m)$$

- Most of the action for the first 10 years of MT was here. Words weren't the problem. Word *order* was hard.

# ALIGNMENT

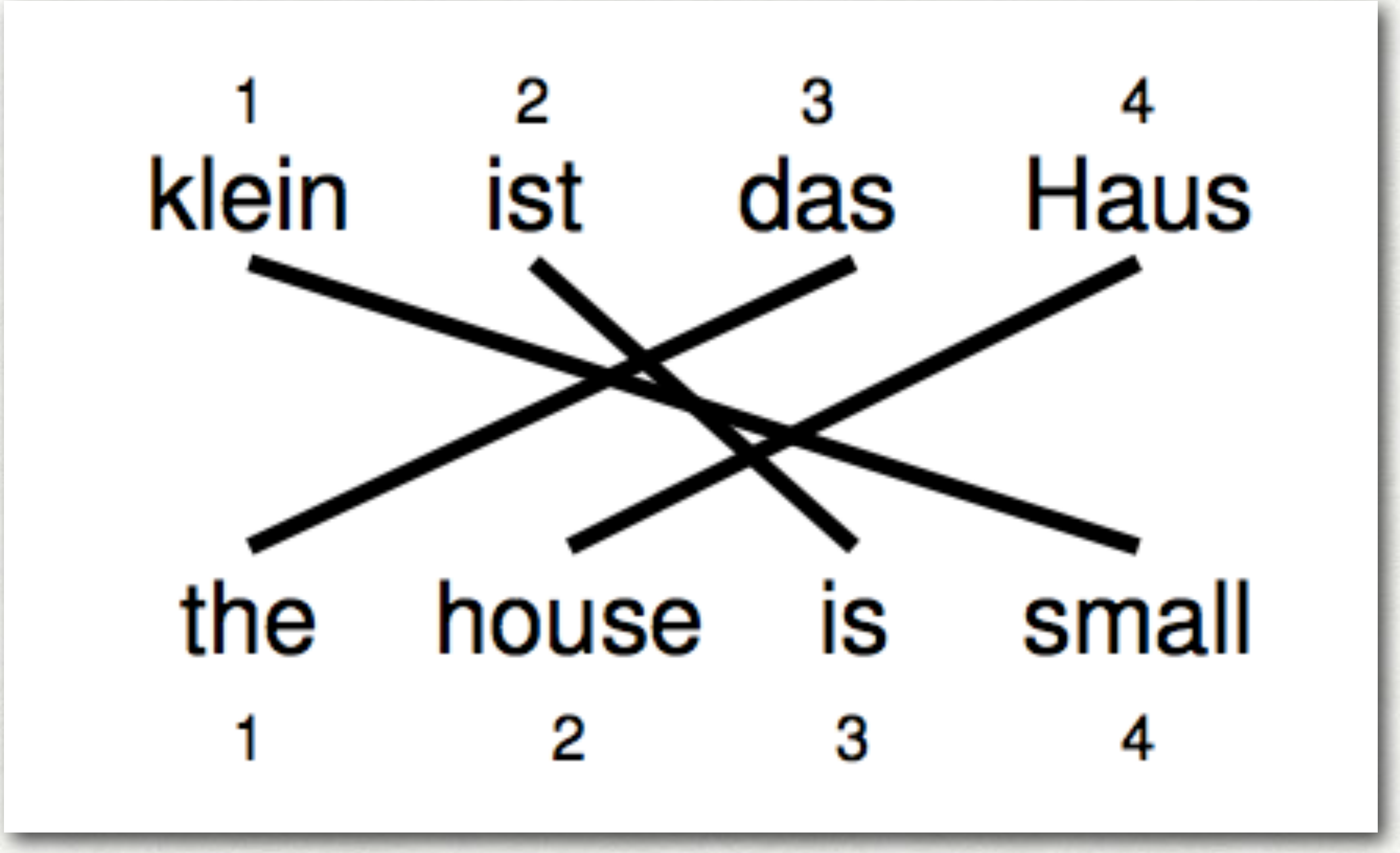
- Alignments can be visualized by drawing links between two sentences, and they are represented as vectors of positions:




$$\mathbf{a} = (1, 2, 3, 4)^\top$$

# REORDERING

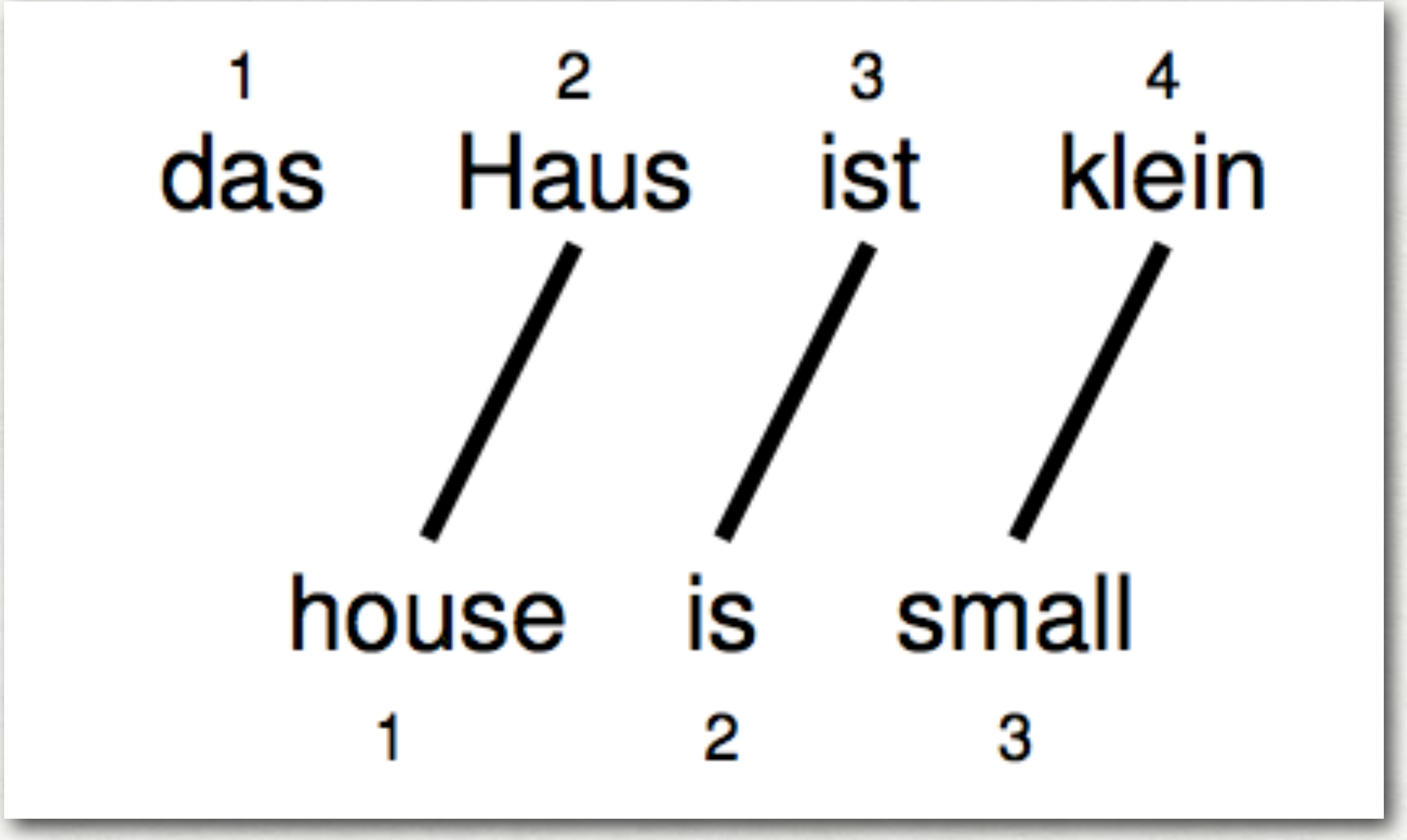
- Words may be reordered during translation




$$\mathbf{a} = (3, 4, 2, 1)^\top$$

# WORD DROPPING

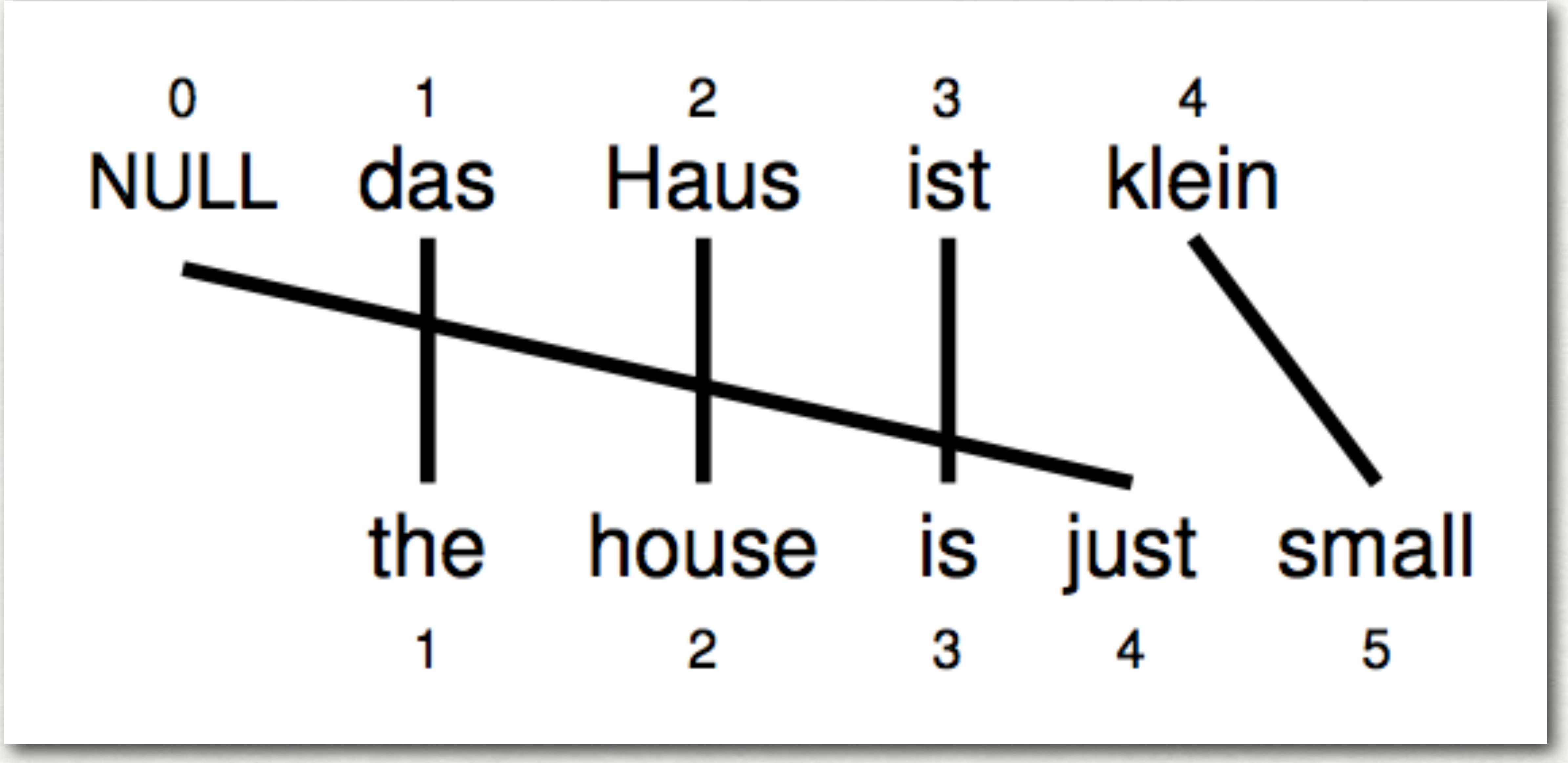
- A source word may not be translated at all




$$\mathbf{a} = (2, 3, 4)^\top$$

# WORD INSERTION

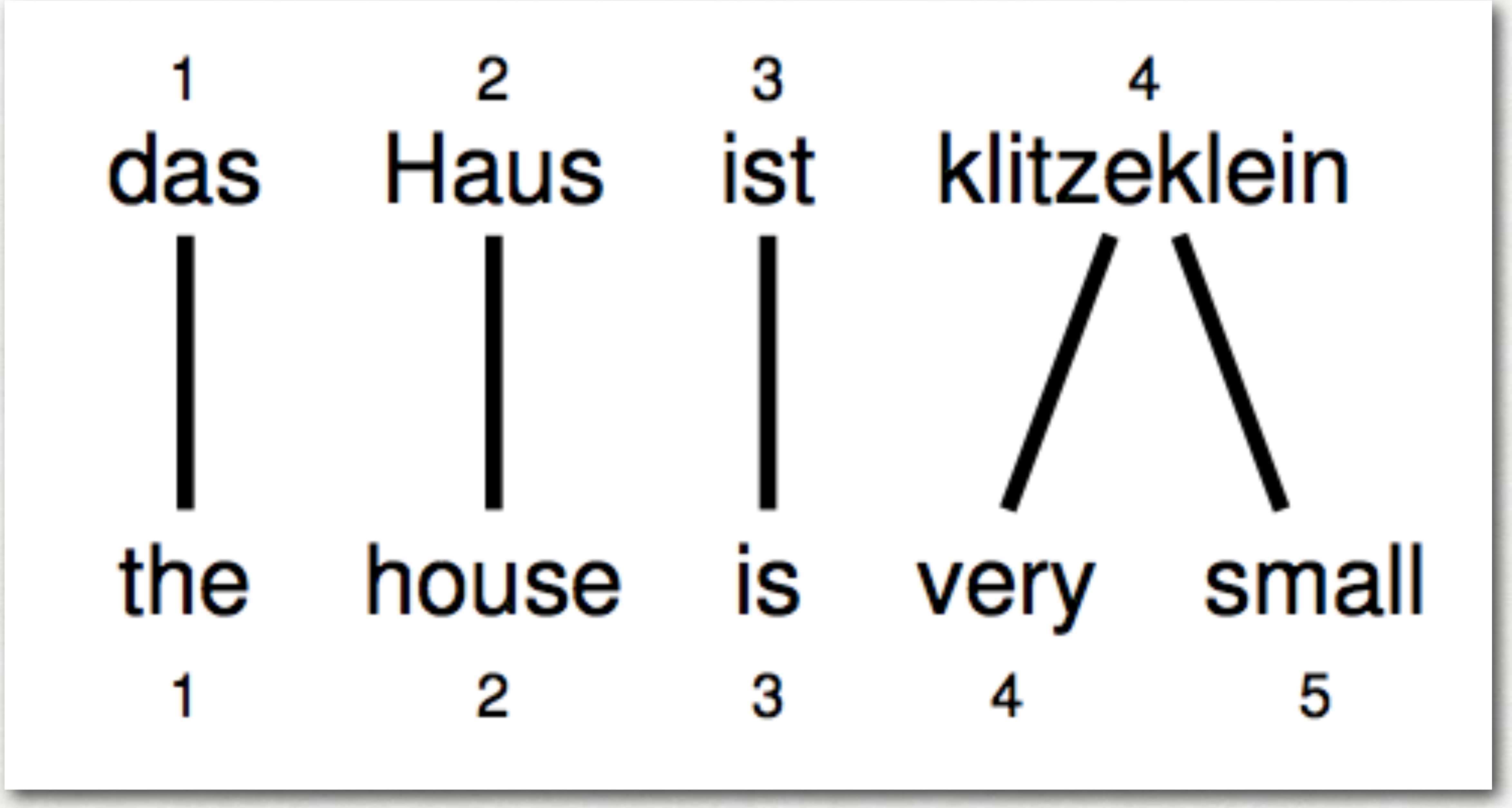
- Words may be inserted during translation
- E.g. English just does not have an equivalent
- But these words must be explained – we typically assume every source sentence contains a NULL token




$$\mathbf{a} = (1, 2, 3, 0, 4)^\top$$

# ONE-TO-MANY TRANSLATION

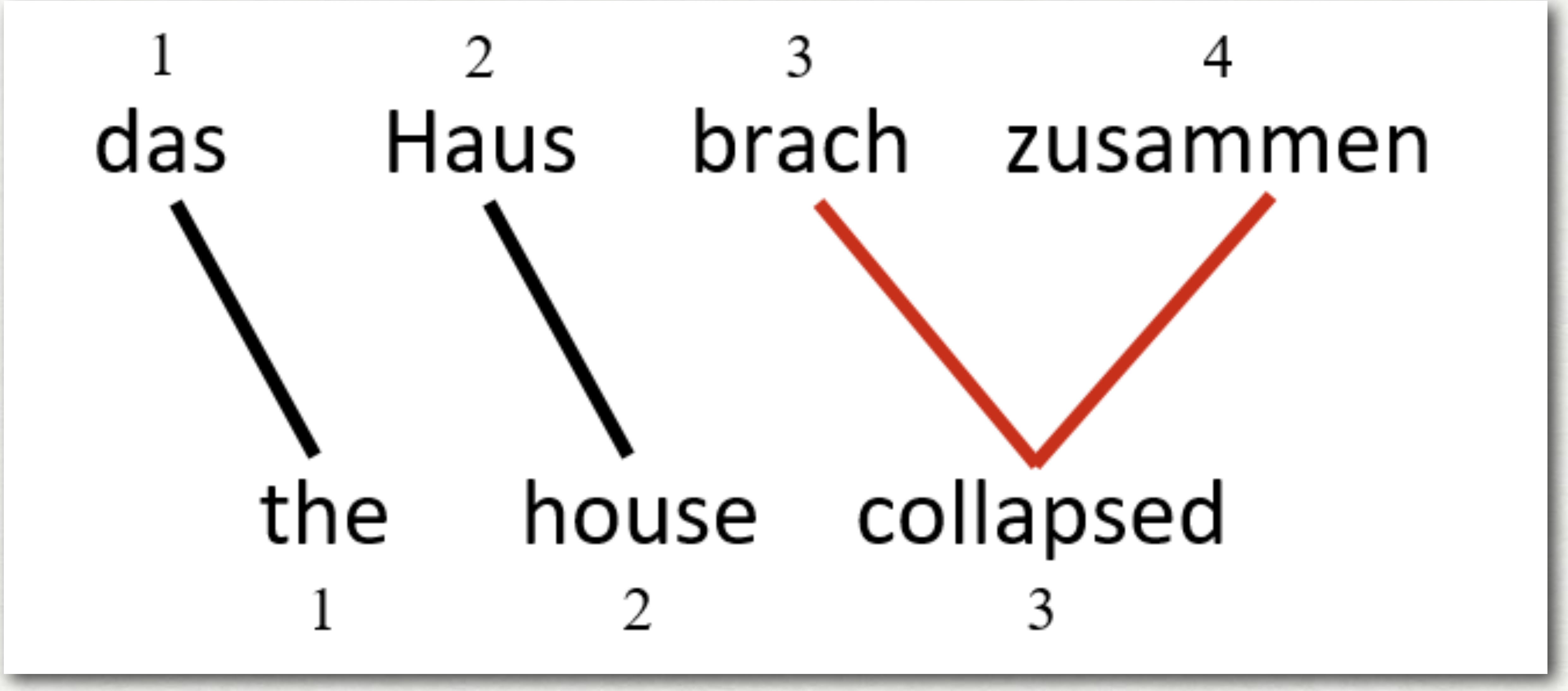
- A source word may translate into **more than one** target word




$$\mathbf{a} = (1, 2, 3, 4, 4)^\top$$

# MANY-TO-ONE TRANSLATION

- More than one source word may **not** translate as a unit in lexical translation




$\mathbf{a} = ???$ $\qquad\qquad$ $\mathbf{a} = (1, 2, (3, 4)^\top)^\top$ ?

# IBM MODEL 1

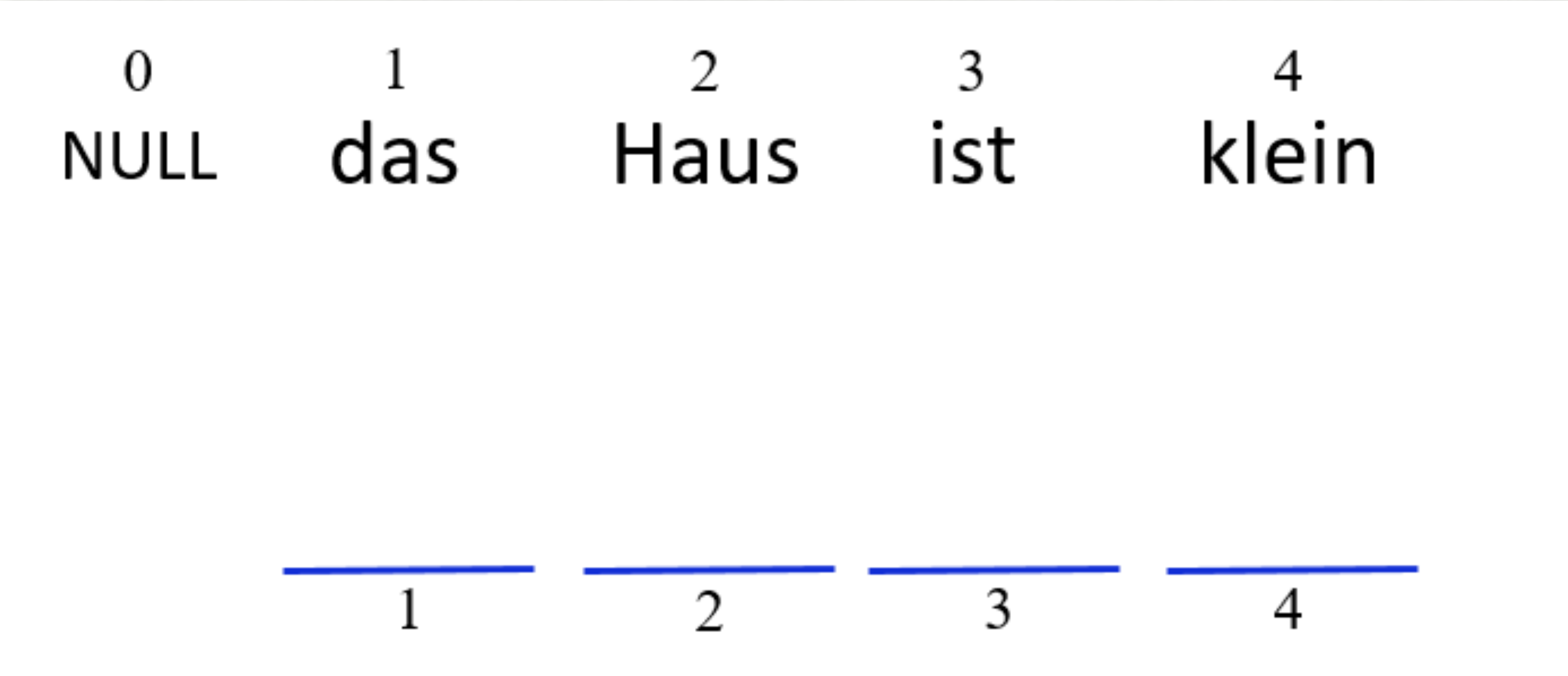
Simplest possible lexical translation model

Additional assumptions:

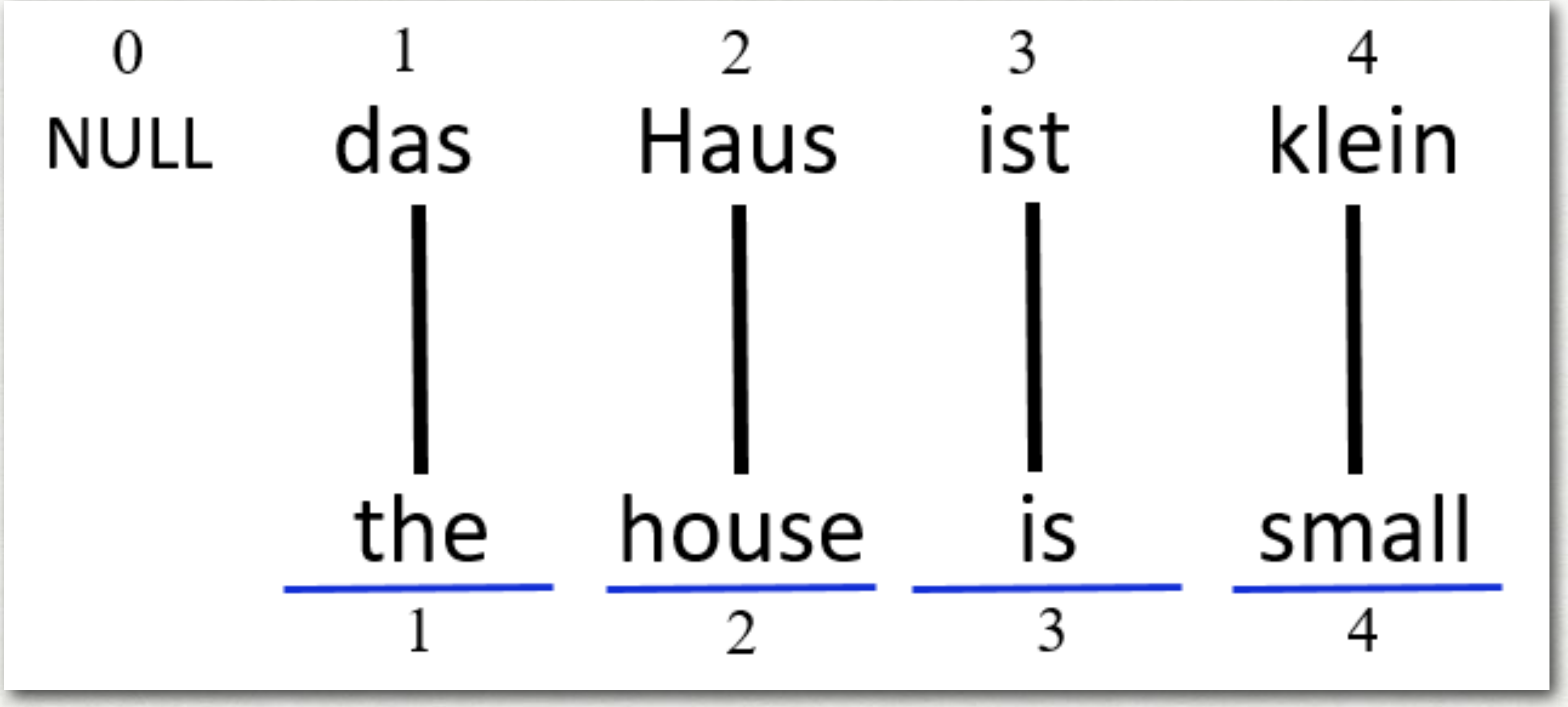
- The *m* alignment decisions are independent
- The alignment distribution for each $\mathbf{a}_i$ is uniform over all source words and NULL

$$
\begin{aligned}
&\text{for each } i \in [1, 2, \ldots, m] \\
&\qquad a_i \sim \text{Uniform}(0, 1, 2, \ldots, n) \\
&\qquad e_i \sim \text{Categorical}(\boldsymbol{\theta}_{f_{a_i}})
\end{aligned}
$$

# TRANSLATING WITH MODEL 1

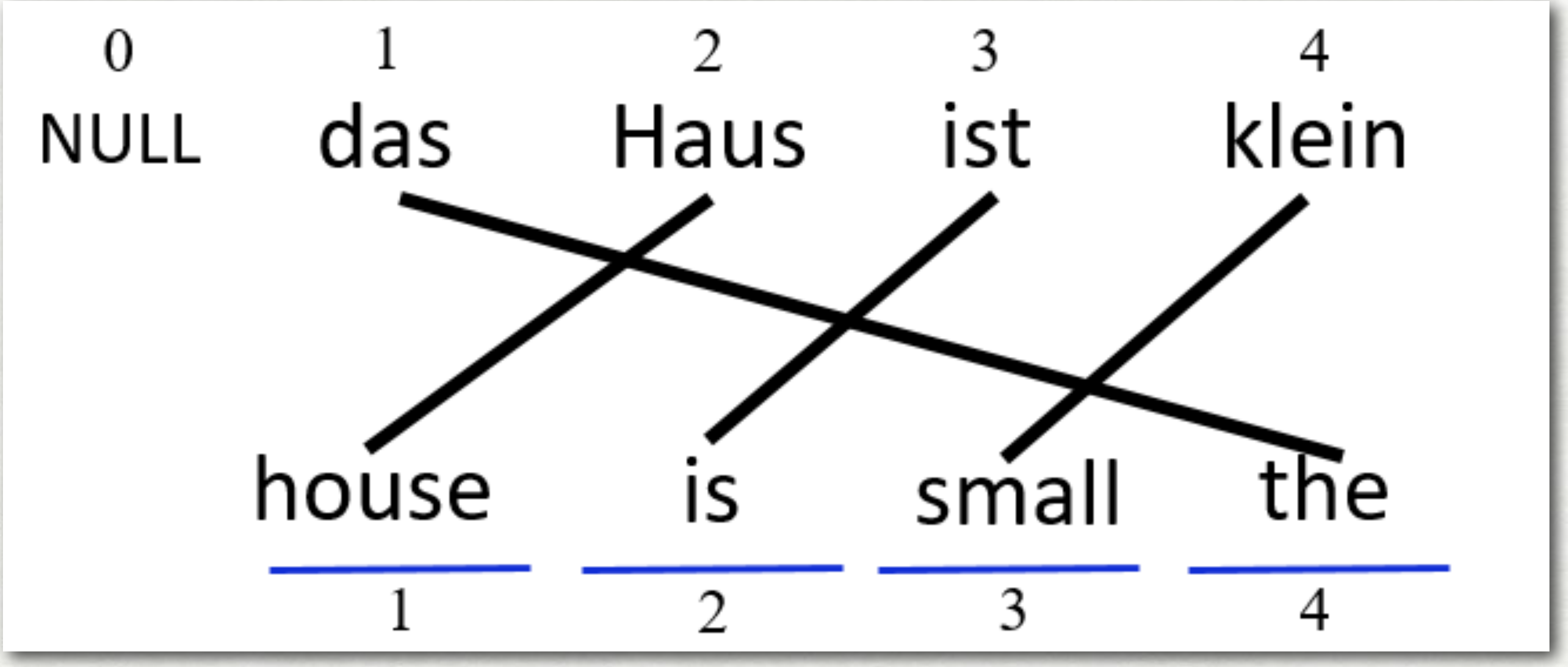


# TRANSLATING WITH MODEL 1




Language model says: ☺

# TRANSLATING WITH MODEL 1




Language model says: ☹

# LEARNING LEXICAL TRANSLATION MODELS

How do we learn the parameters *p(e|f)*?

"Chicken and egg" problem:

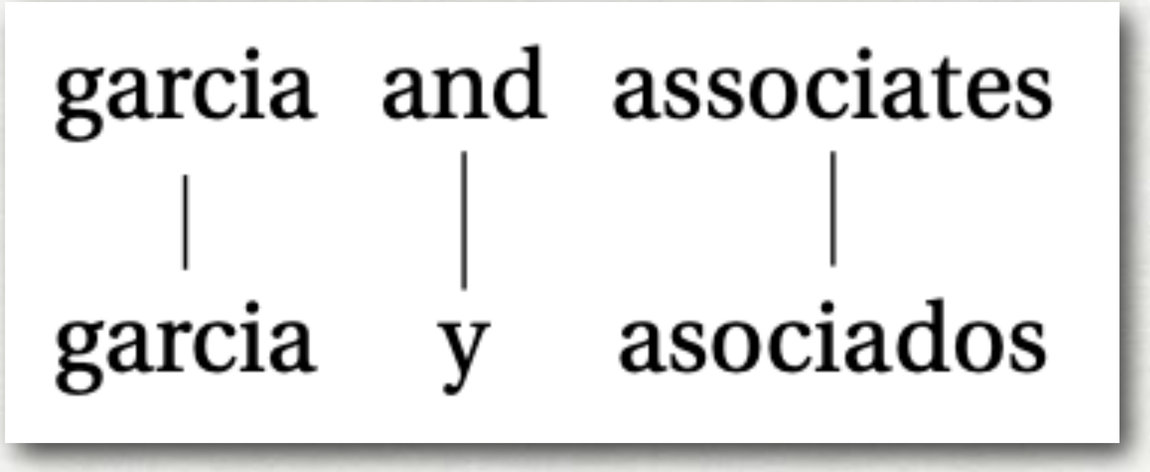
If we had the alignments, we could estimate the translation probabilities (MLE estimation)

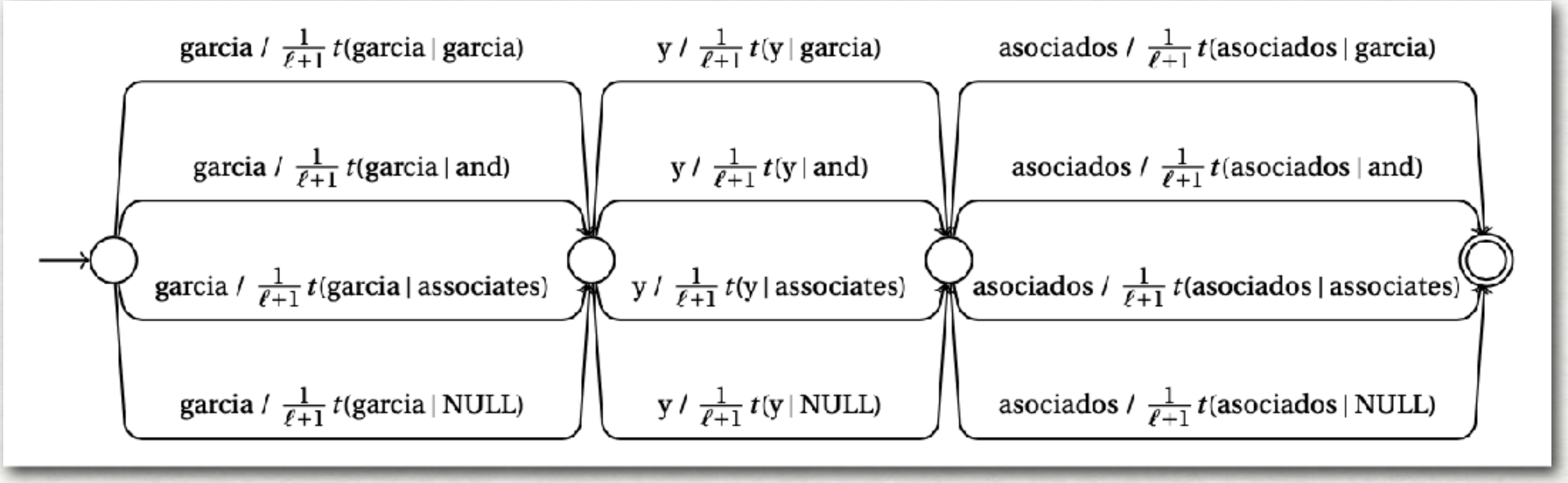
If we had the translation probabilities we could find the most likely alignments (greedy)

# IBM 1 - GENERATIVE STORY

We start with an English Sentence $\mathbf{e} = e_1 e_2 \ldots e_n$

1. Choose the length of the Spanish sentence $m$, with uniform probability $\epsilon = \frac{1}{M}$, where $M$ is the maximum allowed length of any Spanish sentence in the corpus.
2. Generate an alignment $a_1, \ldots, a_m$ again with uniform probability.
3. Generate Spanish words $f_1, \ldots, f_m$ each with probability $t(f_j \mid e_{a_j})$ or $t(f_j \mid \text{NULL})$




How can we estimate the $t(f \mid e)$ parameters?

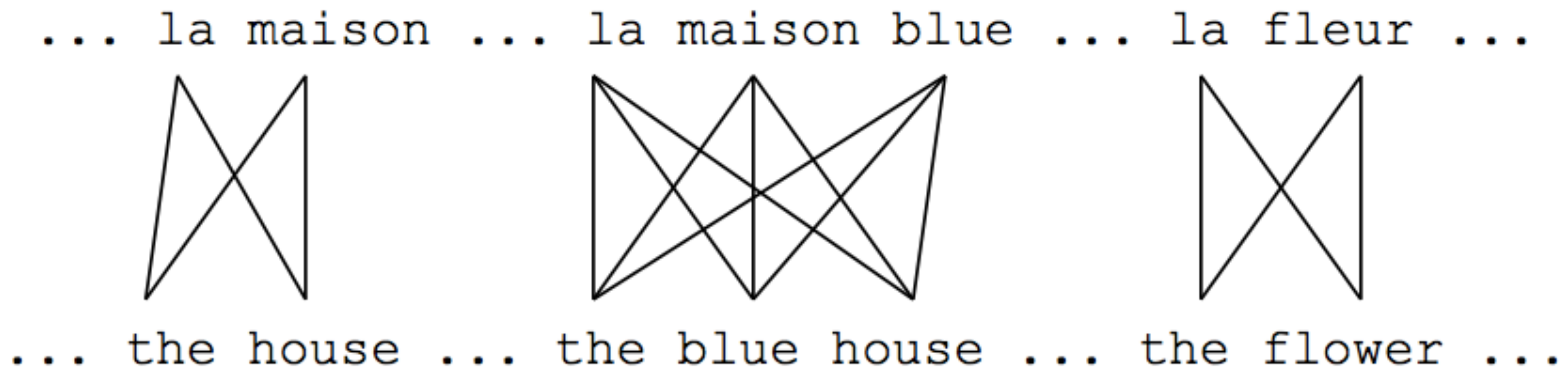
# EM ALGORITHM

Pick some random (or uniform) starting parameters

Repeat until bored (~5 iterations for lexical translation models):

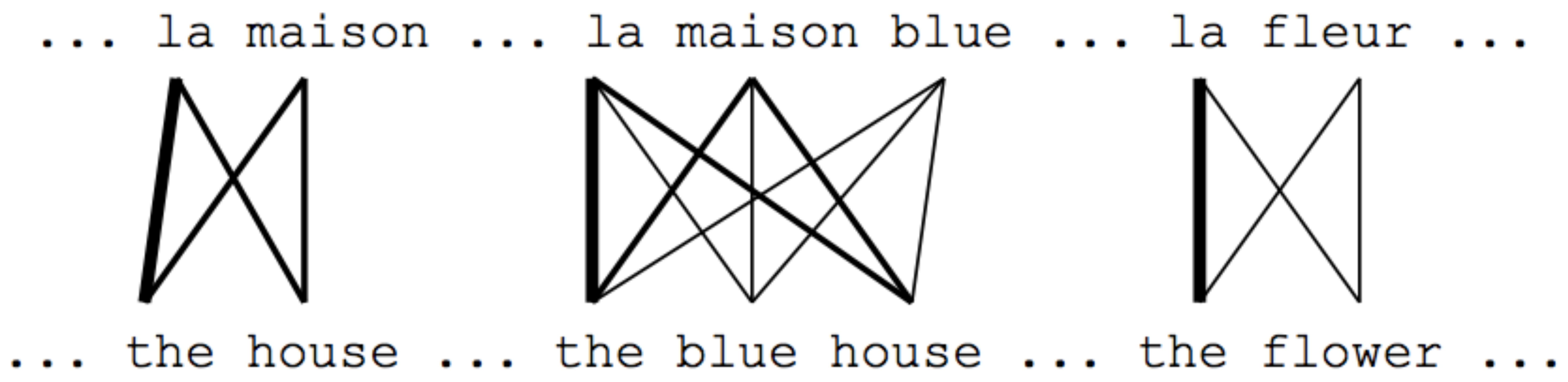
1. Using the current parameters, compute "expected" alignments $p(\mathbf{a}_i|\mathbf{e}, \mathbf{f})$ for every target word token in the training data
2. Keep track of the expected number of times *f* translates into *e* throughout the whole corpus
3. Keep track of the number of times *f* is used in the source of any translation
4. Use these estimates in the standard MLE equation to get a better set of parameters

# EM FOR MODEL 1



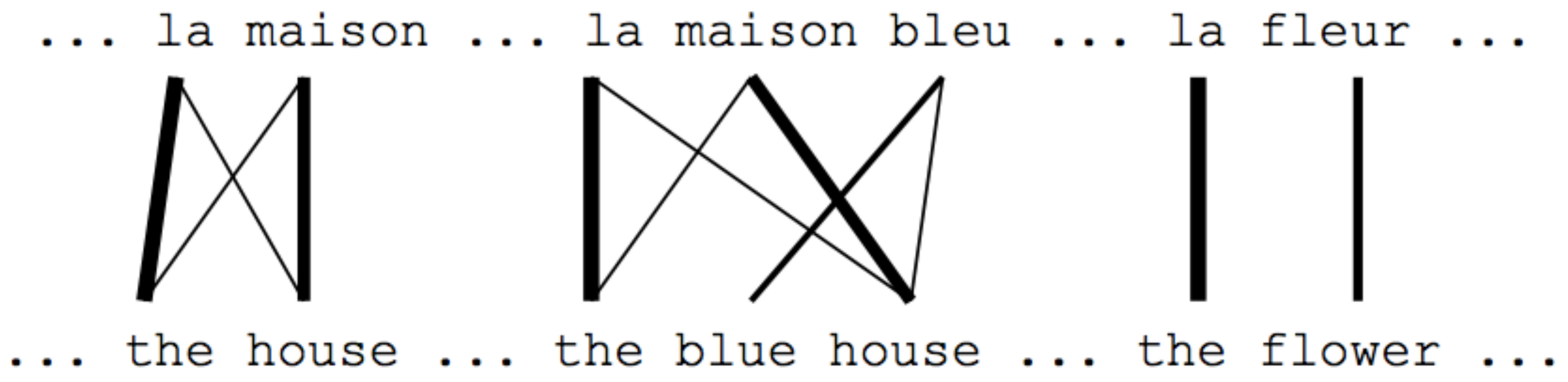

- Initial step: all alignments equally likely
- Model learns that, e.g., la is often aligned with the

# EM FOR MODEL 1




- After one iteration
- Alignments, e.g., between la and the are more likely

# EM FOR MODEL 1




- After another iteration
- It becomes apparent that alignments, e.g., between fleur and flower are more likely (pigeon hole principle)

# EM FOR MODEL 1


p(la|the) = 0.453
p(le|the) = 0.334
p(maison|house) = 0.876
p(bleu|blue) = 0.563
...

- Parameter estimation from the aligned corpus

# EM ALGORITHM - PSEUDOCODE

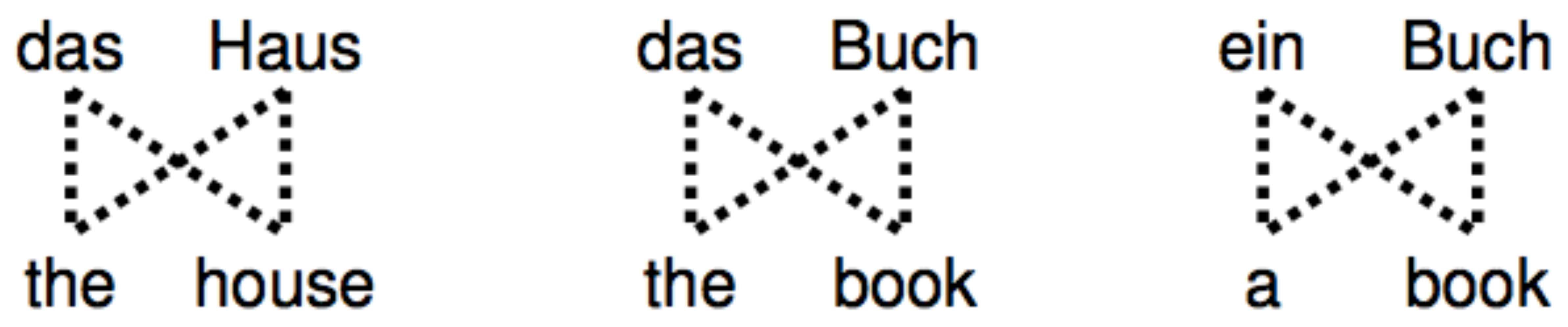
1. Initialize $t(\cdot \mid e)$ to uniform: $t(f \mid e) = \dfrac{1}{|V_f|}$ where $V_f$ is the Spanish vocabulary, and $e$ is any English word or NULL.
2. E-step: Calculate the *expected* number of times that word $e$ is translated as $f$.
   For each $i, j$ the transition that generates $f_j$ from $e_i$ "competes" with the transitions that generate $f_j$ from the other English words (or NULL). So we update our expected counts $c(f, e)$ as follows:

$$c(f_j, e_i) \leftarrow c(f_j, e_i) + \frac{t(f_j \mid e_i)}{t(f_j \mid \text{NULL}) + \Sigma_{i'} t(f_j \mid e_{i'})} \qquad c(f_j, \text{NULL}) \leftarrow c(f_j, \text{NULL}) + \frac{t(f_j \mid \text{NULL})}{t(f_j \mid \text{NULL}) + \Sigma_{i'} t(f_j \mid e_{i'})}$$

3. M-step: Estimate the model's parameters based on the expected counts.
   Let $t(f \mid e) \leftarrow \dfrac{c(f, e)}{\Sigma_f c(f, e)}$ where $e$ is any English word or NULL.

4. Go to step 2.

# CONVERGENCE



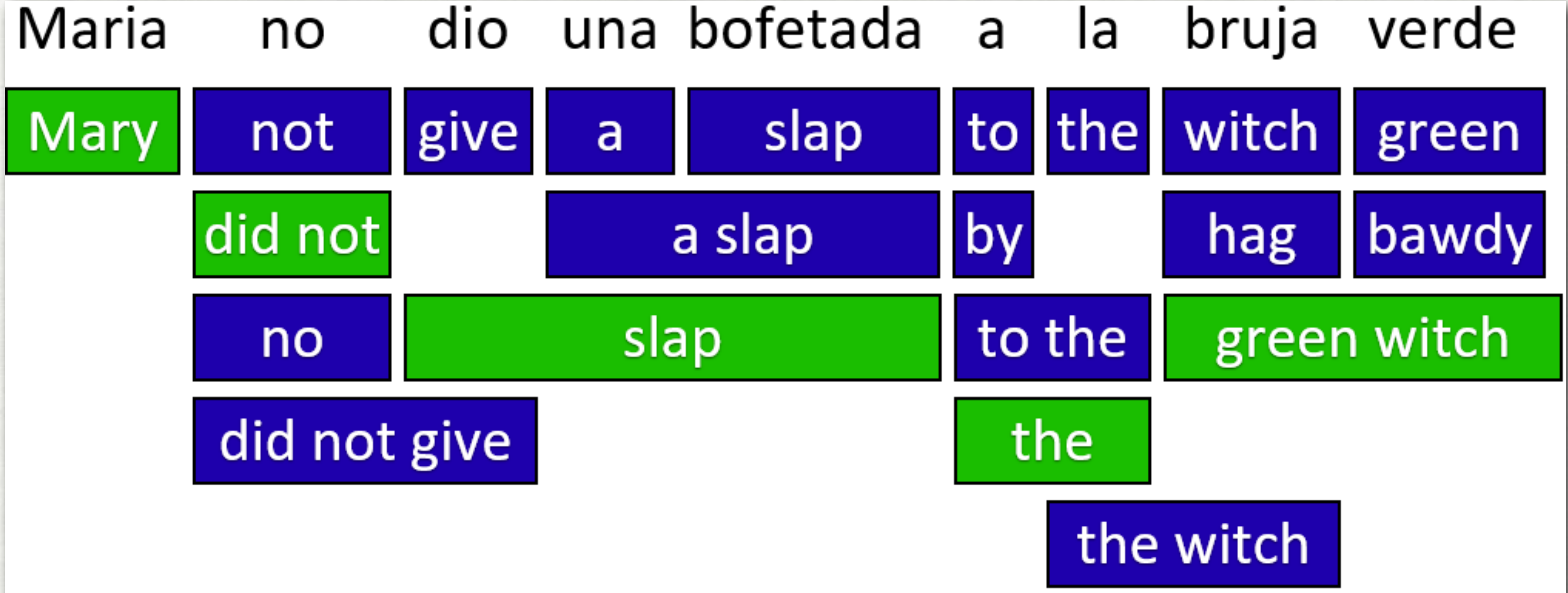

| $e$ | $f$ | initial | 1st it. | 2nd it. | 3rd it. | ... | final |
|---|---|---|---|---|---|---|---|
| the | das | 0.25 | 0.5 | 0.6364 | 0.7479 | ... | 1 |
| book | das | 0.25 | 0.25 | 0.1818 | 0.1208 | ... | 0 |
| house | das | 0.25 | 0.25 | 0.1818 | 0.1313 | ... | 0 |
| the | buch | 0.25 | 0.25 | 0.1818 | 0.1208 | ... | 0 |
| book | buch | 0.25 | 0.5 | 0.6364 | 0.7479 | ... | 1 |
| a | buch | 0.25 | 0.25 | 0.1818 | 0.1313 | ... | 0 |
| book | ein | 0.25 | 0.5 | 0.4286 | 0.3466 | ... | 0 |
| a | ein | 0.25 | 0.5 | 0.5714 | 0.6534 | ... | 1 |
| the | haus | 0.25 | 0.5 | 0.4286 | 0.3466 | ... | 0 |
| house | haus | 0.25 | 0.5 | 0.5714 | 0.6534 | ... | 1 |

# EXTENSIONS

Phrase-based MT:

Allow multiple words to translate as chunks (including many-to-one)

Introduce another latent variable, the source *segmentation*

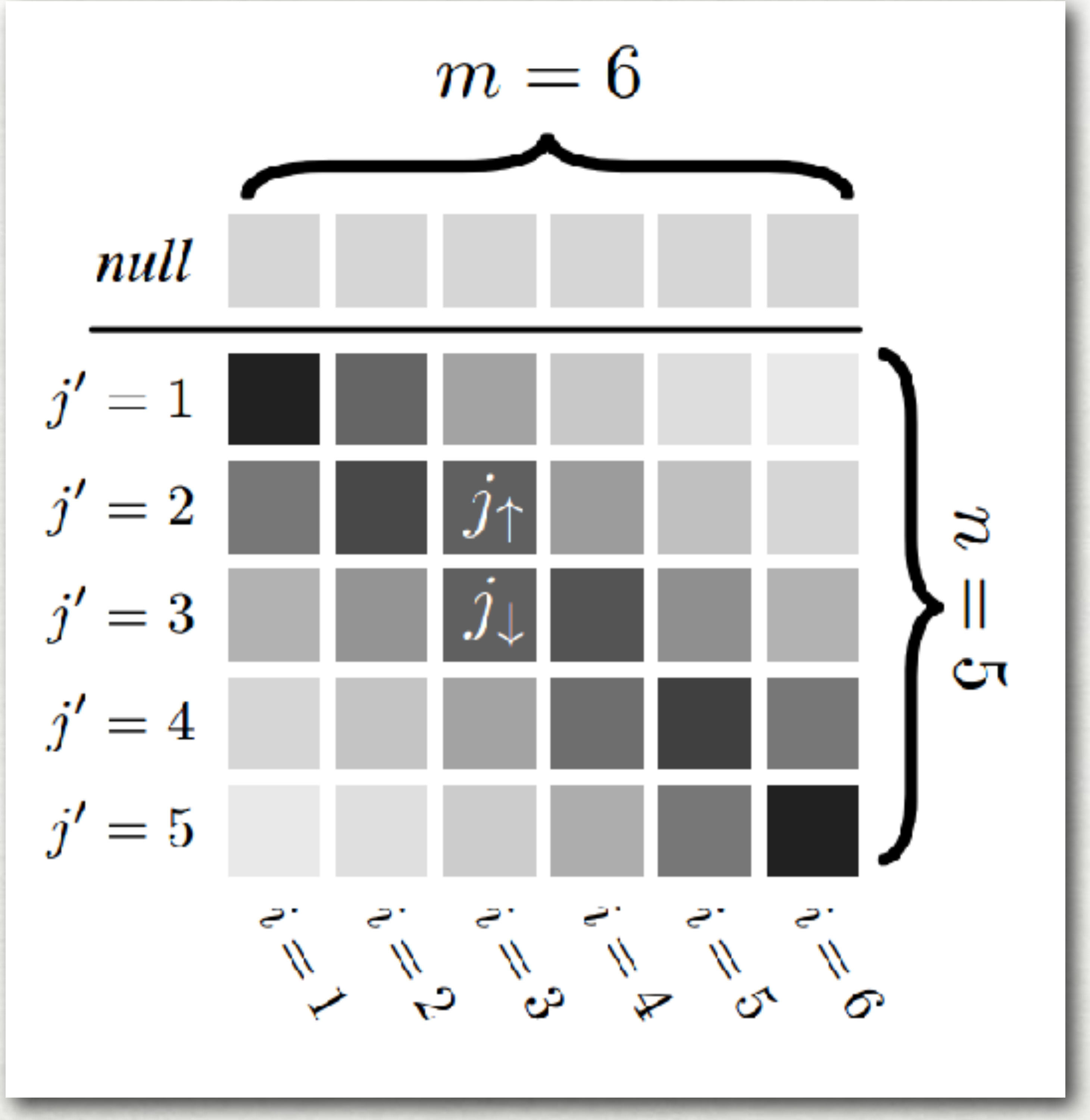



Adapted from Koehn (2006)

# EXTENSIONS

Alignment Priors:

Instead of assuming the alignment decisions are uniform, impose (or learn) a prior over alignment grids:



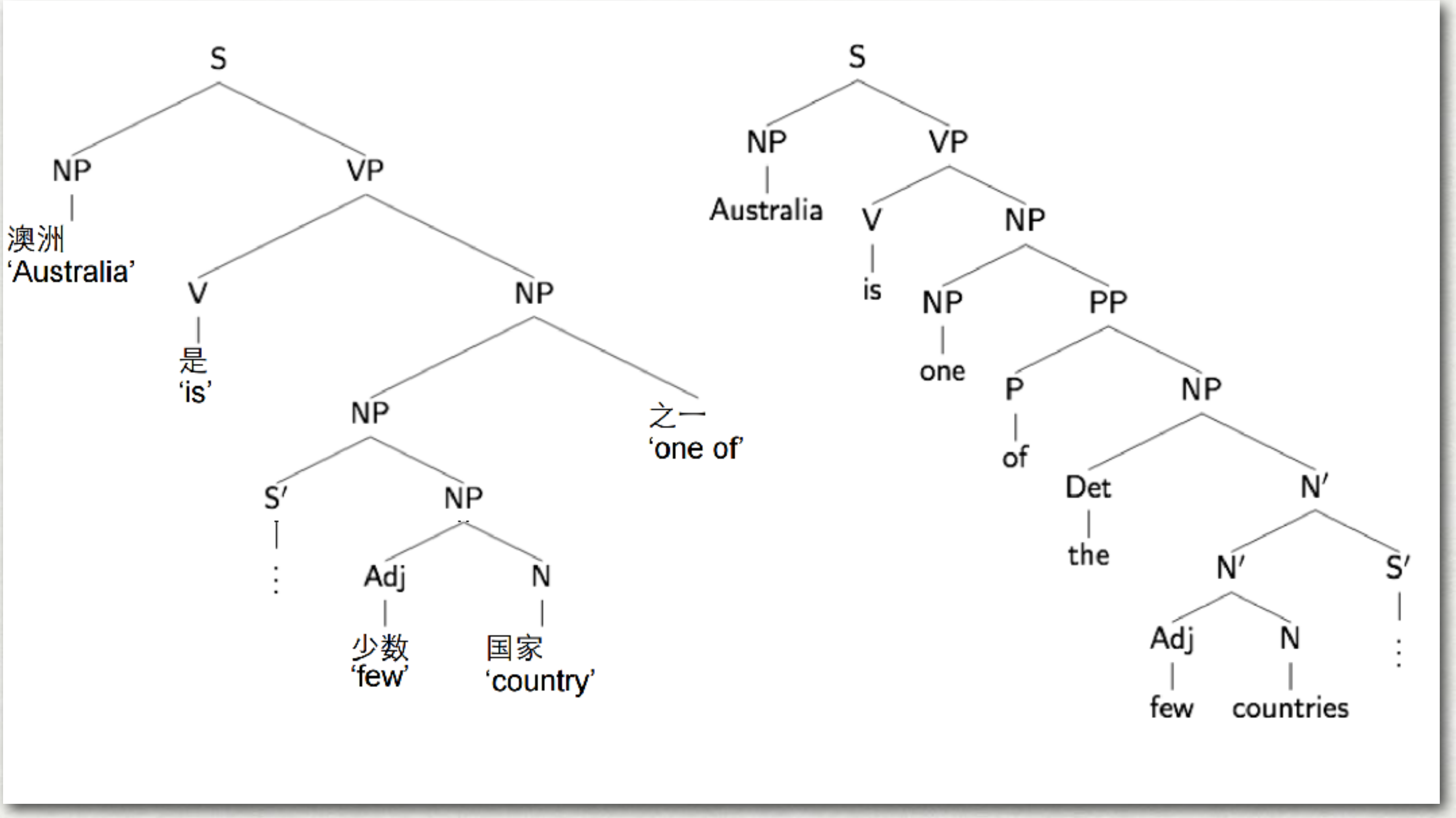

Chahuneau et al. (2013)

# EXTENSIONS

Syntactic structure

Rules of the form:

X之一 → one of the X




Chiang (2005), Galley et al. (2006)

# EVALUATION

How do we evaluate translation systems’ output?

Central idea: “The closer a machine translation is to a professional human translation, the better it is.”

Most commonly used metric is called BLEU

# BLEU: AN EXAMPLE

Candidate 1: *It is a guide to action which ensures that the military always obey the commands of the party.*

Reference 1: *It is a guide to action that ensures that the military will forever heed Party commands.*

Reference 2: *It is the guiding principle which guarantees the military forces always being under the command of the Party.*

Reference 3: *It is the practical guide for the army always to heed directions of the party.*

Unigram Precision : 17/18

Adapted from slides by Arthur Chan

# ISSUE OF N-GRAM PRECISION

What if some words are over-generated?
e.g. "the"

An extreme example

Candidate: *the the the the the the the.*

Reference 1: *The cat is on the mat.*

Reference 2: *There is a cat on the mat.*

N-gram Precision: 7/7

**Solution: reference word should be exhausted after it is matched.**

Adapted from slides by Arthur Chan

# ISSUE OF N-GRAM PRECISION

What if some words are just dropped?

Another extreme example

Candidate: *the.*

Reference 1: *My mom likes the blue flowers.*

Reference 2: *My mother prefers the blue flowers.*

N-gram Precision: 1/1

**Solution: add a penalty if the candidate is too short.**

Adapted from slides by Arthur Chan

# BLEU

Geometric Average

$$\text{BLEU} = (p_1 \cdot p_2 \cdot p_3 \cdot p_4)^{\frac{1}{4}} \underbrace{\max(1,\ e^{1-\frac{r}{c}})}_{\text{Brevity Penalty}}$$

Clipped N-gram precisions for N=1, 2, 3, 4

Ranges from 0.0 to 1.0, but usually shown multiplied by 100

An increase of +1.0 BLEU is usually a conference paper

MT systems usually score in the 10s to 30s (40-50s?)

Human translators usually score in the 70s and 80s

# A SHORT SEGUE

Word- and phrase-based ("symbolic") models were cutting edge for decades (up until ~2014)

- Such models are still the most widely used in commercial applications

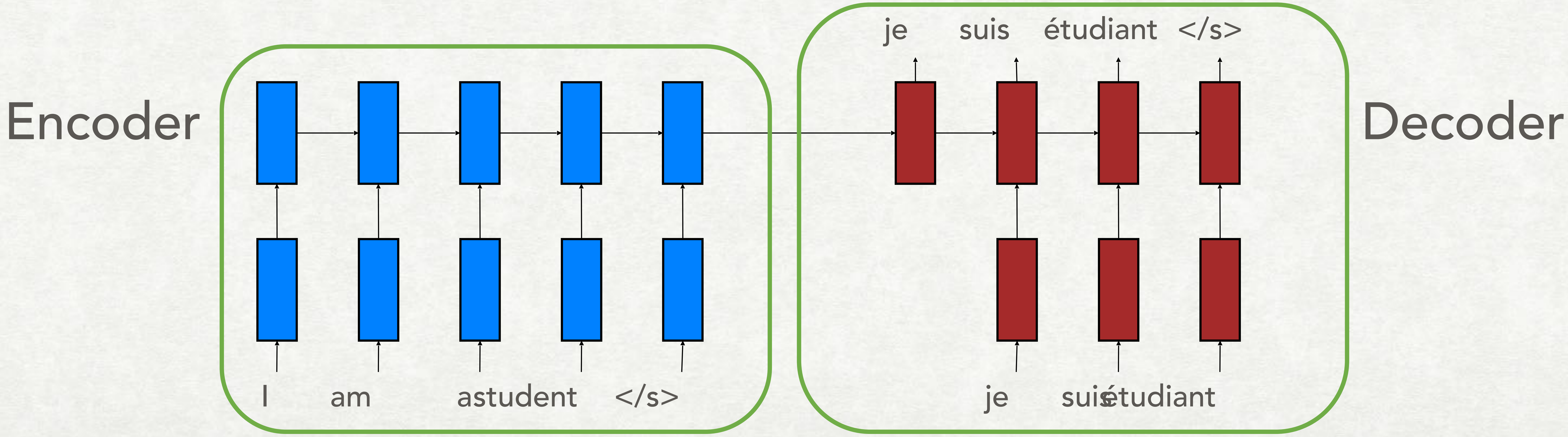
Since 2014 most research on MT has focused on **neural** models

# FULLY NEURAL TRANSLATION

Fully end-to-end RNN-based translation model

Encode the source sentence using one RNN

Generate the target sentence one word at a time using another RNN




Sutskever et al. (2014)

# ATTENTIONAL MODEL

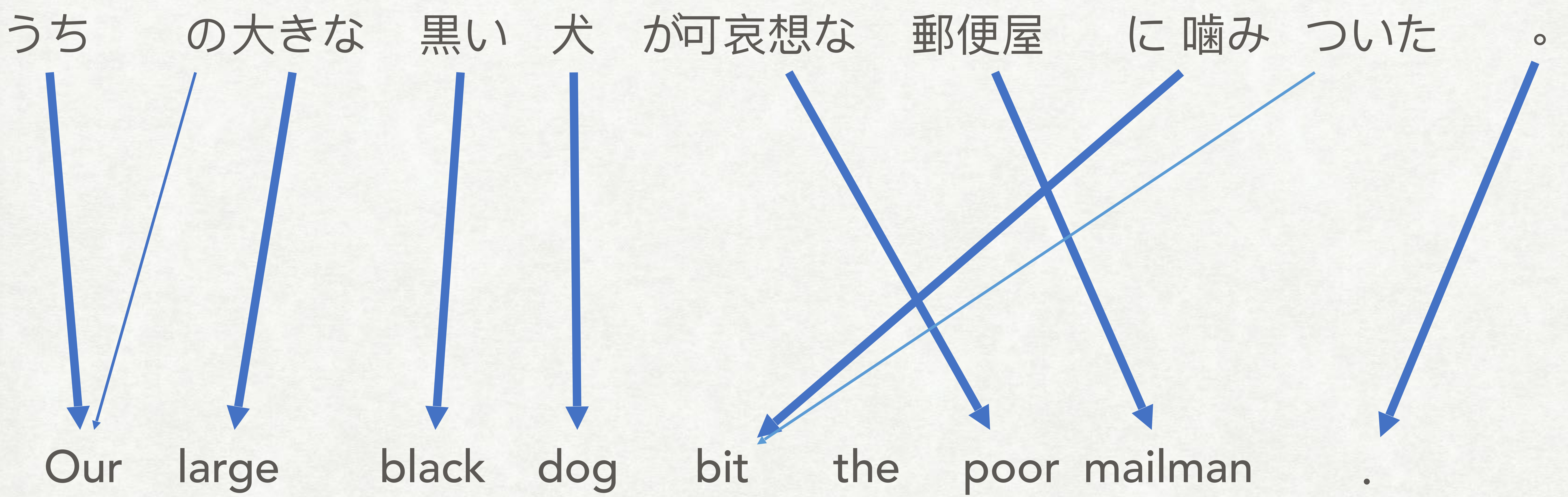
The encoder-decoder model struggles with long sentences

An RNN is trying to compress an arbitrarily long sentence into a finite-length worth vector

What if we only look at one (or a few) source words when we generate each output word?

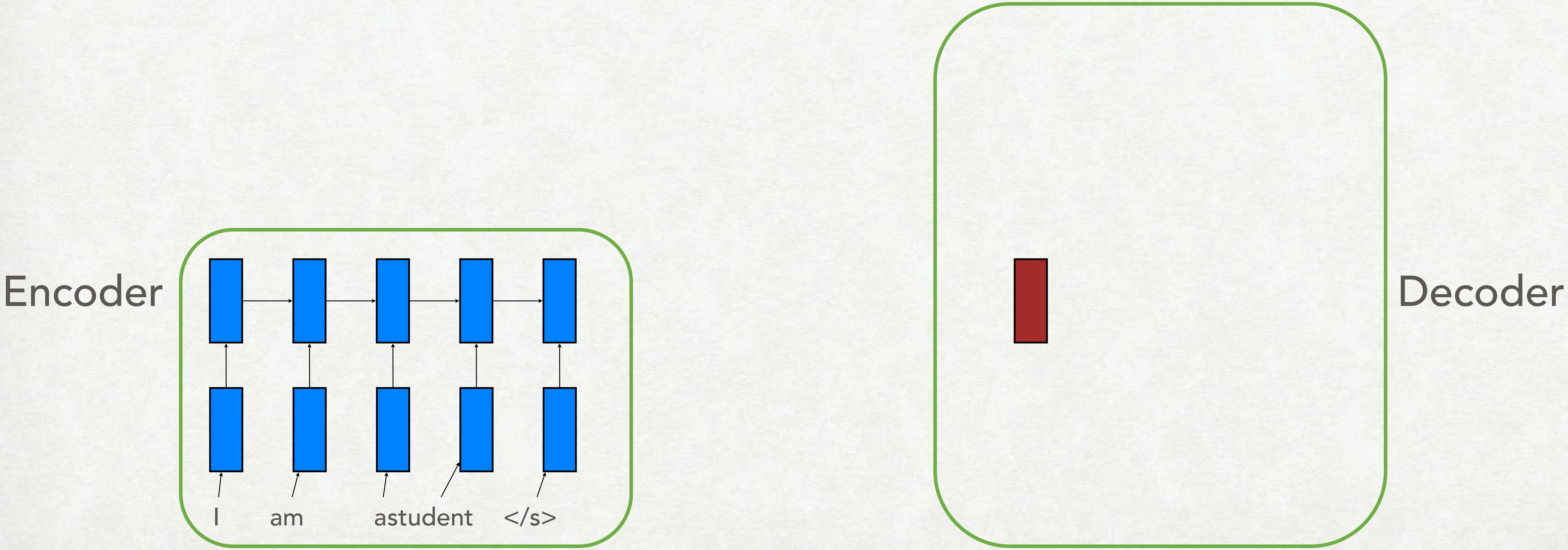
Bahdanau et al. (2014)

# THE INTUITION



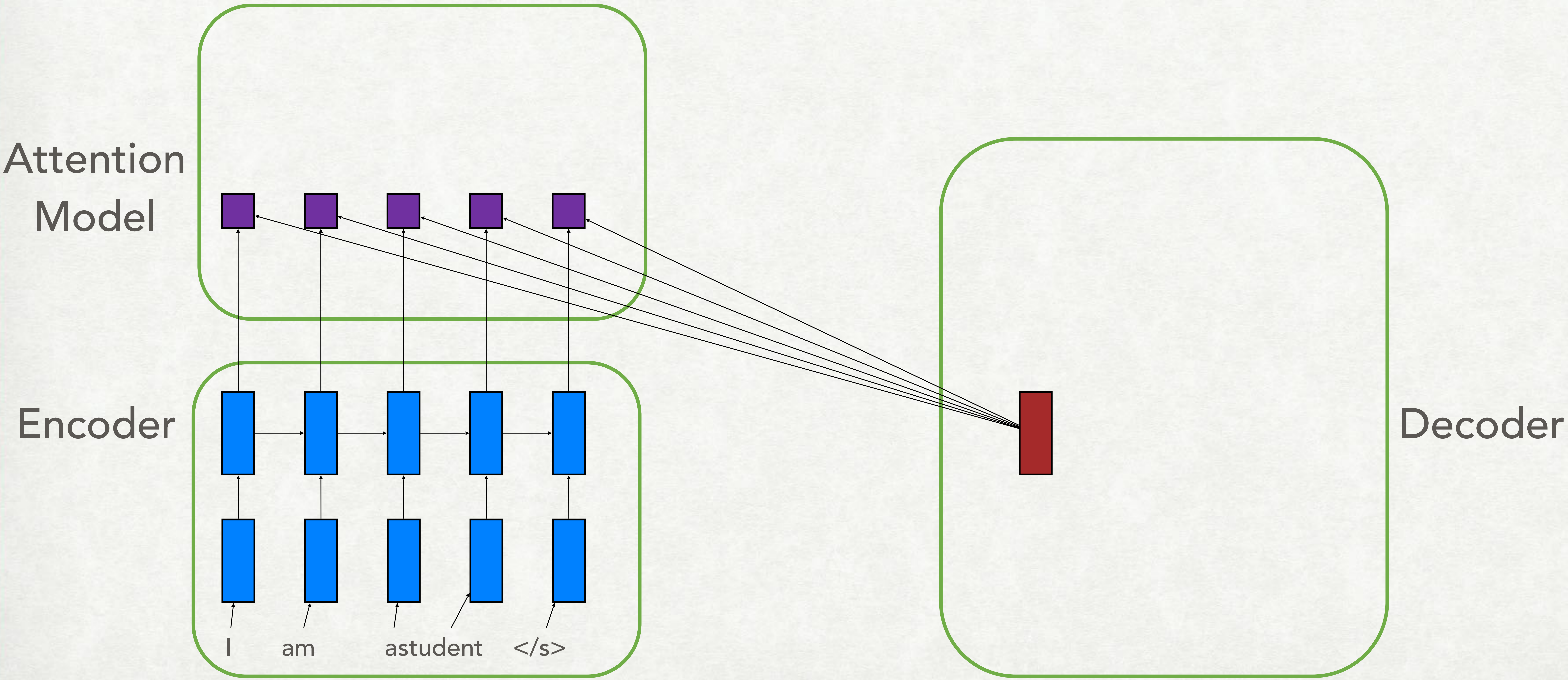

Bahdanau et al. (2014)

# THE ATTENTION MODEL



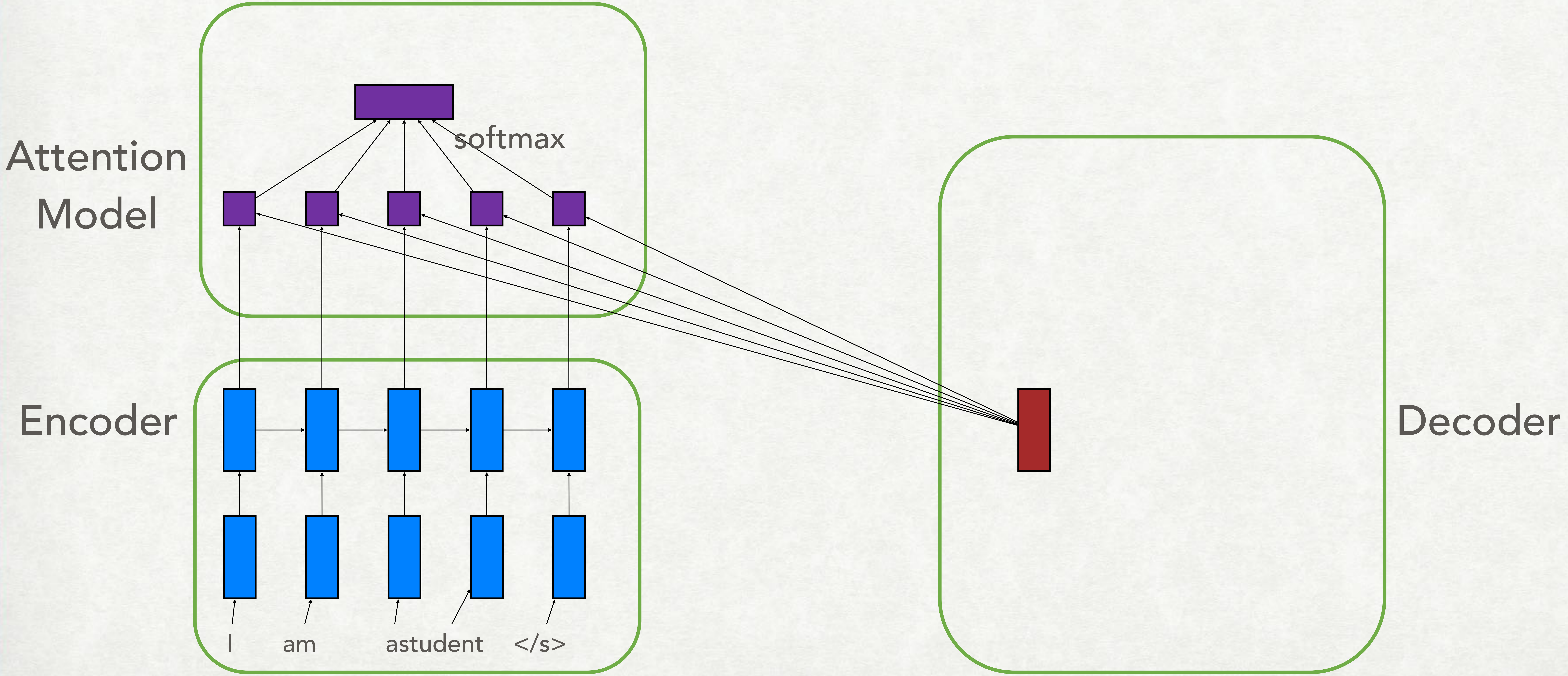

Bahdanau et al. (2014)

# THE ATTENTION MODEL



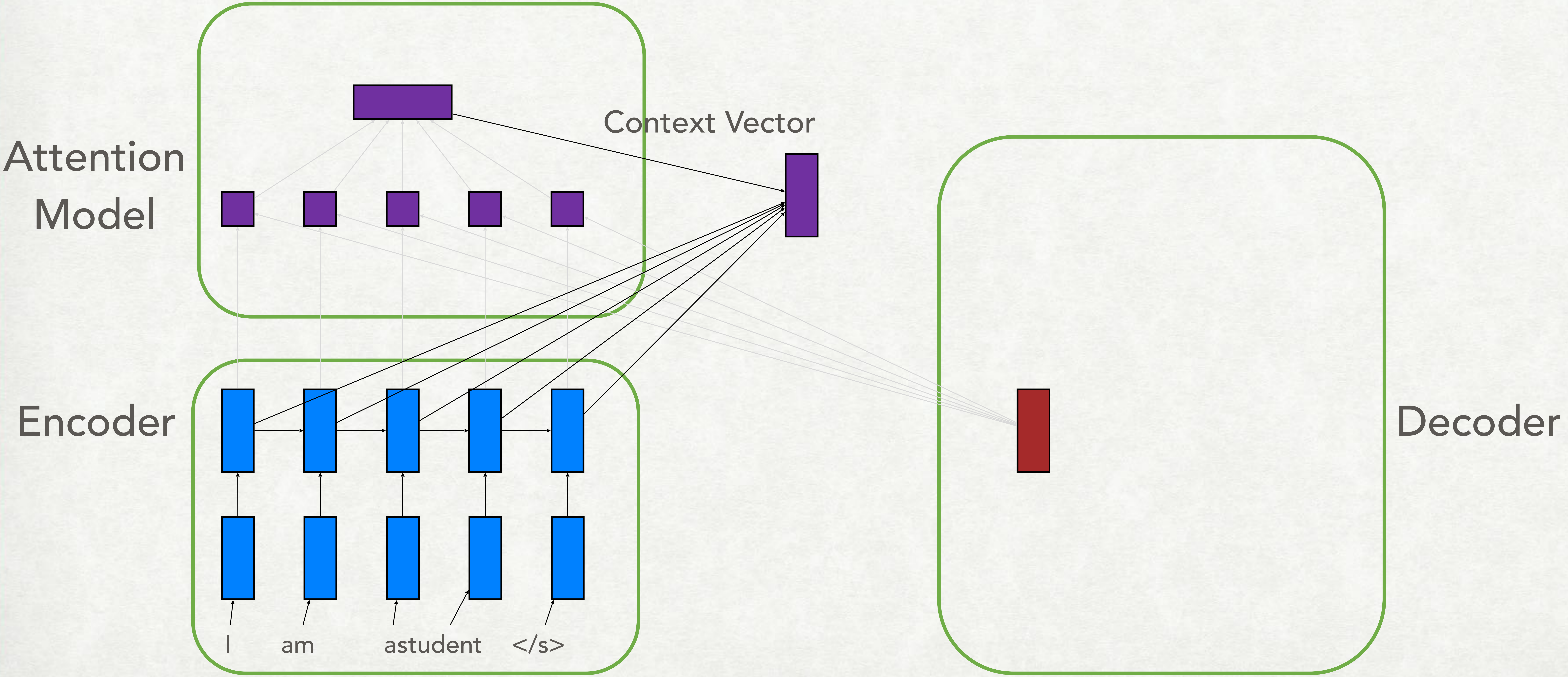

Bahdanau et al. (2014)

# THE ATTENTION MODEL



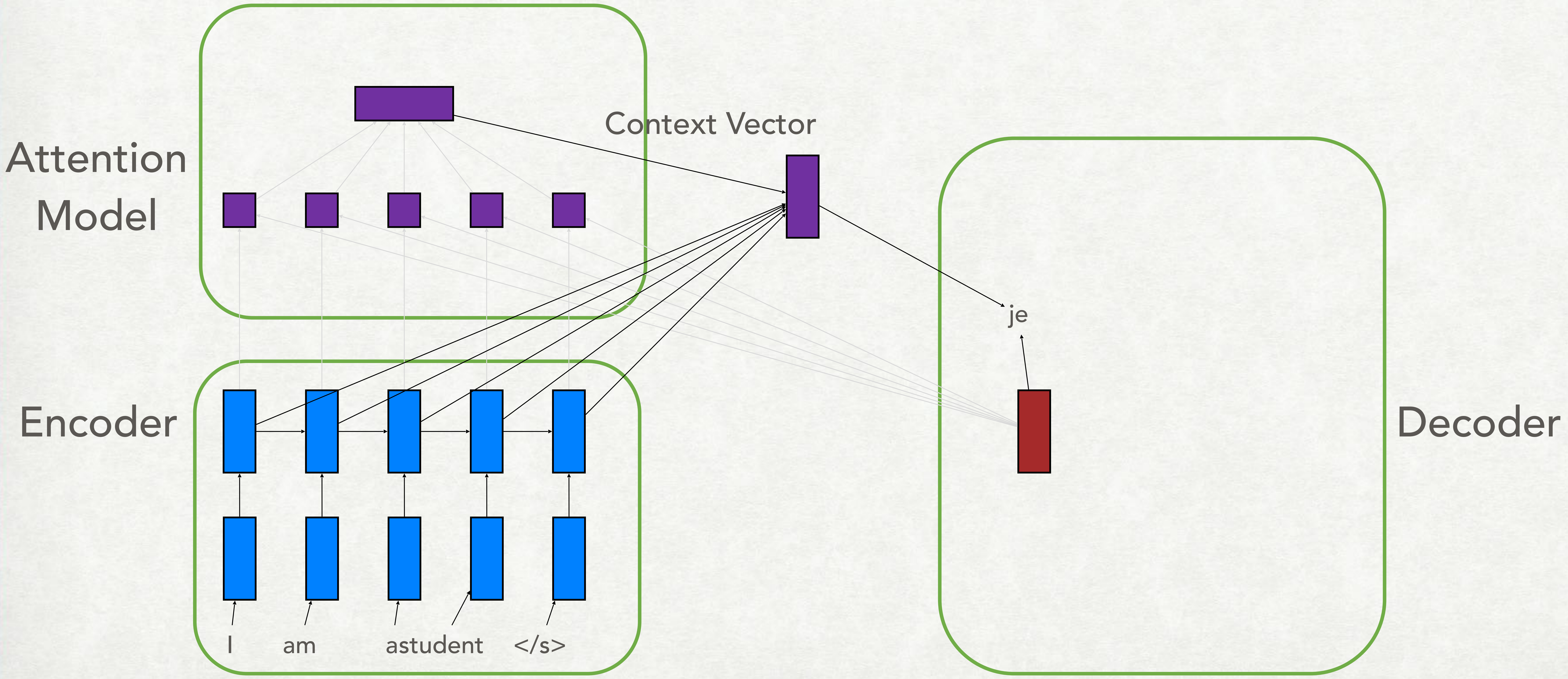

Bahdanau et al. (2014)

# THE ATTENTION MODEL



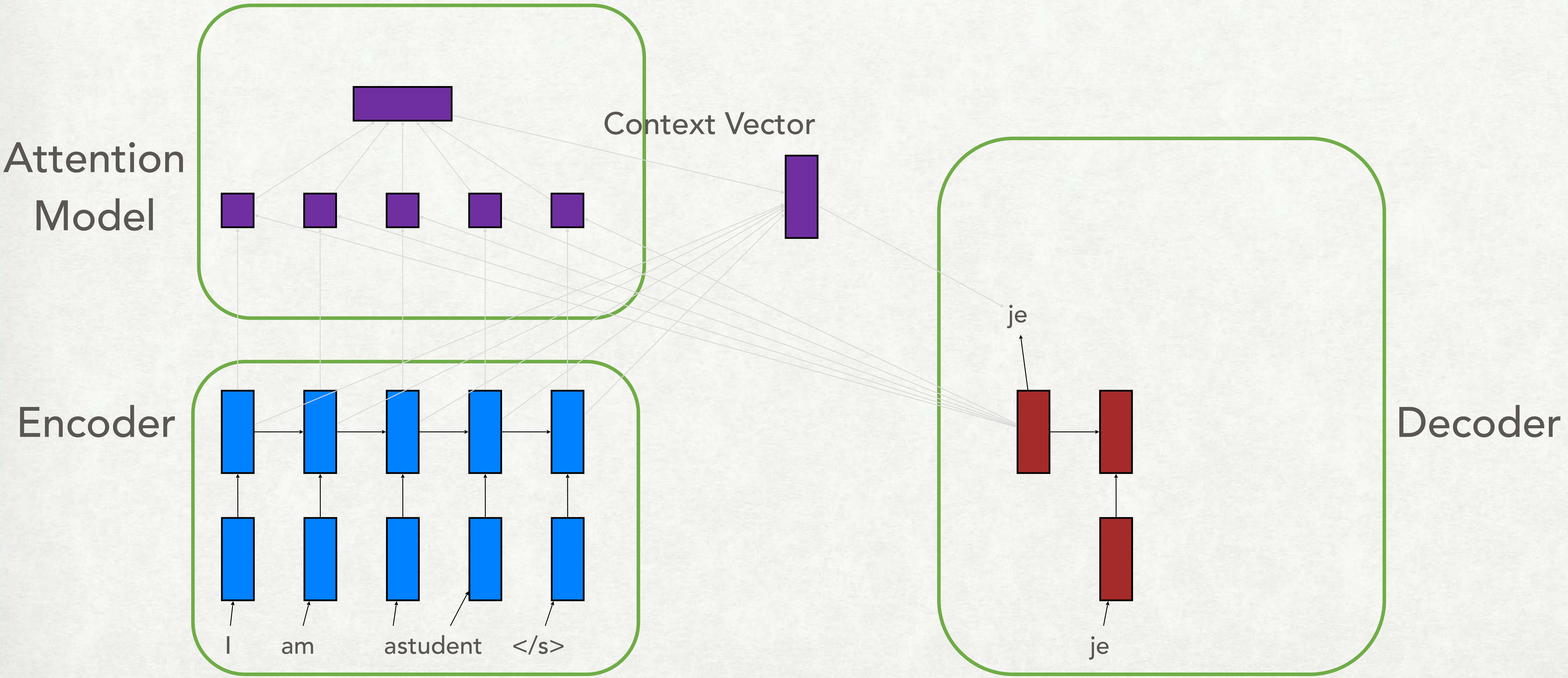

Bahdanau et al. (2014)

# THE ATTENTION MODEL



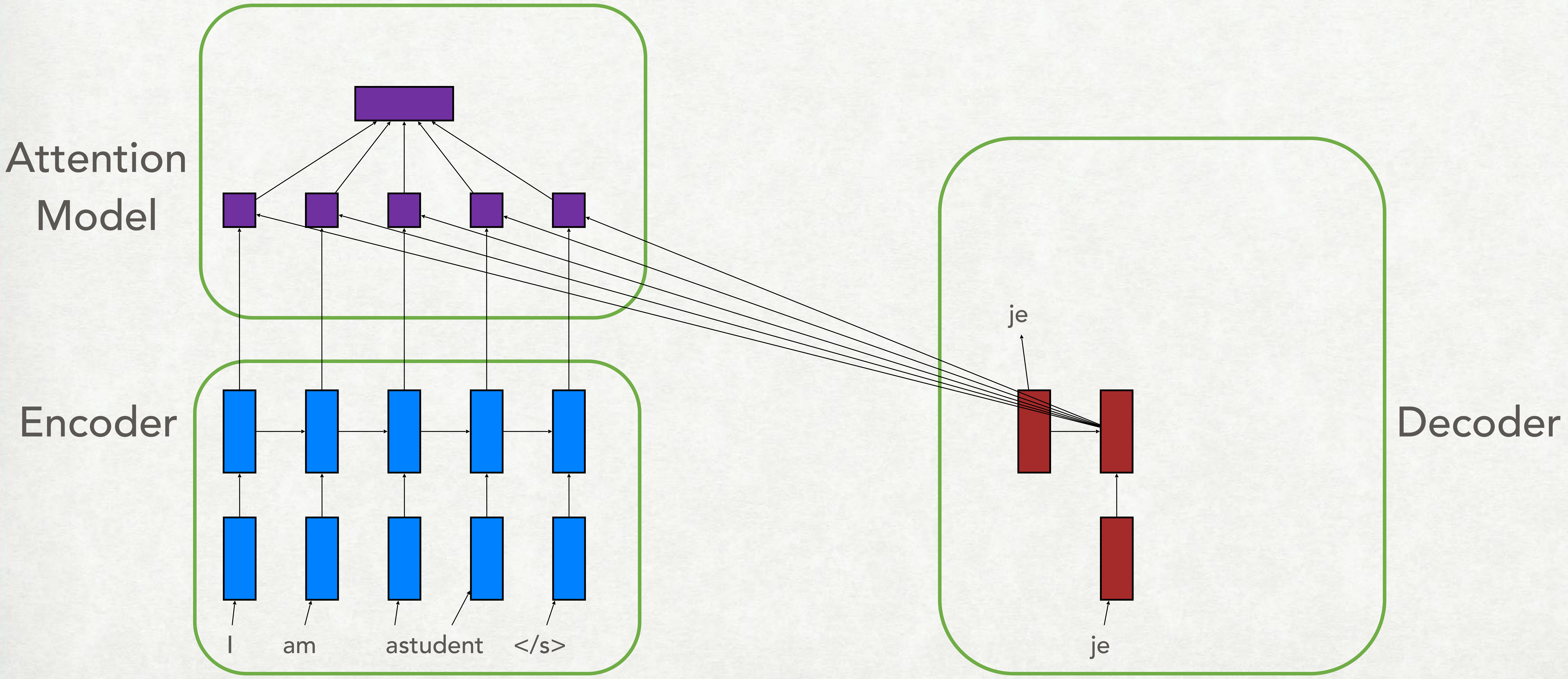

Bahdanau et al. (2014)

# THE ATTENTION MODEL



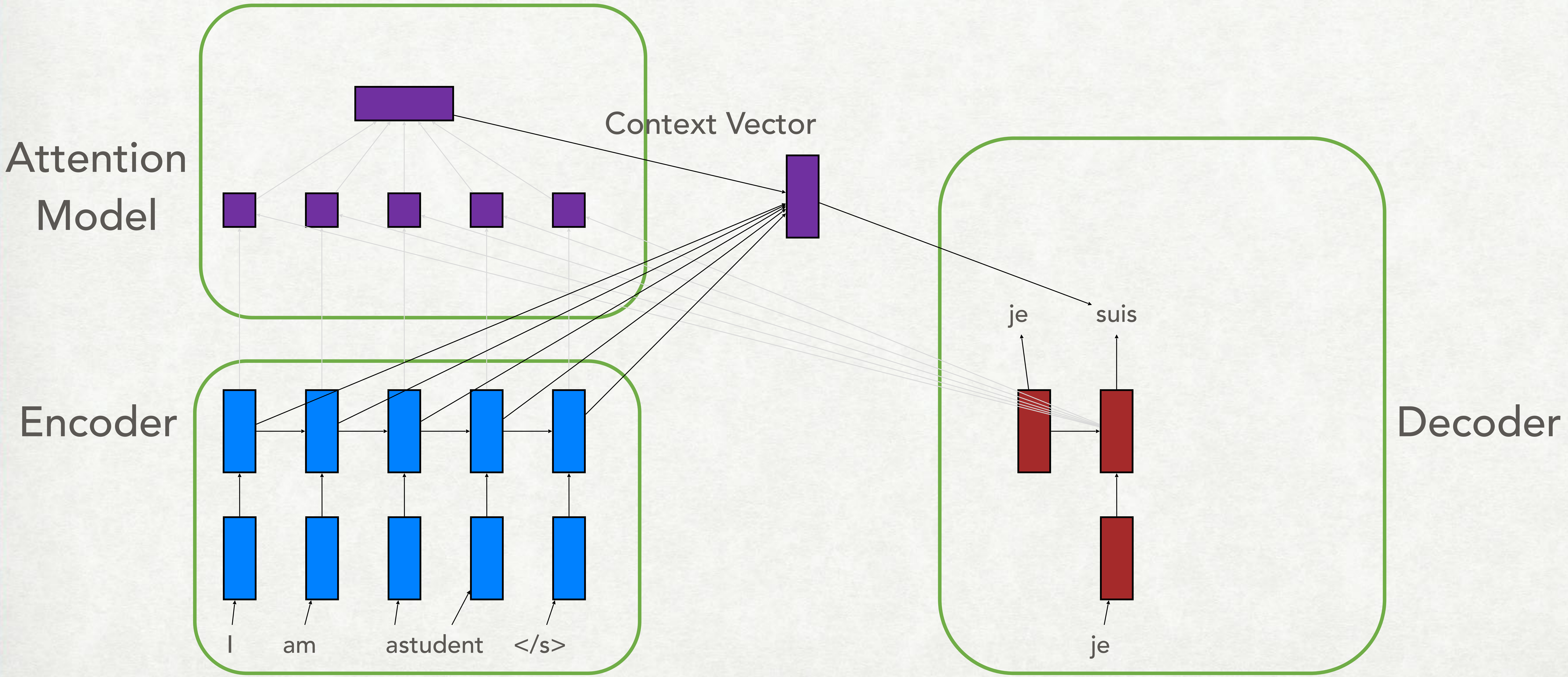

Bahdanau et al. (2014)

# THE ATTENTION MODEL



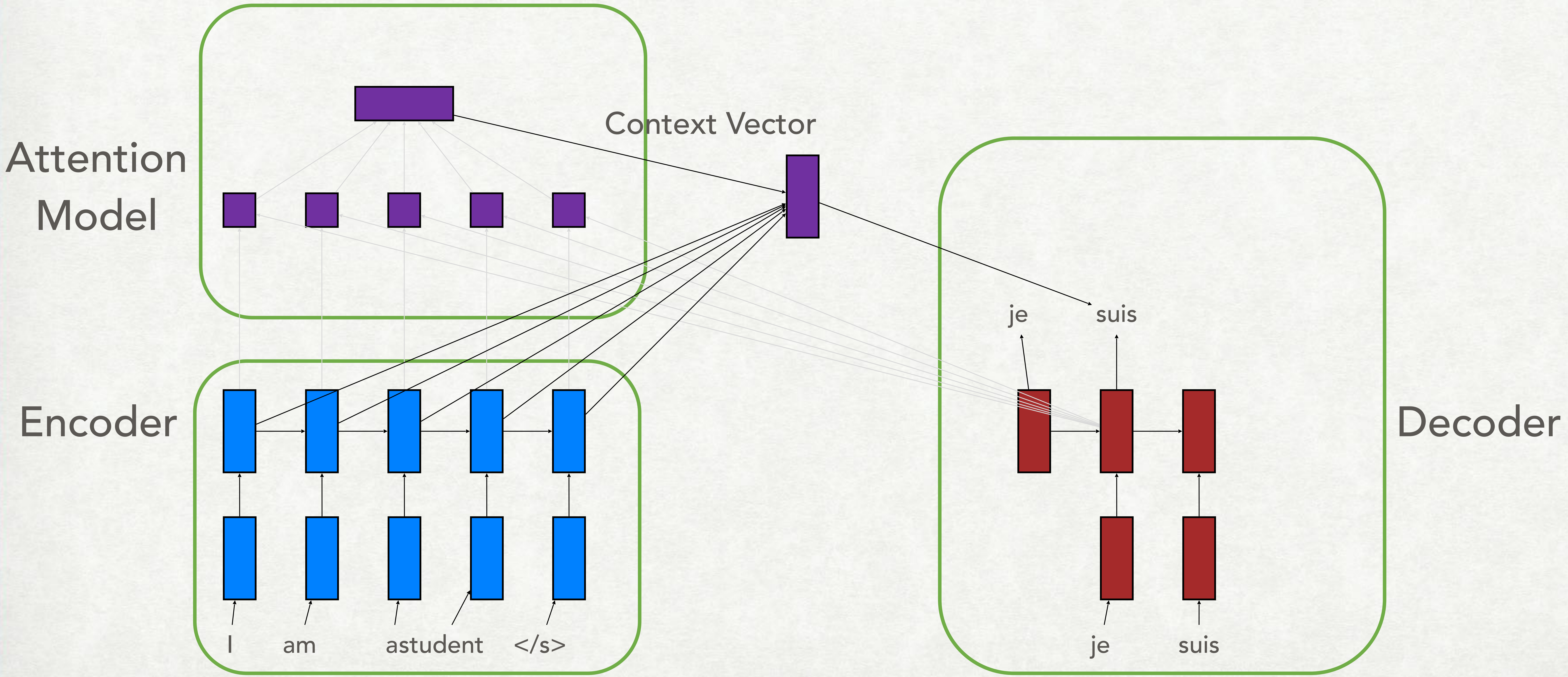

Bahdanau et al. (2014)

# THE ATTENTION MODEL



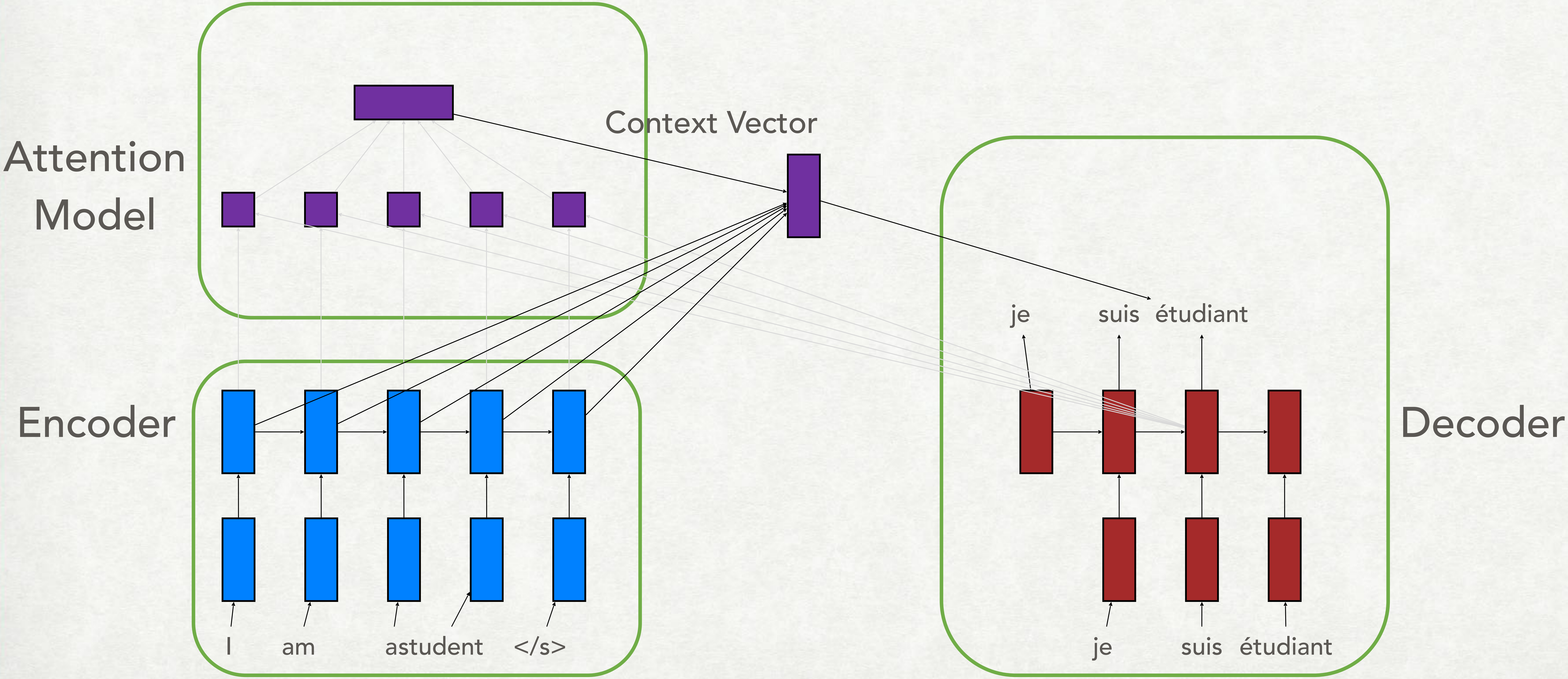

Bahdanau et al. (2014)

# THE ATTENTION MODEL



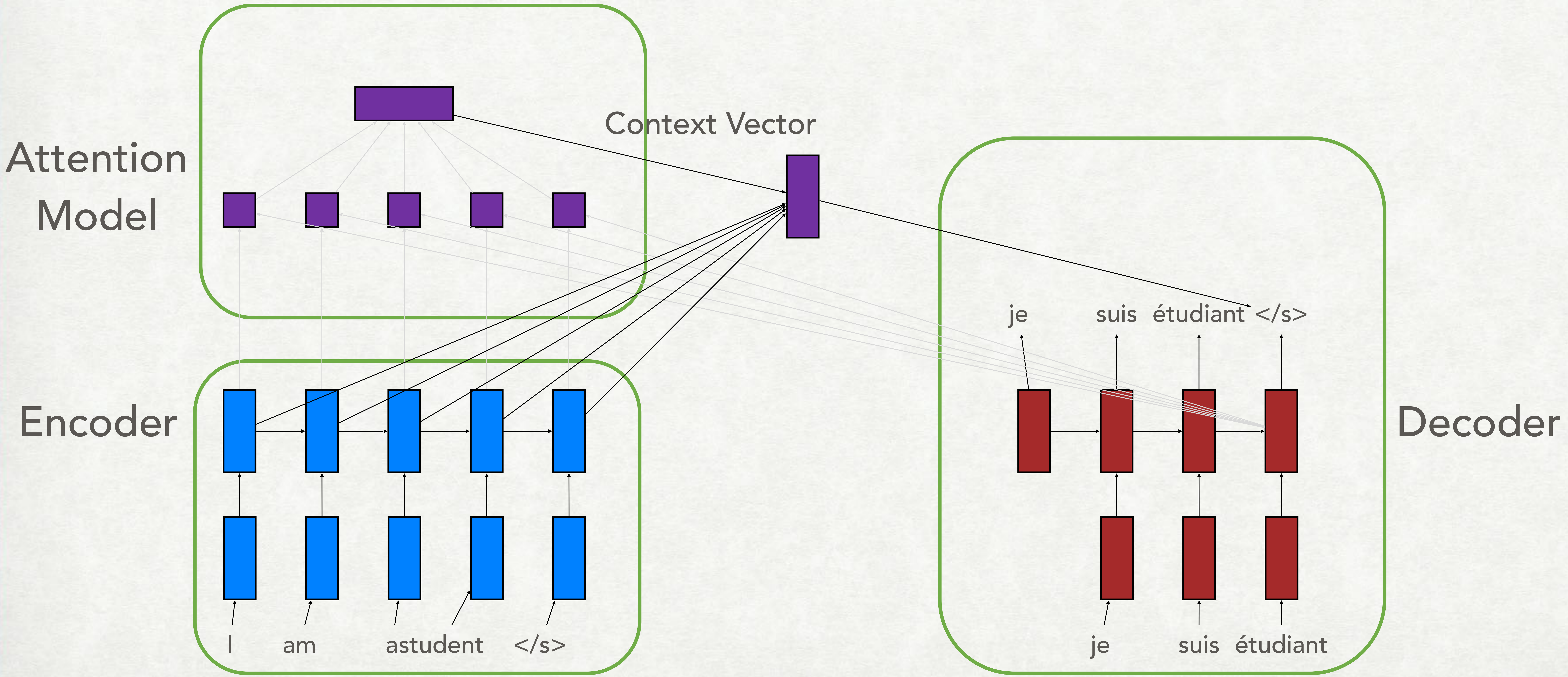

Bahdanau et al. (2014)

# THE ATTENTION MODEL




Bahdanau et al. (2014)

# THE ATTENTION MODEL




Bahdanau et al. (2014)

# CONVOLUTIONAL ENCODER-DECODER

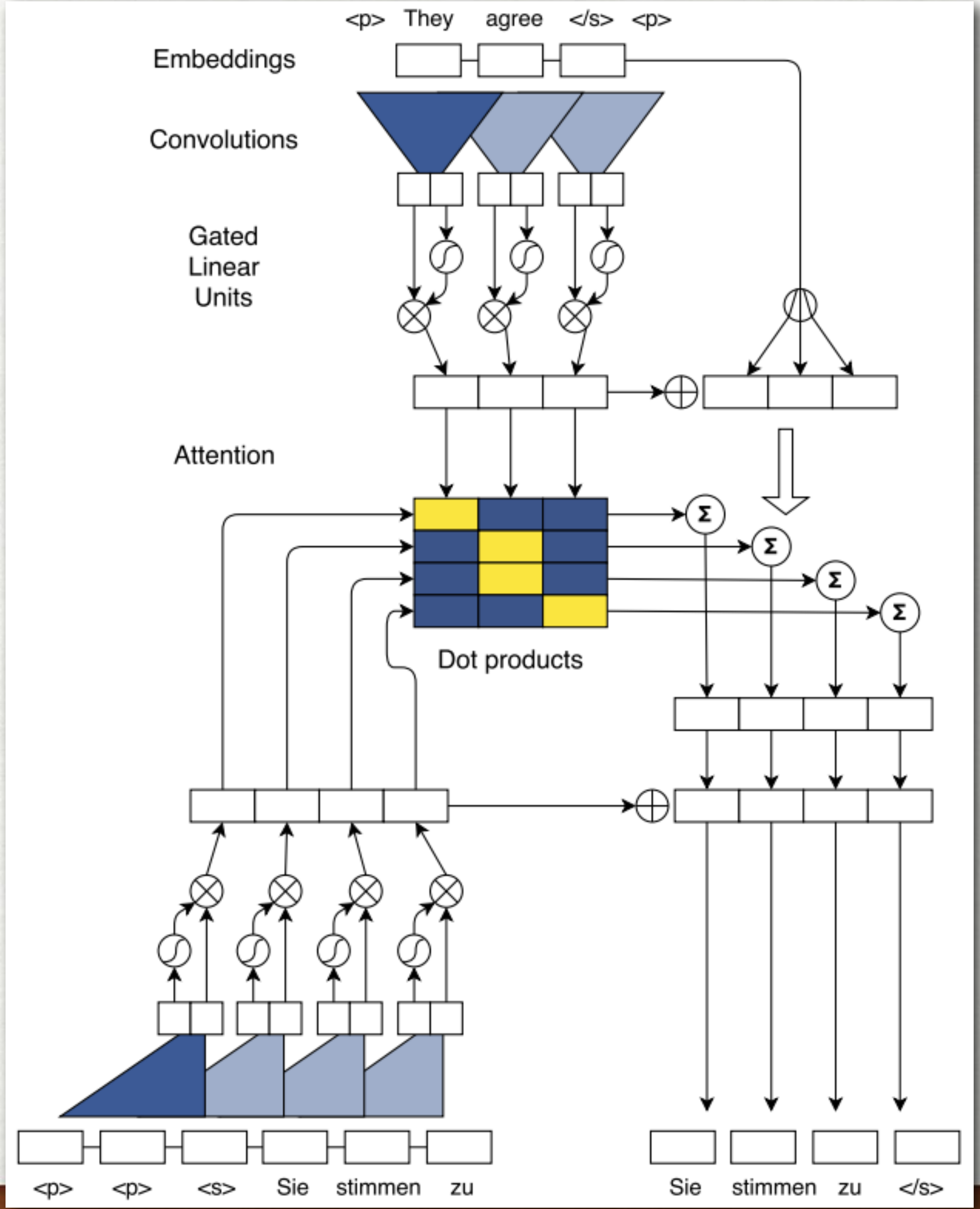
Gehring et. al 2017

CNN:

- encodes words within a fixed size window
- Parallel computation
- Shortest path to cover a wider range of words

RNN:

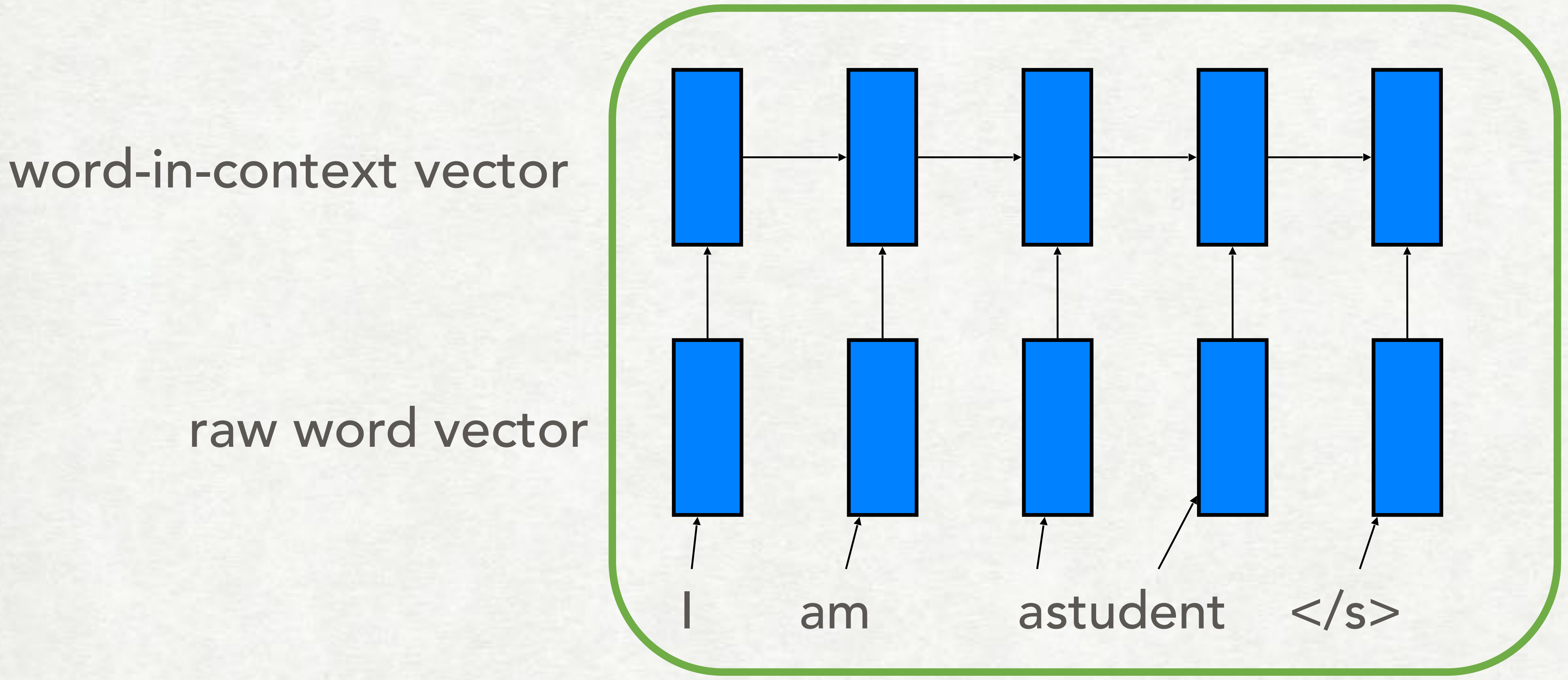
- sequentially encode a sentence from left to right
- Hard to parallelize



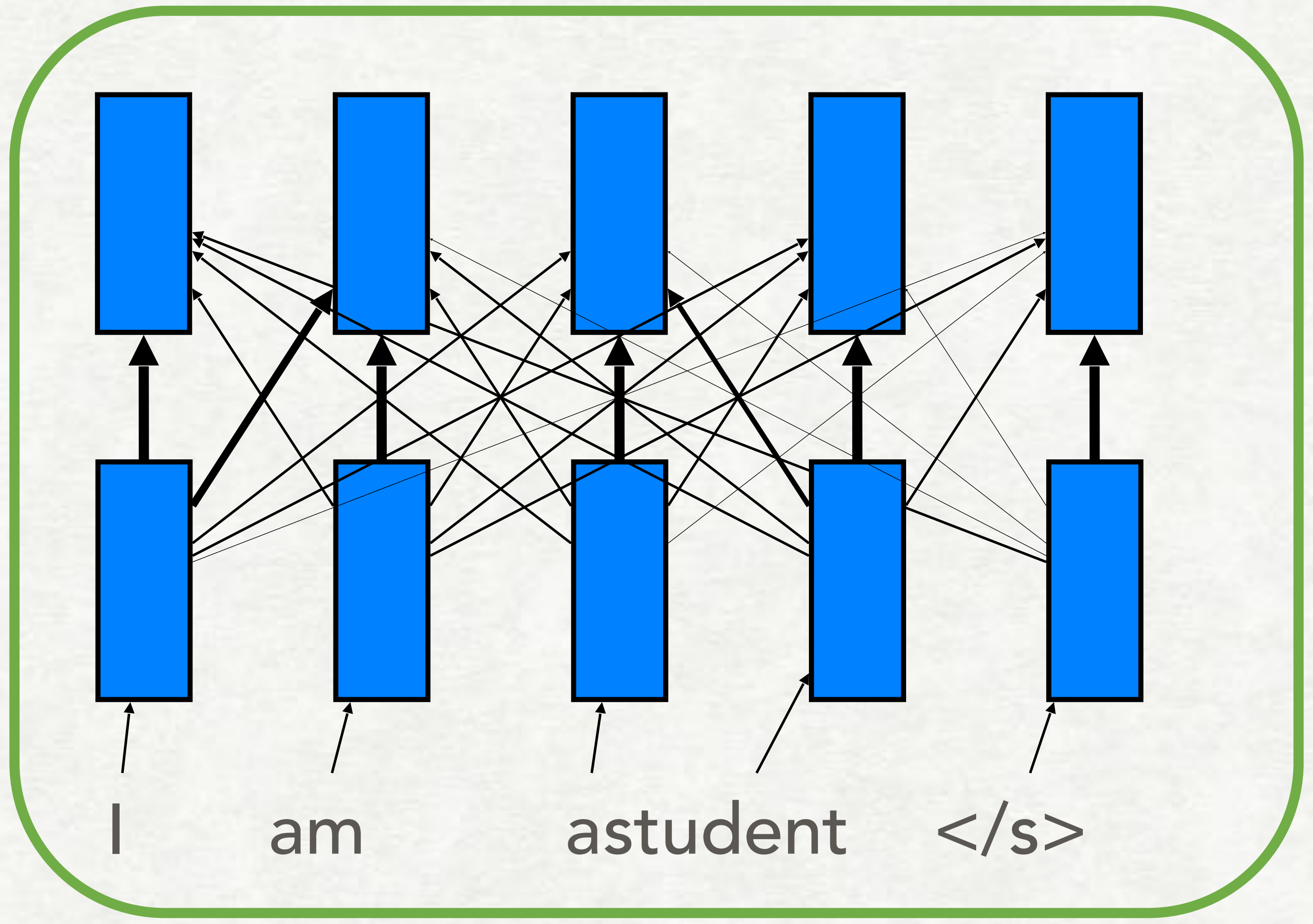
# THE TRANSFORMER

- Idea: Instead of using an RNN to encode the source sentence and the partial target sentence, use self-attention!

Standard RNN Encoder

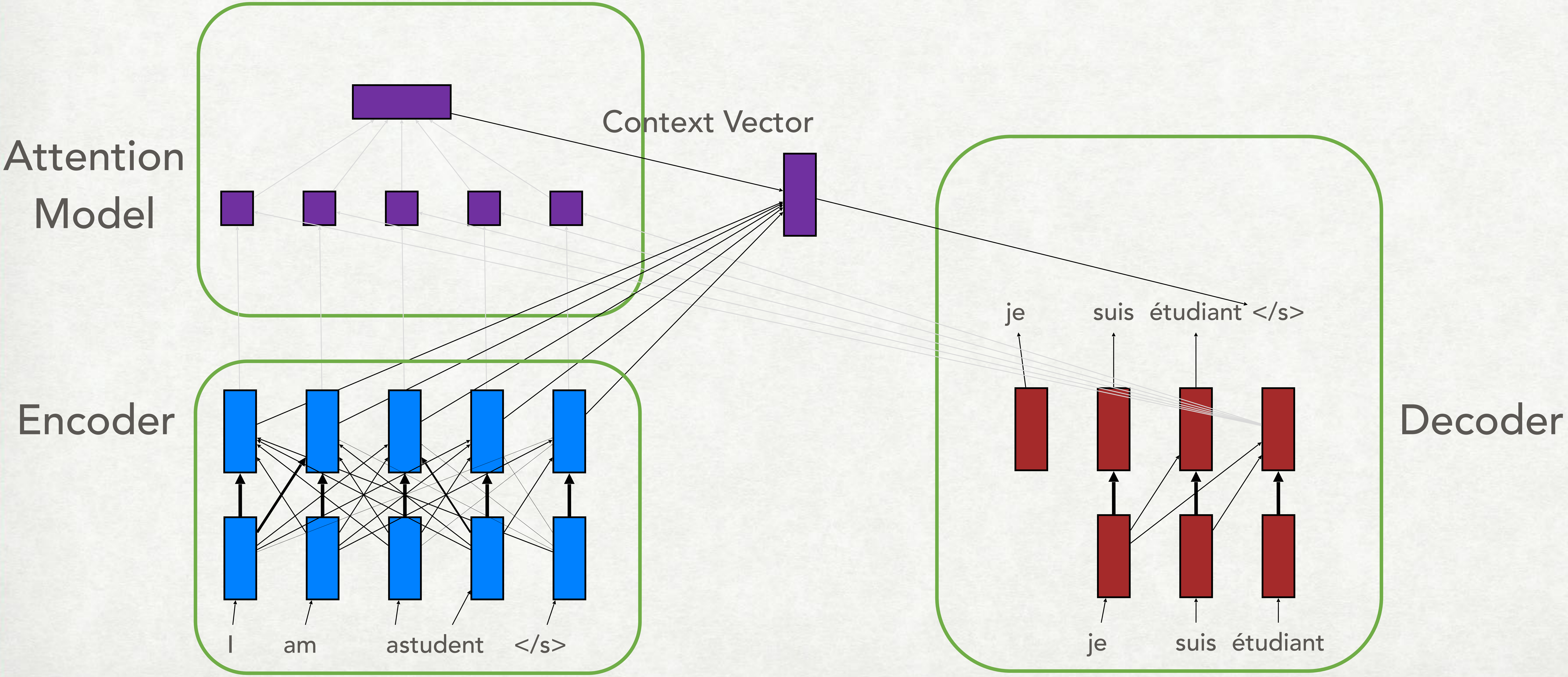



Self Attention Encoder




Vaswani et al. (2017)

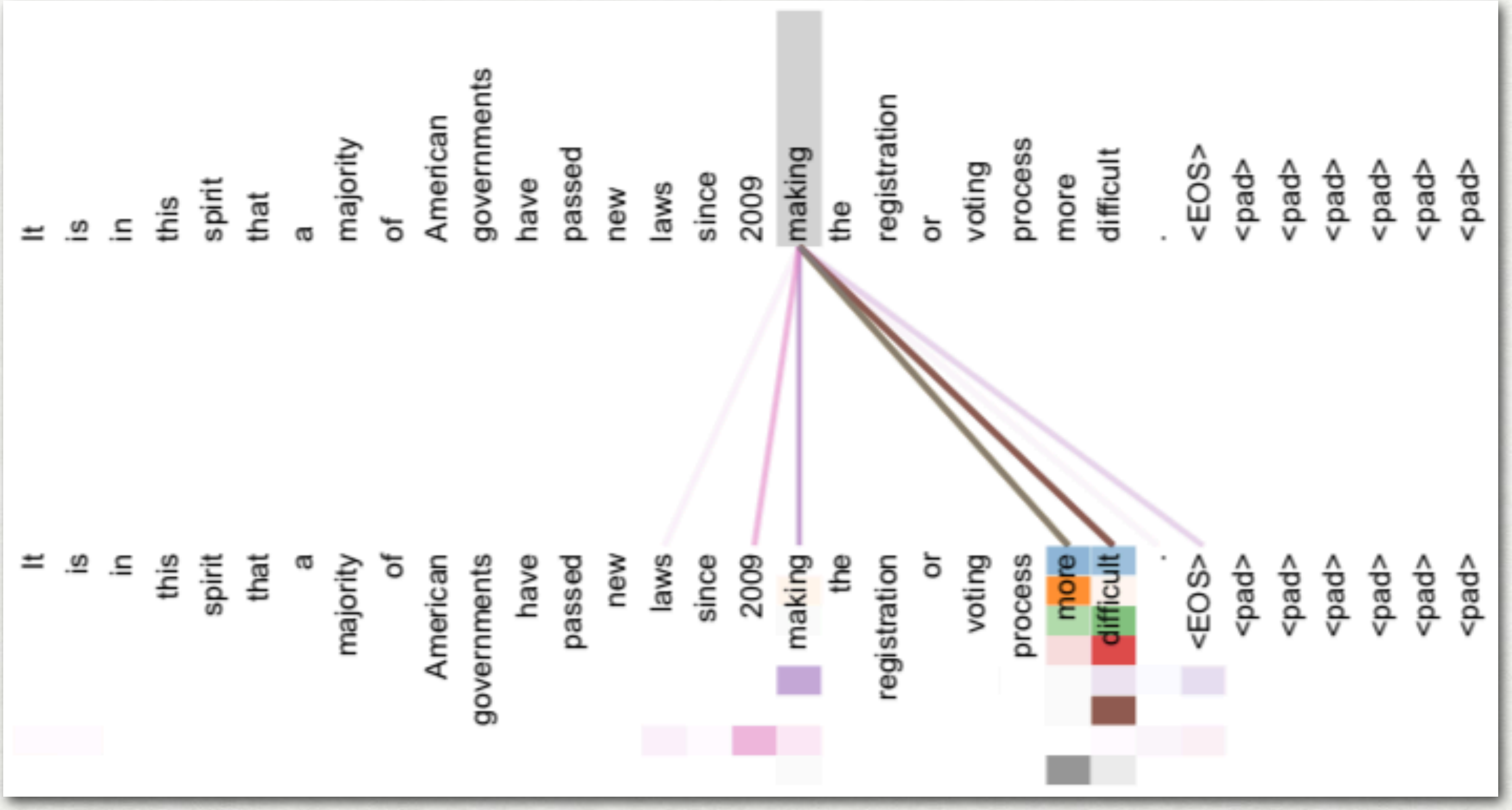
# THE TRANSFORMER




Vaswani et al. (2017)

# VISUALIZATION OF ATTENTION WEIGHT

- Self-attention weight can detect long-term dependencies within a sentence, e.g., make … more difficult



# THE TRANSFORMER

Computation is easily parallelizable

Shorter path from each target word to each source word → stronger gradient signals

Empirically stronger translation performance

Empirically trains substantially faster than more serial models

| Model | BLEU | | Training Cost (FLOPs) | |
|---|---|---|---|---|
| | EN-DE | EN-FR | EN-DE | EN-FR |
| ByteNet [17] | 23.75 | | | |
| Deep-Att + PosUnk [37] | | 39.2 | | $1.0 \cdot 10^{20}$ |
| GNMT + RL [36] | 24.6 | 39.92 | $2.3 \cdot 10^{19}$ | $1.4 \cdot 10^{20}$ |
| ConvS2S [9] | 25.16 | 40.46 | $9.6 \cdot 10^{18}$ | $1.5 \cdot 10^{20}$ |
| MoE [31] | 26.03 | 40.56 | $2.0 \cdot 10^{19}$ | $1.2 \cdot 10^{20}$ |
| Deep-Att + PosUnk Ensemble [37] | | 40.4 | | $8.0 \cdot 10^{20}$ |
| GNMT + RL Ensemble [36] | 26.30 | 41.16 | $1.8 \cdot 10^{20}$ | $1.1 \cdot 10^{21}$ |
| ConvS2S Ensemble [9] | 26.36 | **41.29** | $7.7 \cdot 10^{19}$ | $1.2 \cdot 10^{21}$ |
| Transformer (base model) | 27.3 | 38.1 | $\mathbf{3.3 \cdot 10^{18}}$ | |
| Transformer (big) | **28.4** | **41.0** | $2.3 \cdot 10^{19}$ | |

# CURRENT RESEARCH DIRECTIONS ON NEURAL MT

- Incorporation syntax into Neural MT
- Handling of morphologically rich languages
- Optimizing translation quality (instead of corpus probability)
- Multilingual models

# CURRENT RESEARCH DIRECTIONS ON NEURAL MT

- Incorporation syntax into Neural MT
- Handling of morphologically rich languages
- Optimizing translation quality (instead of corpus probability)
- Multilingual models

Exploring Massively Multilingual, Massive Neural Machine Translation

Friday, October 11, 2019

Posted by Ankur Bapna, Software Engineer and Orhan Firat, Research Scientist, Google Research

# CURRENT RESEARCH DIRECTIONS ON NEURAL MT

- Incorporation syntax into Neural MT
- Handling of morphologically rich languages
- Optimizing translation quality (instead of corpus probability)
- Multilingual models

Achieving Human Parity on Automatic
Chinese to English News Translation

Hany Hassan*, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, William Lewis, Mu Li, Shujie Liu, Tie-Yan Liu, Renqian Luo, Arul Menezes, Tao Qin, Frank Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia, Dongdong Zhang, Zhirui Zhang, and Ming Zhou

Microsoft AI & Research

# CURRENT RESEARCH DIRECTIONS ON NEURAL MT

- Incorporation syntax into Neural MT
- Handling of morphologically rich languages
- Optimizing translation quality (instead of corpus probability)
- Multilingual models
- Document-level translation

**Has Machine Translation Achieved Human Parity?**
**A Case for Document-level Evaluation**

**Samuel Läubli**[1] **Rico Sennrich**[1,2] **Martin Volk**[1]

[1]Institute of Computational Linguistics, University of Zurich
{laeubli,volk}@cl.uzh.ch

# CURRENT RESEARCH DIRECTIONS ON NEURAL MT

- Incorporation syntax into Neural MT
- Handling of morphologically rich languages
- Optimizing translation quality (instead of corpus probability)
- Multilingual models
- Document-level translation
- Domain adaptation and robustness