自然语言处理

week-12

凌震华

2024年5月30日

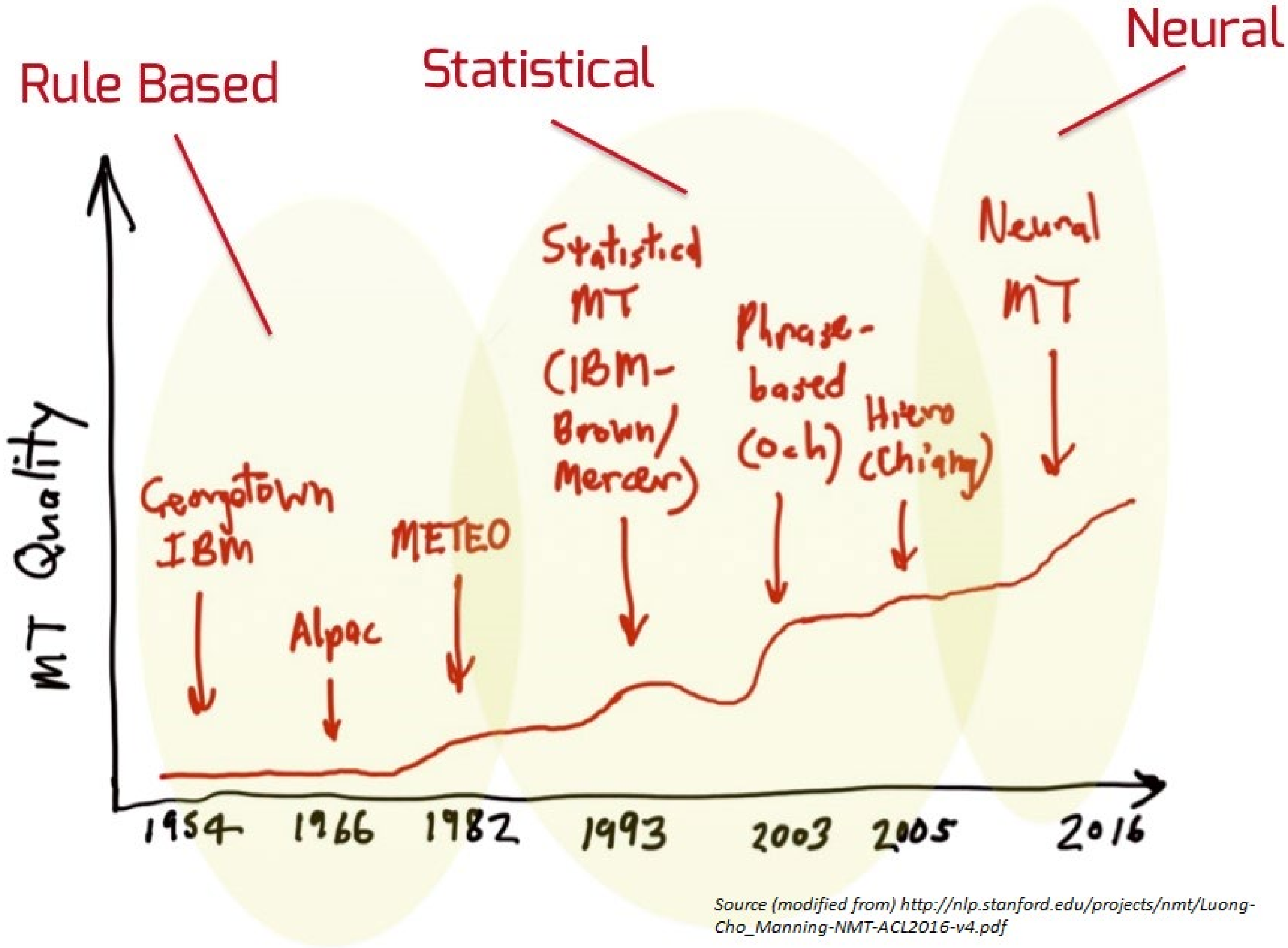
□Machine Translation

□MT Evaluation

□Statistical Machine Translation

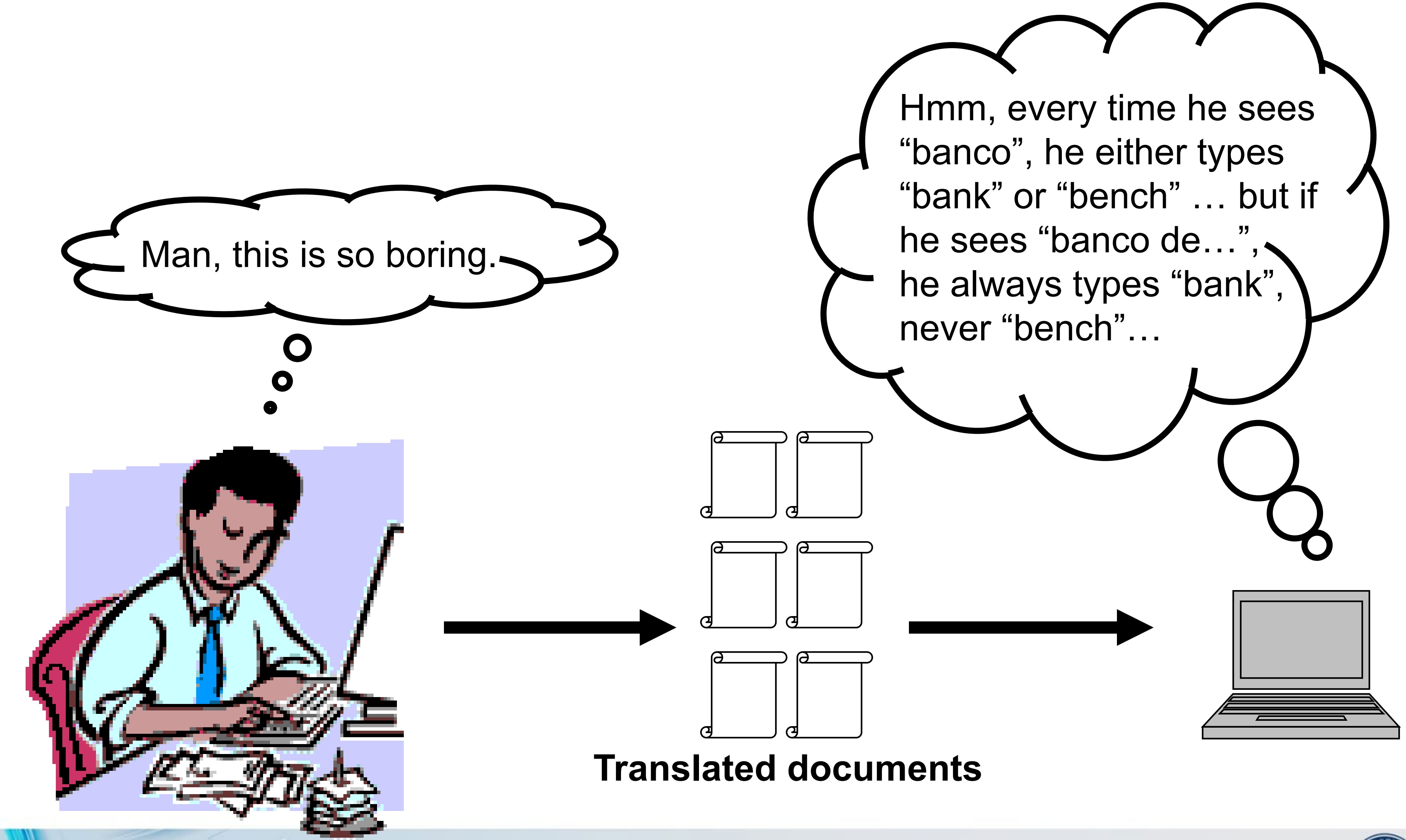
□Neural Machine Translation

Machine Translation Progress



Source (modified from) http://nlp.stanford.edu/projects/nmt/Luong-Cho_Manning-NMT-ACL2016-v4.pdf

Statistical Machine Translation


Man, this is so boring.
Hmm, every time he sees “banco”, he either types “bank” or “bench” … but if he sees “banco de…”, he always types “bank”, never “bench”…
Translated documents

Current Approaches

- Same old noisy channel model…
- If we're translating French to English, the French we're seeing is just a weird garbled 篡改 version of English
- There must have been some process that generated the French from the original English
- The key is to decode the garbles back into the original English by…
- argmax P(E | F) by Bayes
- A very old idea

Warren Weaver (1947)

When I look at an article in Russian, I say to myself: This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok **farok** ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok **farok** izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok .

1b. at-voon bichat dat .

2a. ok-drubel ok-voon anok plok sprok .

2b. at-drubel at-voon pippat rrat dat .

3a. erok sprok izok hihok ghirok .

3b. totat dat arrat vat hilat .

4a. ok-voon anok drok brok jok .

4b. at-voon krat pippat sat lat .

5a. wiwok **farok** izok stok .

5b. totat jjat quat cat .

6a. lalok sprok izok jok stok .

6b. wat dat krat quat cat .

7a. lalok **farok** ororok lalok sprok izok enemok .

7b. wat jjat bichat wat dat vat eneat .

8a. lalok brok anok plok nok .

8b. iat lat pippat rrat nnat .

9a. wiwok nok izok kantok ok-yurp .

9b. totat nnat quat oloat at-yurp .

10a. lalok mok nok yorok ghirok clok .

10b. wat nnat gat mat bat hilat .

11a. lalok nok crrrok hihok yorok zanzanok .

11b. wat nnat arrat mat zanzanat .

12a. lalok rarok nok izok hihok mok .

12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok .

1b. at-voon bichat dat .

2a. ok-drubel ok-voon anok plok sprok .

2b. at-drubel at-voon pippat rrat dat .

3a. erok sprok izok hihok ghirok .

3b. totat dat arrat vat hilat .

4a. ok-voon anok drok brok jok .

4b. at-voon krat pippat sat lat .

5a. wiwok **farok** izok stok .

5b. totat **jjat** quat cat .

6a. lalok sprok izok jok stok .

6b. wat dat krat quat cat .

7a. lalok **farok** ororok lalok sprok izok enemok .

7b. wat **jjat** bichat wat dat vat eneat .

8a. lalok brok anok plok nok .

8b. iat lat pippat rrat nnat .

9a. wiwok nok izok kantok ok-yurp .

9b. totat nnat quat oloat at-yurp .

10a. lalok mok nok yorok ghirok clok .

10b. wat nnat gat mat bat hilat .

11a. lalok nok crrrok hihok yorok zanzanok .

11b. wat nnat arrat mat zanzanat .

12a. lalok rarok nok izok hihok mok .

12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok .

1b. at-voon bichat dat .

2a. ok-drubel ok-voon anok plok sprok .

2b. at-drubel at-voon pippat rrat dat .

3a. erok sprok izok hihok ghirok .

3b. totat dat arrat vat hilat .

4a. ok-voon anok drok brok jok .

4b. at-voon krat pippat sat lat .

5a. wiwok **farok** izok stok .

5b. totat jjat quat cat .

6a. lalok sprok izok jok stok .

6b. wat dat krat quat cat .

7a. lalok **farok** ororok lalok sprok izok enemok .

7b. wat jjat bichat wat dat vat eneat .

8a. lalok brok anok plok nok .

8b. iat lat pippat rrat nnat .

9a. wiwok nok izok kantok ok-yurp .

9b. totat nnat quat oloat at-yurp .

10a. lalok mok nok yorok ghirok clok .

10b. wat nnat gat mat bat hilat .

11a. lalok nok **crrrok** hihok yorok zanzanok .

???

11b. wat nnat arrat mat zanzanat .

12a. lalok rarok nok izok hihok mok .

12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok **hihok** ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok **hihok** yorok zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok **hihok** mok . 12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: **farok** crrrok **hihok** yorok clok kantok ok-yurp

1a. ok-voon ororok sprok .

1b. at-voon bichat dat .

2a. ok-drubel ok-voon anok plok sprok .

2b. at-drubel at-voon pippat rrat dat .

3a. erok sprok izok **hihok** ghirok .

3b. totat dat arrat vat hilat .

4a. ok-voon anok drok brok jok .

4b. at-voon krat pippat sat lat .

5a. wiwok farok izok stok .

5b. totat jjat quat cat .

6a. lalok sprok izok jok stok .

6b. wat dat krat quat cat .

7a. lalok farok ororok lalok sprok izok enemok .

7b. wat jjat bichat wat dat vat eneat .

8a. lalok brok anok plok nok .

8b. iat lat pippat rrat nnat .

9a. wiwok nok izok kantok ok-yurp .

9b. totat nnat quat oloat at-yurp .

10a. lalok mok nok **yorok** ghirok clok .

10b. wat nnat gat mat bat hilat .

11a. lalok nok crrrok **hihok** **yorok** zanzanok .

11b. wat nnat arrat mat zanzanat .

12a. lalok rarok nok izok **hihok** mok .

12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: **farok** crrrok **hihok yorok** clok kantok ok-yurp

1a. ok-voon ororok sprok .

1b. at-voon bichat dat .

2a. ok-drubel ok-voon anok plok sprok .

2b. at-drubel at-voon pippat rrat dat .

3a. erok sprok izok hihok ghirok .

3b. totat dat arrat vat hilat .

4a. ok-voon anok drok brok jok .

4b. at-voon krat pippat sat lat .

5a. wiwok farok izok stok .

5b. totat jjat quat cat .

6a. lalok sprok izok jok stok .

6b. wat dat krat quat cat .

7a. lalok farok ororok lalok sprok izok enemok .

7b. wat jjat bichat wat dat vat eneat .

8a. lalok brok anok plok nok .

8b. iat lat pippat rrat nnat .

9a. wiwok nok izok kantok ok-yurp .

9b. totat nnat quat oloat at-yurp .

10a. lalok mok nok yorok ghirok **clok** .

10b. wat nnat gat mat bat hilat .

11a. lalok nok crrrok hihok yorok zanzanok .

11b. wat nnat arrat mat zanzanat .

12a. lalok rarok nok izok hihok mok .

12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: **farok** crrrok **hihok yorok** clok kantok ok-yurp

1a. ok-voon ororok sprok .

1b. at-voon bichat dat .

2a. ok-drubel ok-voon anok plok sprok .

2b. at-drubel at-voon pippat rrat dat .

3a. erok sprok izok hihok ghirok .

3b. totat dat arrat vat hilat .

4a. ok-voon anok drok brok jok .

4b. at-voon krat pippat sat lat .

5a. wiwok farok izok stok .

5b. totat jjat quat cat .

6a. lalok sprok izok jok stok .

6b. wat dat krat quat cat .

7a. lalok farok ororok lalok sprok izok enemok .

7b. wat jjat bichat wat dat vat eneat .

8a. lalok brok anok plok nok .

8b. iat lat pippat rrat nnat .

9a. wiwok nok izok kantok ok-yurp .

9b. totat nnat quat oloat at-yurp .

10a. lalok mok nok yorok ghirok **clok** .

???

10b. wat nnat gat mat bat hilat .

11a. lalok nok crrrok hihok yorok zanzanok .

11b. wat nnat arrat mat zanzanat .

12a. lalok rarok nok izok hihok mok .

12b. wat nnat forat arrat vat gat .

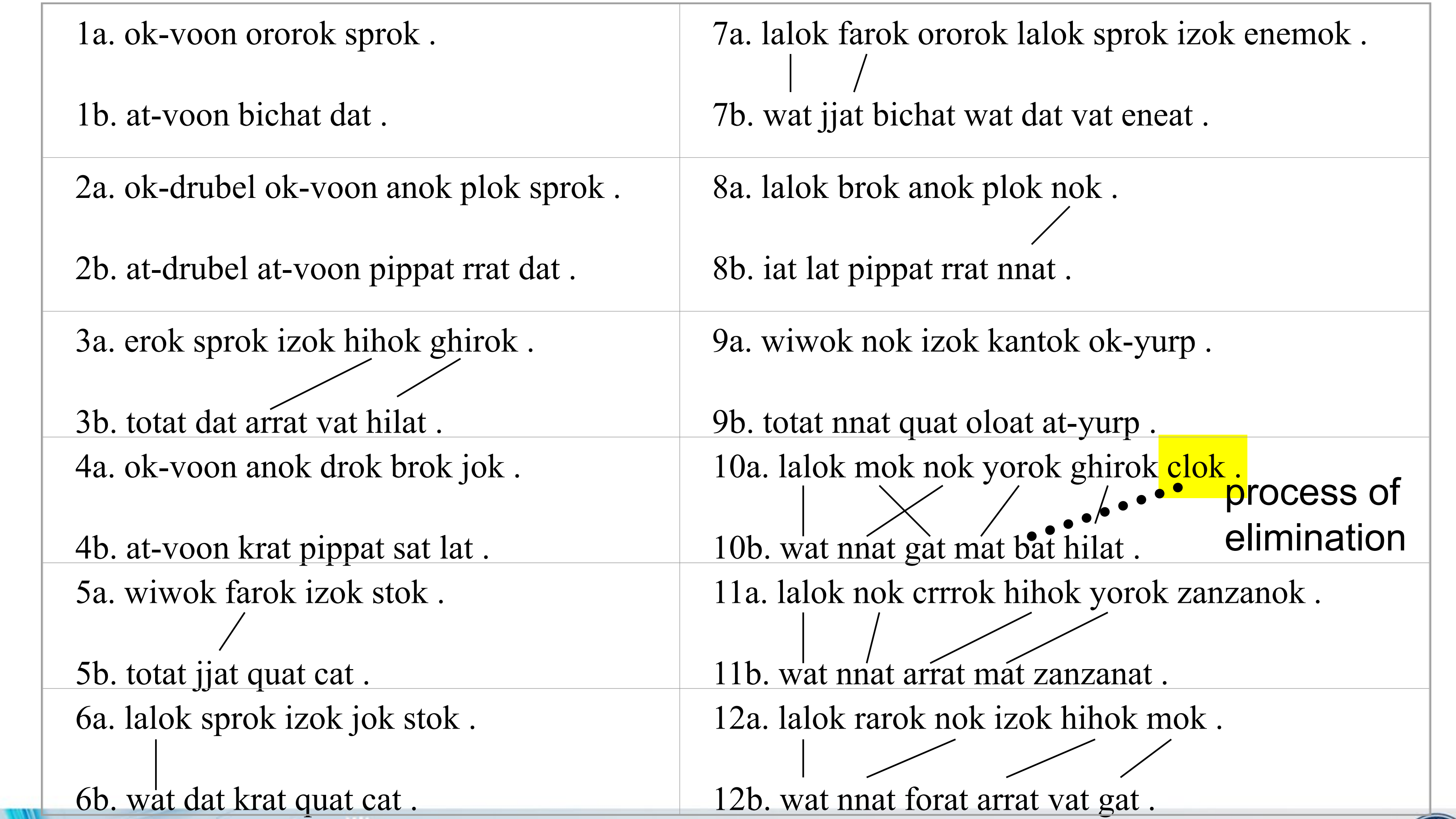
Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

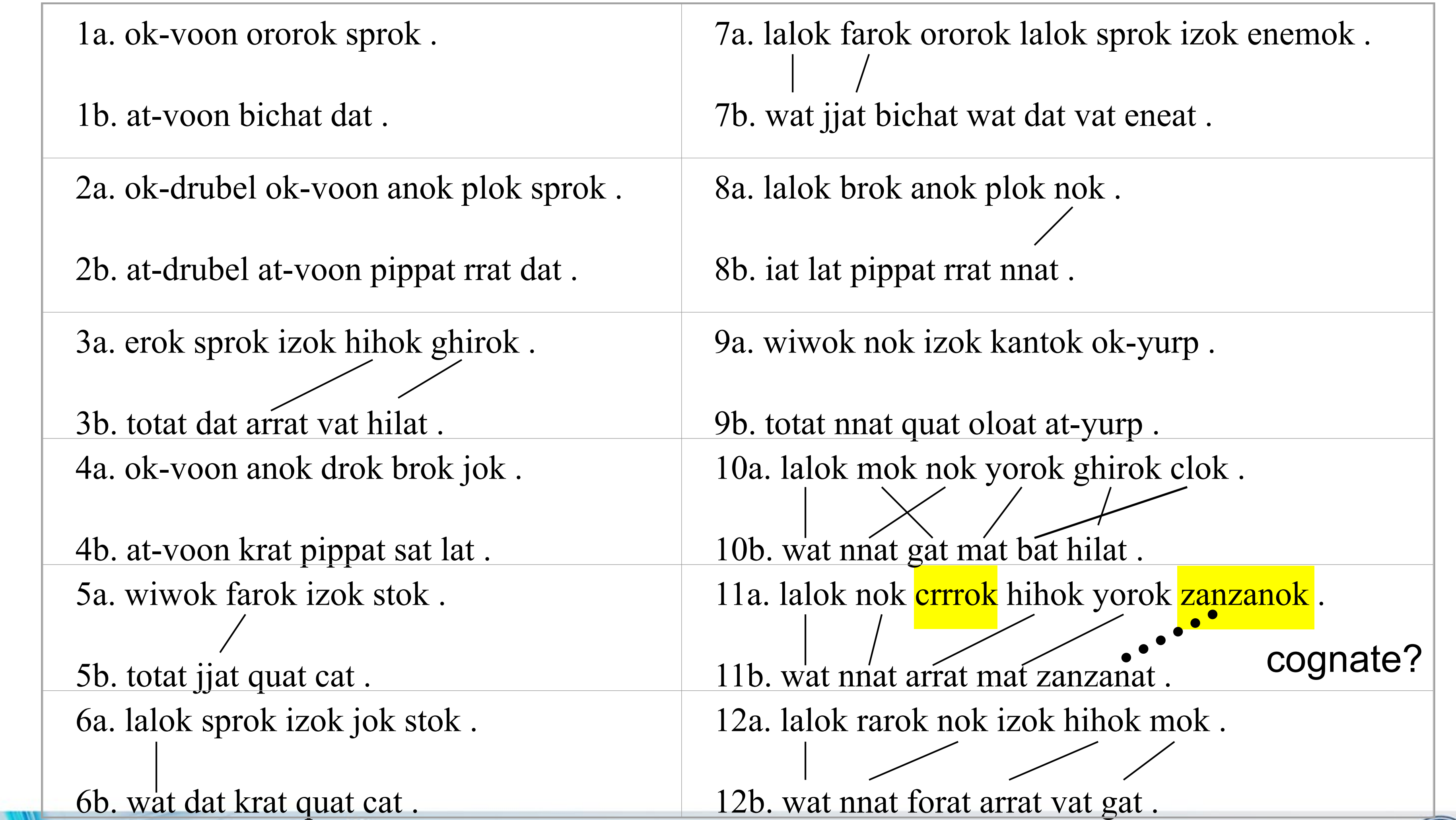


1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

process of elimination

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

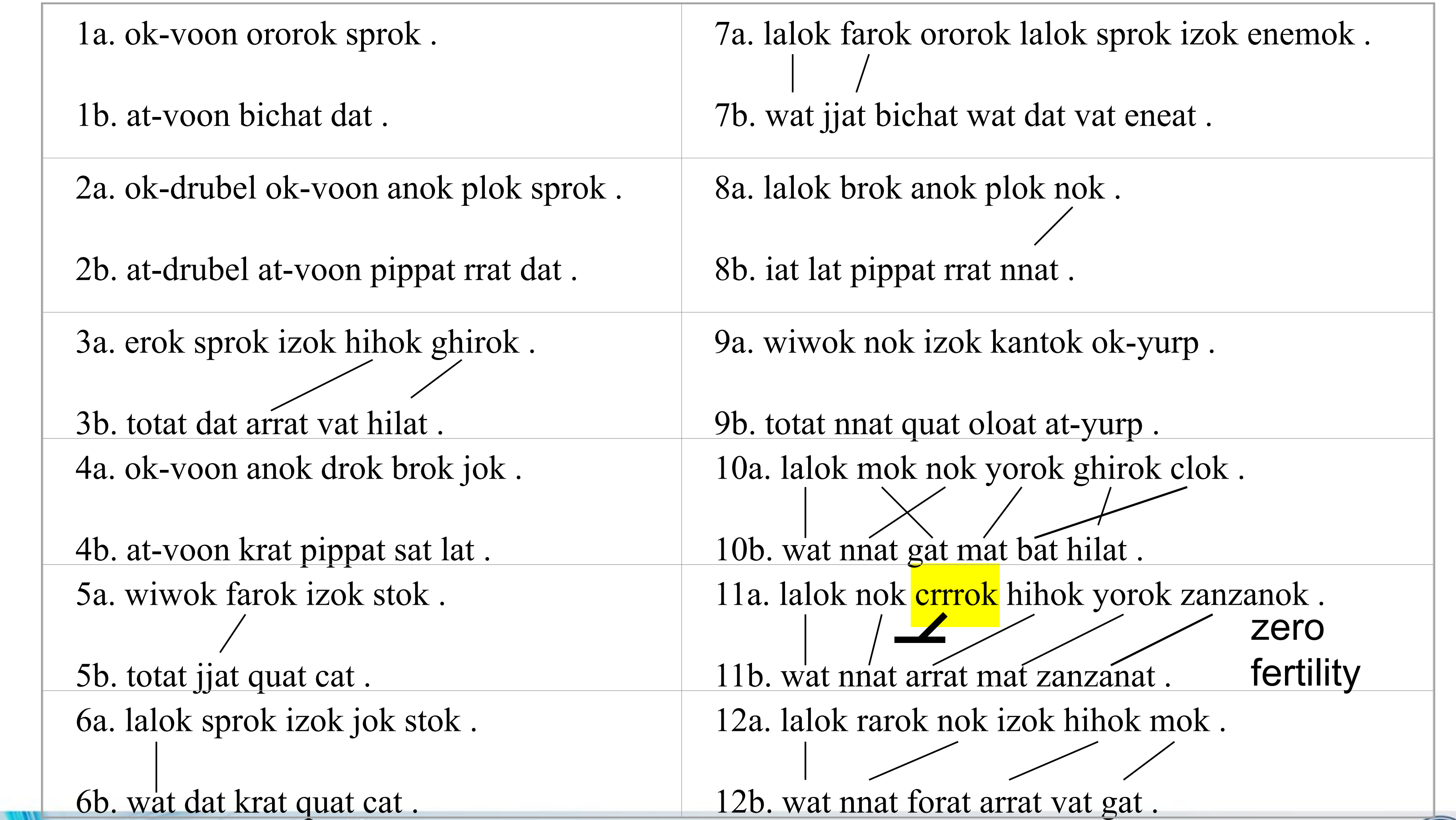


1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

cognate?

Centauri/Arcturan [Knight, 1997]

Your assignment, put these words in order: { jjat, arrat, mat, bat, oloat, at-yurp }



1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

zero fertility

Spanish/English text

Translate: Clients do not sell pharmaceuticals in Europe.

1a. Garcia and associates .
1b. Garcia y asociados .

2a. Carlos Garcia has three associates .
2b. Carlos Garcia tiene tres asociados .

3a. his associates are not strong .
3b. sus asociados no son fuertes .

4a. Garcia has a company also .
4b. Garcia tambien tiene una empresa .

5a. its clients are angry .
5b. sus clientes estan enfadados .

6a. the associates are also angry .
6b. los asociados tambien estan enfadados .

7a. the clients and the associates are enemies .
7b. los clients y los asociados son enemigos .

8a. the company has three groups .
8b. la empresa tiene tres grupos .

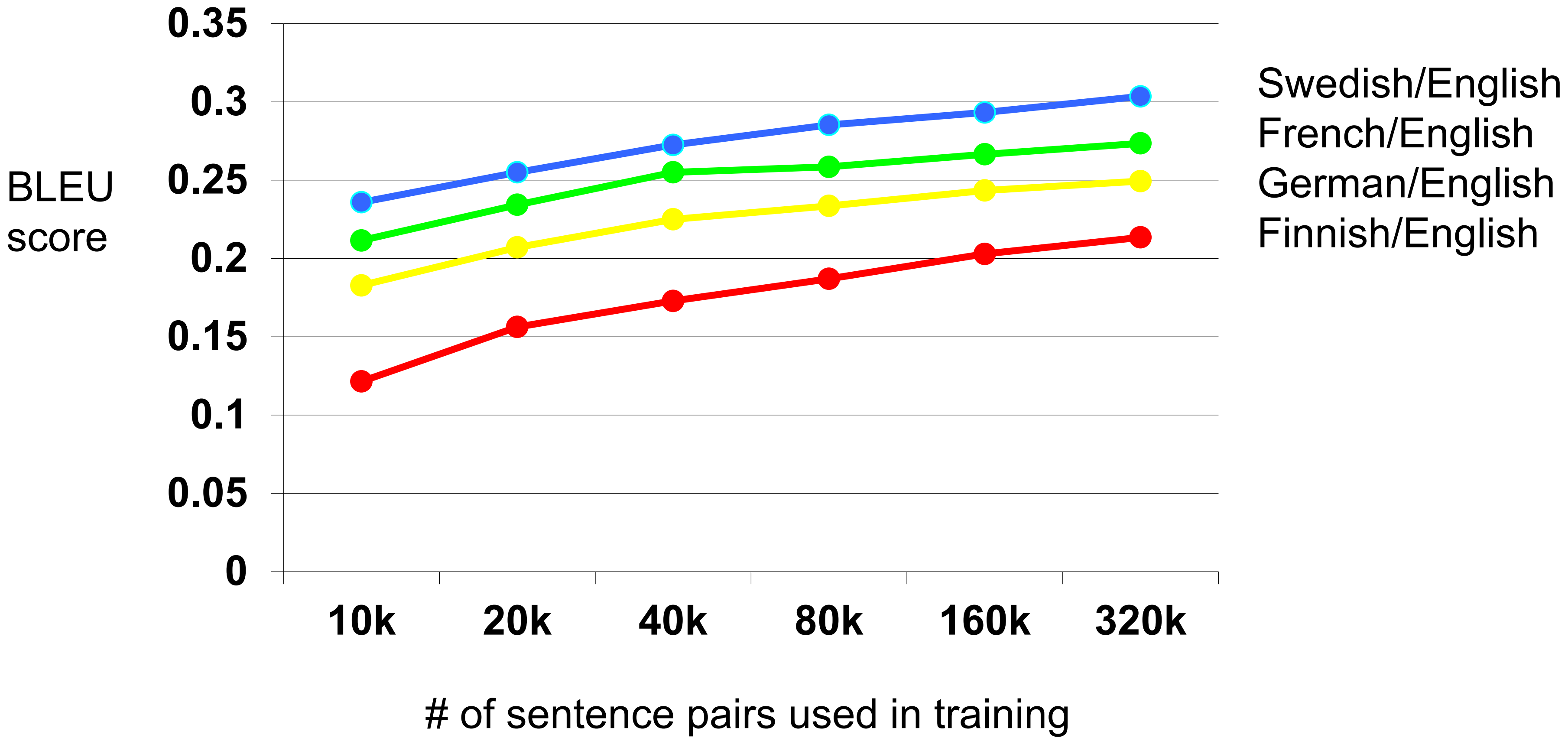
9a. its groups are in Europe .
9b. sus grupos estan en Europa .

10a. the modern groups sell strong pharmaceuticals .
10b. los grupos modernos venden medicinas fuertes .

11a. the groups do not sell zenzanine .
11b. los grupos no venden zanzanina .

12a. the small groups are not modern .
12b. los grupos pequenos no son modernos .

Sample Learning Curves



Experiments by Philipp Koehn

□Machine Translation

□MT Evaluation

□Statistical Machine Translation

□Neural Machine Translation

MT Evaluation

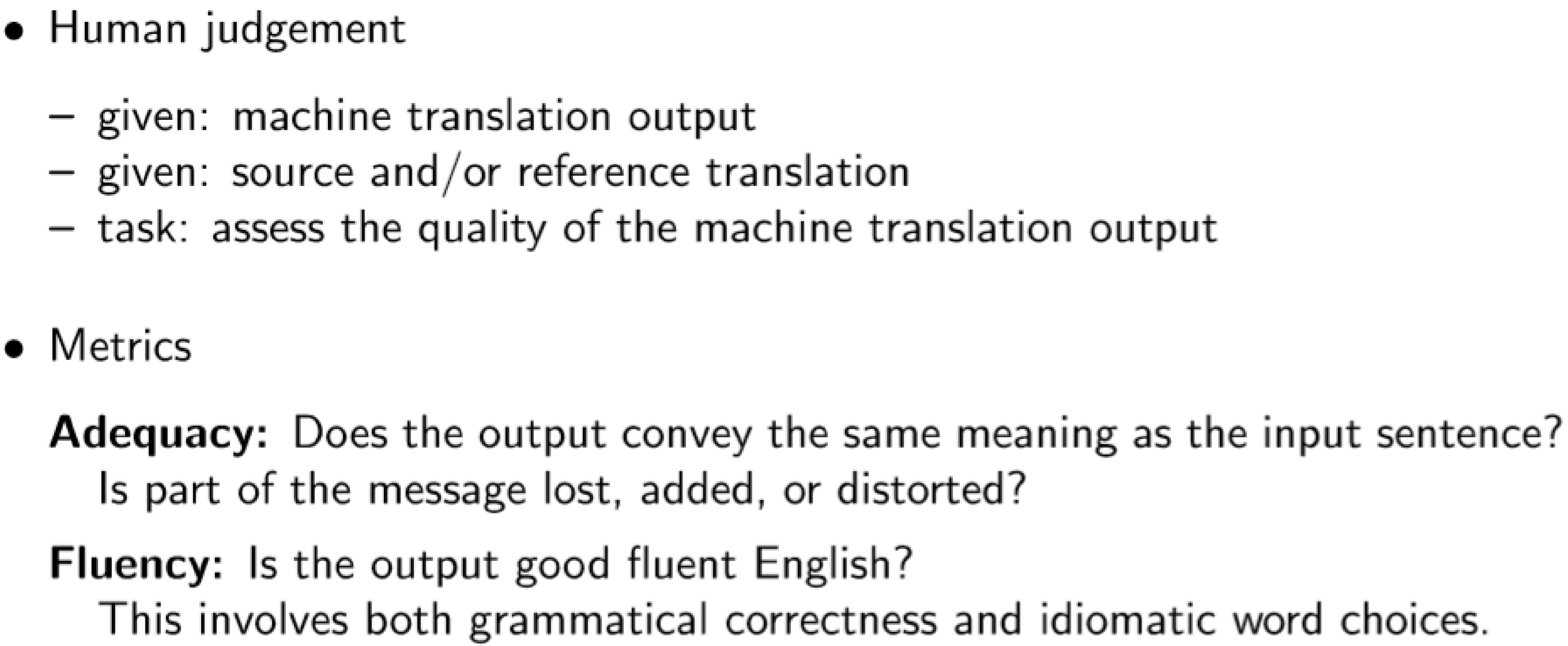
Traditionally difficult because there is no single “right answer”.

20 human translators will translate the same sentence 20 different ways.

Evaluation Metrics

- subjective judgments by human evaluators
- automatic evaluation metrics
- task-based evaluation, e.g.:
  - how much post-editing effort?
  - does information come across?

Adequacy 忠实度 and Fluency 流畅度



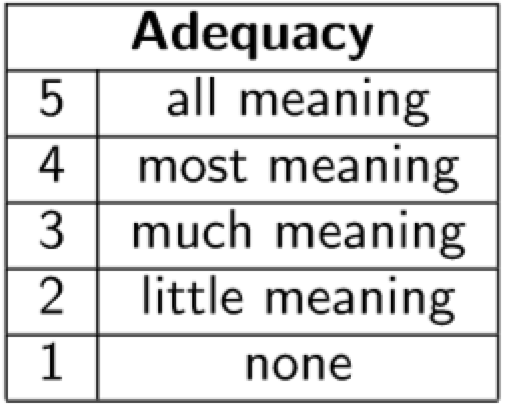
- Human judgement
  - given: machine translation output
  - given: source and/or reference translation
  - task: assess the quality of the machine translation output

- Metrics

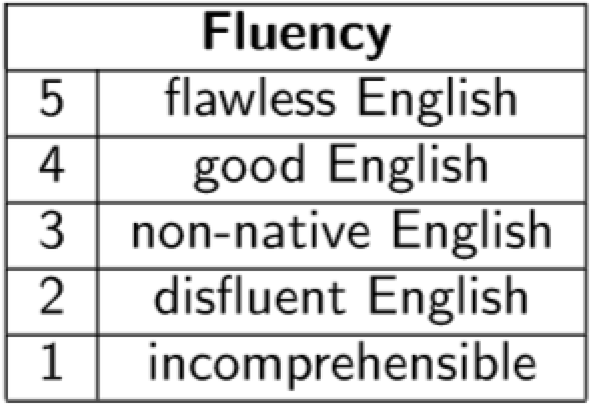
  Adequacy: Does the output convey the same meaning as the input sentence?
  Is part of the message lost, added, or distorted?

  Fluency: Is the output good fluent English?
  This involves both grammatical correctness and idiomatic word choices.

Adequacy and Fluency Scales

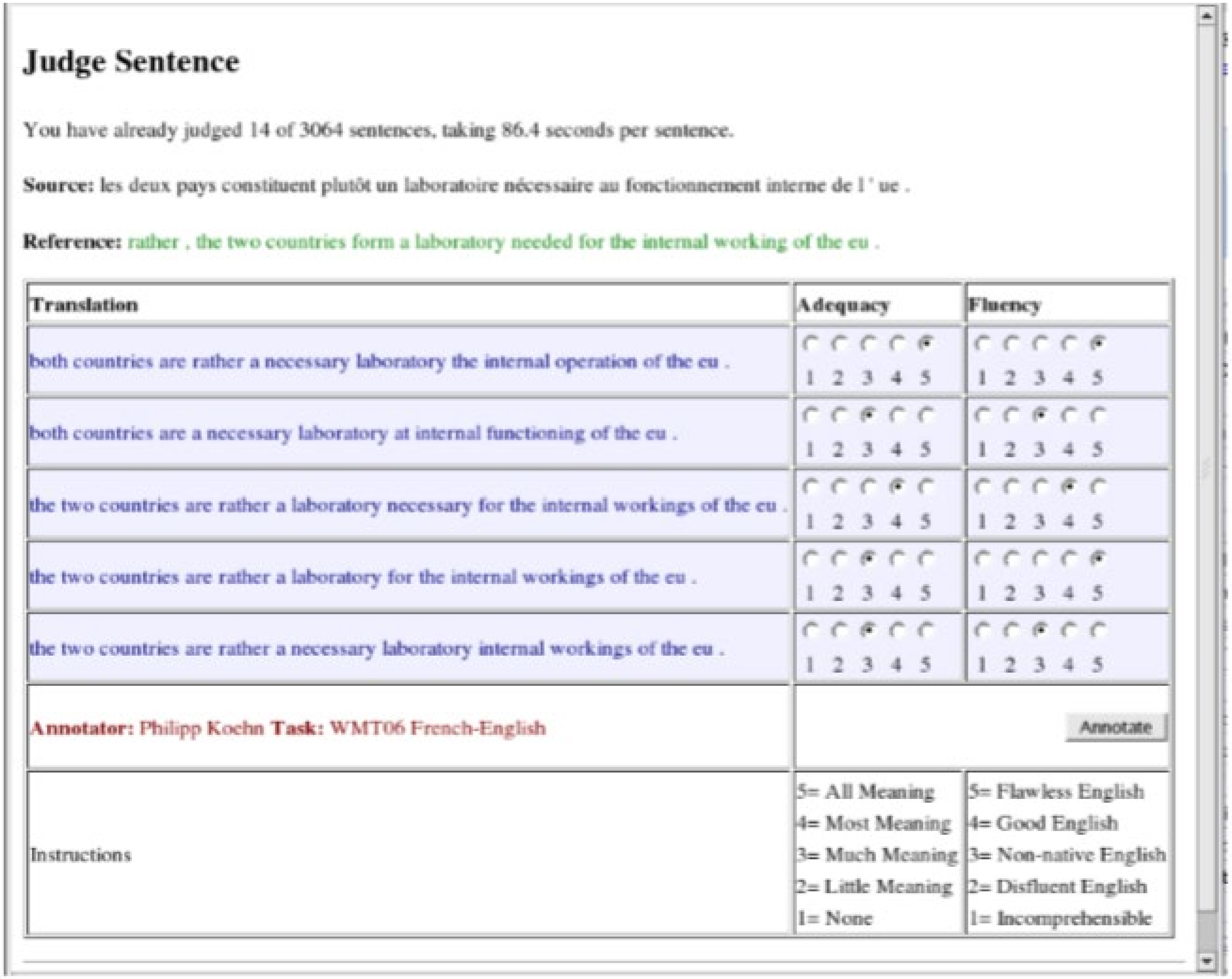


Adequacy	
5	all meaning
4	most meaning
3	much meaning
2	little meaning
1	none



Fluency	
5	flawless English
4	good English
3	non-native English
2	disfluent English
1	incomprehensible

Annotation Tools



Judge Sentence

You have already judged 14 of 3064 sentences, taking 86.4 seconds per sentence.

Source: les deux pays constituent plutôt un laboratoire nécessaire au fonctionnement interne de l ' ue .

Reference: rather , the two countries form a laboratory needed for the internal working of the eu .

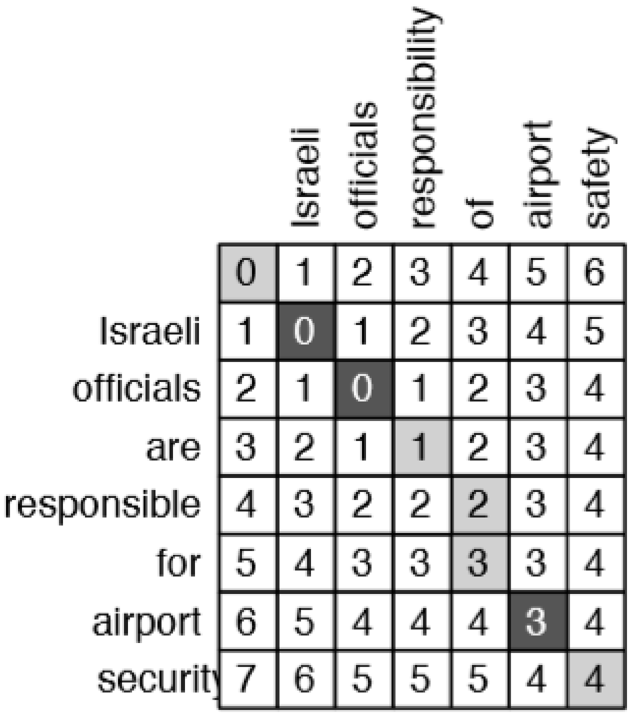
Translation	Adequacy	Fluency
both countries are rather a necessary laboratory the internal operation of the eu .	○ ○ ○ ○ ● 1 2 3 4 5	○ ○ ○ ○ ● 1 2 3 4 5
both countries are a necessary laboratory at internal functioning of the eu .	○ ○ ● ○ ○ 1 2 3 4 5	○ ○ ● ○ ○ 1 2 3 4 5
the two countries are rather a laboratory necessary for the internal workings of the eu .	○ ○ ○ ● ○ 1 2 3 4 5	○ ○ ○ ● ○ 1 2 3 4 5
the two countries are rather a laboratory for the internal workings of the eu .	○ ○ ● ○ ○ 1 2 3 4 5	○ ○ ○ ○ ● 1 2 3 4 5
the two countries are rather a necessary laboratory internal workings of the eu .	○ ○ ● ○ ○ 1 2 3 4 5	○ ○ ● ○ ○ 1 2 3 4 5
Annotator: Philipp Koehn **Task:** WMT06 French-English	Annotate	
Instructions	5= All Meaning 4= Most Meaning 3= Much Meaning 2= Little Meaning 1= None	5= Flawless English 4= Good English 3= Non-native English 2= Disfluent English 1= Incomprehensible

Automatic Evaluation Metrics

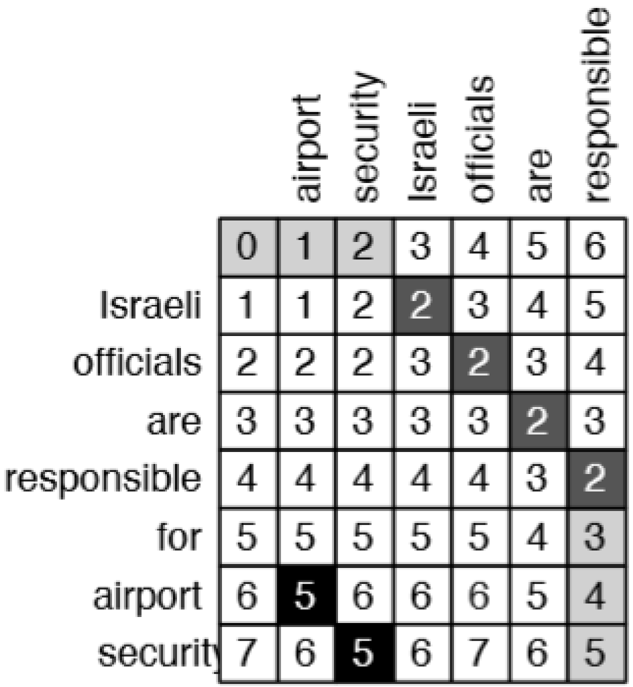
- Given: MT output, Human ref translation
- Task: compute similarity between them
- Precision/Recall of words fails to include word ordering issues
- WER (word error rate) can be useful
  - Find Min num of edits for MT->Human transform

$$\mathrm{WER} = \frac{\textit{substitutions} + \textit{insertions} + \textit{deletions}}{\textit{reference-length}}$$

WER example



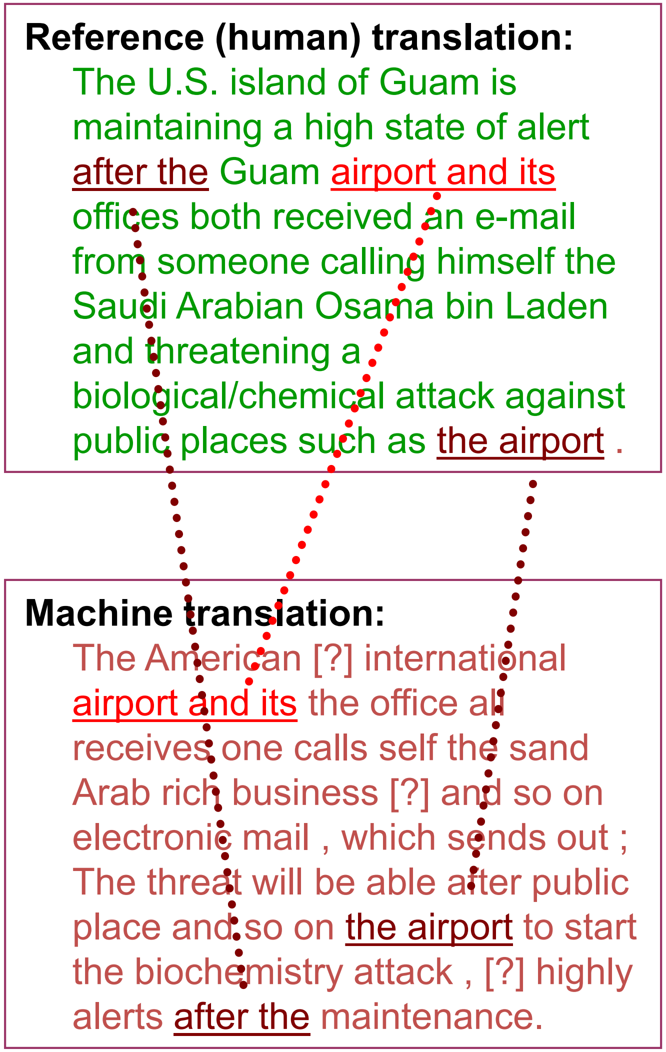
		Israeli	officials	responsibility	of	airport	safety
	0	1	2	3	4	5	6
Israeli	1	0	1	2	3	4	5
officials	2	1	0	1	2	3	4
are	3	2	1	1	2	3	4
responsible	4	3	2	2	2	3	4
for	5	4	3	3	3	3	4
airport	6	5	4	4	4	3	4
securit	7	6	5	5	5	4	4



		airport	security	Israeli	officials	are	responsible
	0	1	2	3	4	5	6
Israeli	1	1	2	2	3	4	5
officials	2	2	2	3	2	3	4
are	3	3	3	3	3	2	3
responsible	4	4	4	4	4	3	2
for	5	5	5	5	5	4	3
airport	6	5	6	6	6	5	4
securit	7	6	5	6	7	6	5

Metric	System A	System B
word error rate (WER)	57%	71%

Evaluation Metric (BLEU)



Reference (human) translation:

The U.S. island of Guam is maintaining a high state of alert after the Guam airport and its offices both received an e-mail from someone calling himself the Saudi Arabian Osama bin Laden and threatening a biological/chemical attack against public places such as the airport .

Machine translation:

The American [?] international airport and its the office all receives one calls self the sand Arab rich business [?] and so on electronic mail , which sends out ; The threat will be able after public place and so on the airport to start the biochemistry attack , [?] highly alerts after the maintenance.

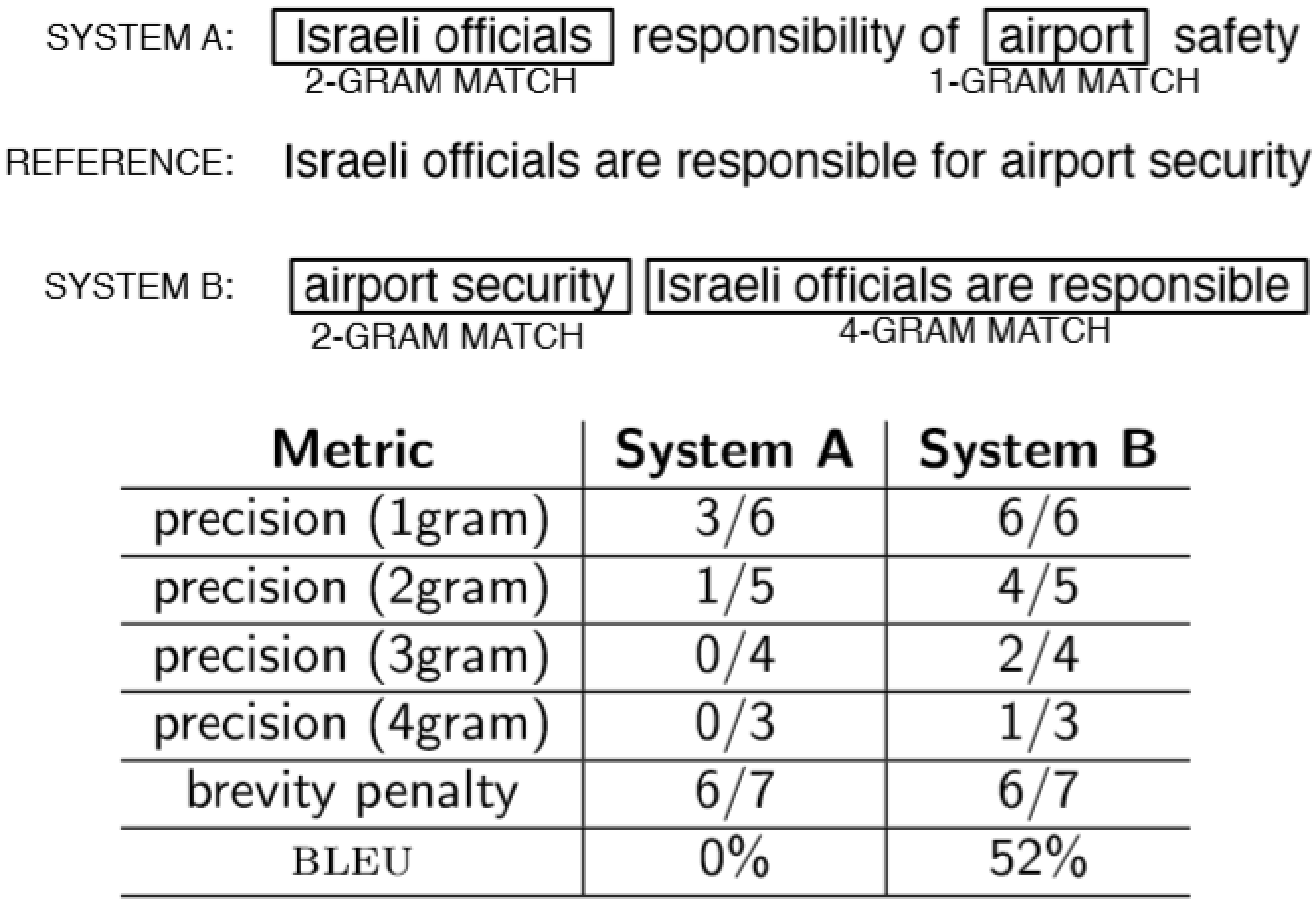
- N-gram precision (score is between 0 & 1)
  - What percentage of machine n-grams can be found in the reference translation?
    - An n-gram is an sequence of n words
  - Not allowed to use same portion of reference translation twice (can't cheat by typing out "the the the the the")
- Brevity penalty
  - Can't just type out single word "the" (precision 1.0!)
- Amazingly hard to "game" the system (i.e., find a way to change machine output so that BLEU goes up, but quality doesn't)
  - Contra doesn't hold. Can find perfectly good improvements that hurt, or don't help, BLEU

BLEU

- Use all n-grams of size 1 to 4
- Typically compute over entire corpus, not just single sentences

$$\text{BLEU} = \min\left(1, \frac{\textit{output-length}}{\textit{reference-length}}\right) \left(\prod_{i=1}^{4} \textit{precision}_i\right)^{\frac{1}{4}}$$

BLEU in Action



SYSTEM A: Israeli officials responsibility of airport safety

REFERENCE: Israeli officials are responsible for airport security

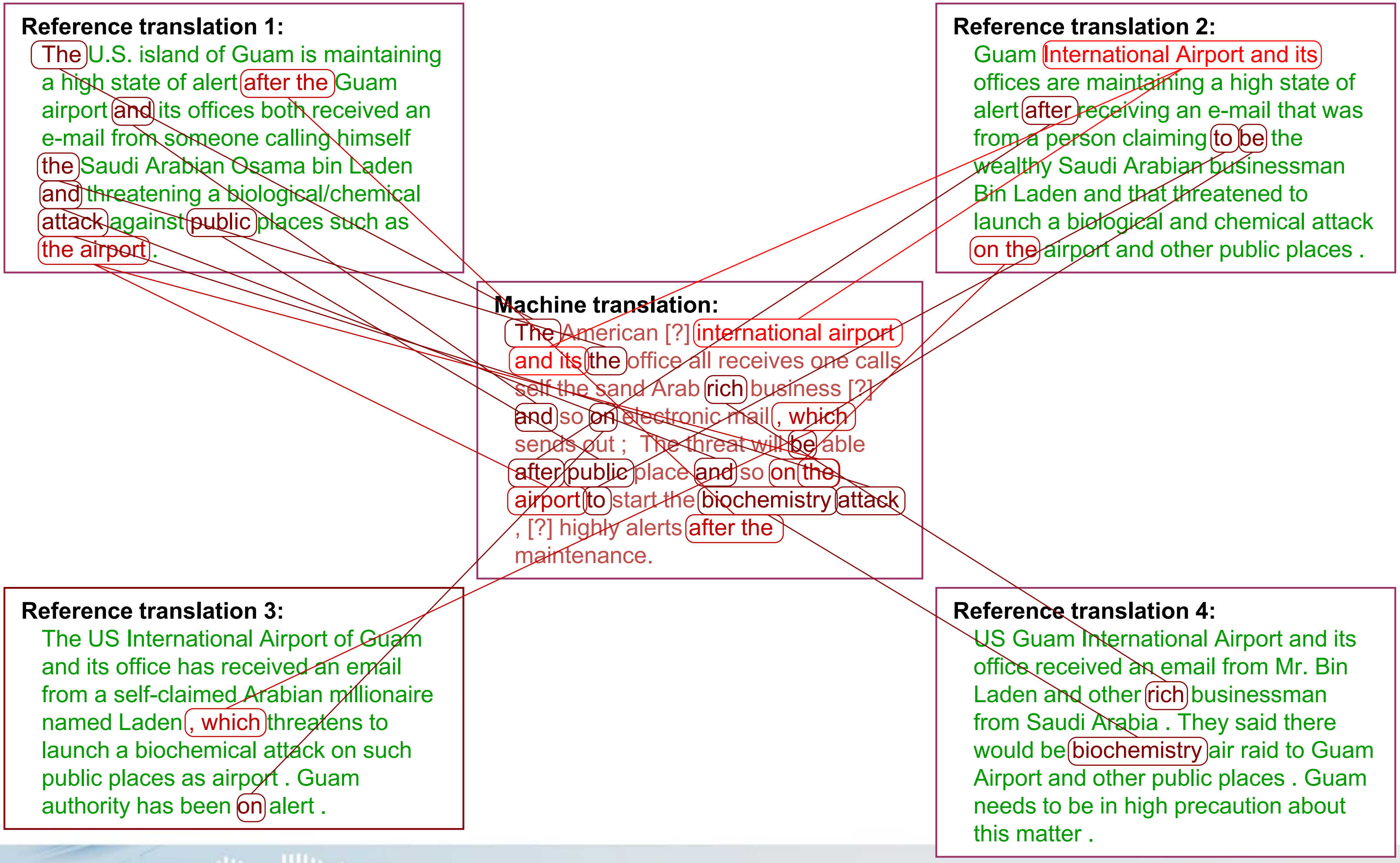
SYSTEM B: airport security Israeli officials are responsible

Metric	System A	System B
precision (1gram)	3/6	6/6
precision (2gram)	1/5	4/5
precision (3gram)	0/4	2/4
precision (4gram)	0/3	1/3
brevity penalty	6/7	6/7
BLEU	0%	52%

Multiple Reference Translations

- To account for variability, use multiple reference translations
  - N-grams may match in any of the references
  - Closest reference length is used.

Multiple Reference Translations



Reference translation 1:
The U.S. island of Guam is maintaining a high state of alert after the Guam airport and its offices both received an e-mail from someone calling himself the Saudi Arabian Osama bin Laden and threatening a biological/chemical attack against public places such as the airport .

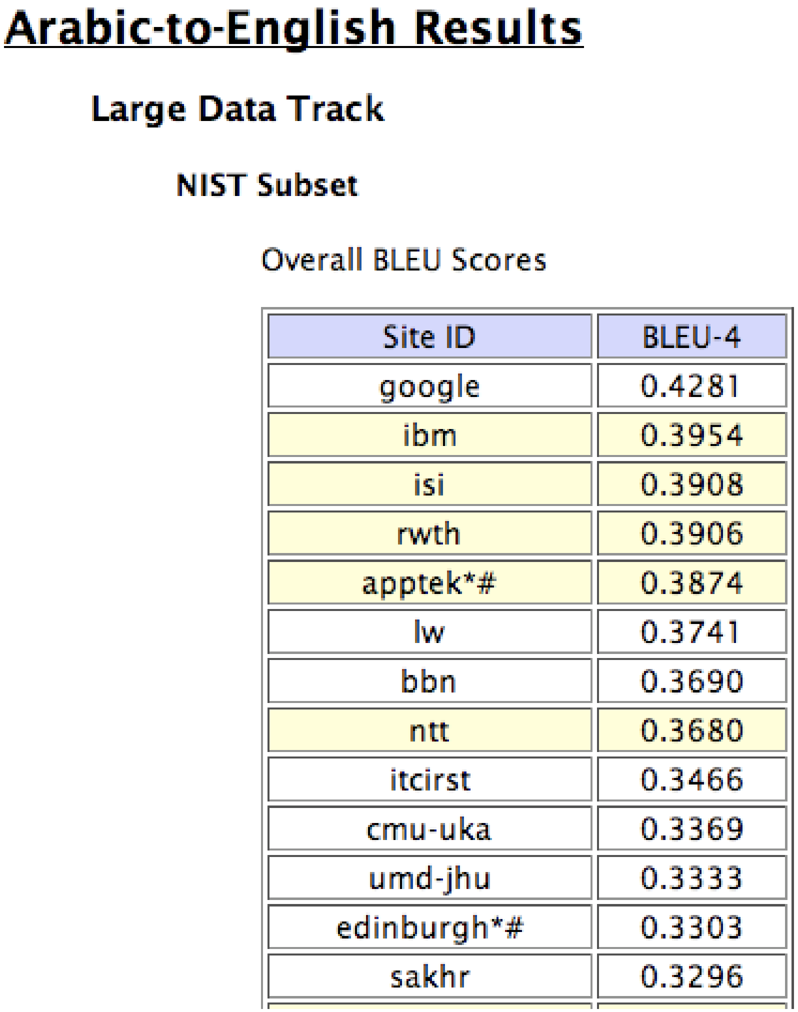
Reference translation 2:
Guam International Airport and its offices are maintaining a high state of alert after receiving an e-mail that was from a person claiming to be the wealthy Saudi Arabian businessman Bin Laden and that threatened to launch a biological and chemical attack on the airport and other public places .

Machine translation:
The American [?] international airport and its the office all receives one calls self the sand Arab rich business [?] and so on electronic mail , which sends out ; The threat will be able after public place and so on the airport to start the biochemistry attack , [?] highly alerts after the maintenance.

Reference translation 3:
The US International Airport of Guam and its office has received an email from a self-claimed Arabian millionaire named Laden , which threatens to launch a biochemical attack on such public places as airport . Guam authority has been on alert .

Reference translation 4:
US Guam International Airport and its office received an email from Mr. Bin Laden and other rich businessman from Saudi Arabia . They said there would be biochemistry air raid to Guam Airport and other public places . Guam needs to be in high precaution about this matter .

NIST 2006 Results



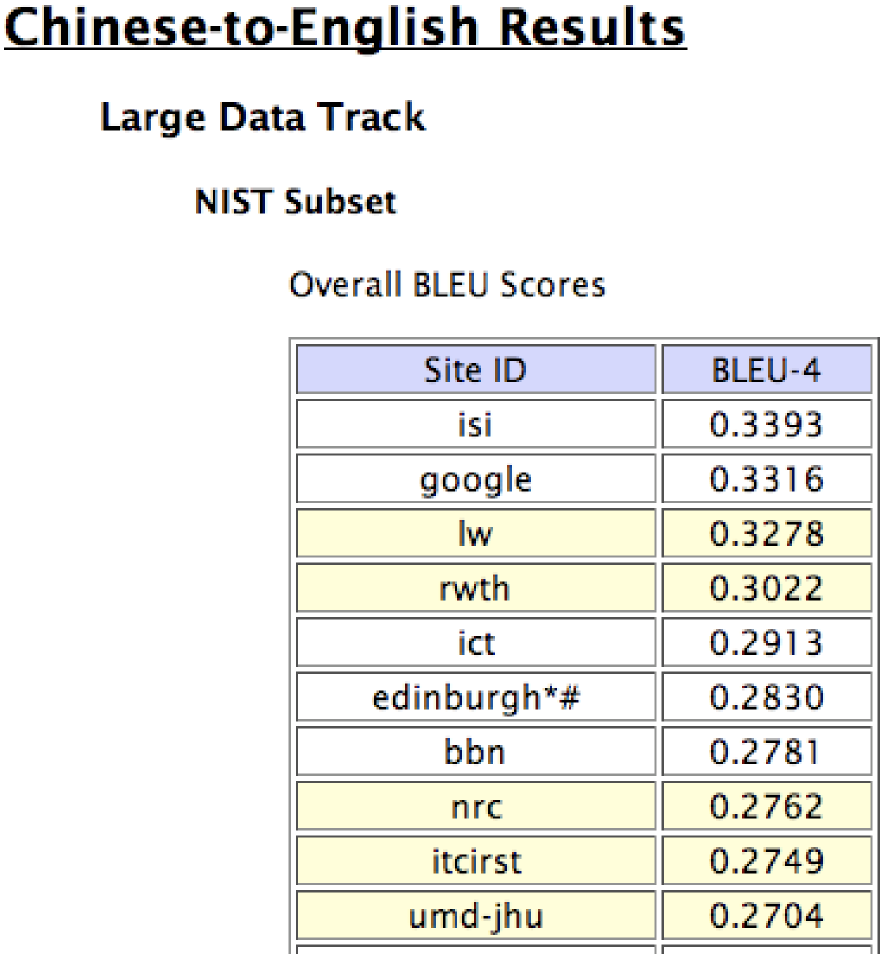
Arabic-to-English Results

Large Data Track

NIST Subset

Overall BLEU Scores

Site ID	BLEU-4
google	0.4281
ibm	0.3954
isi	0.3908
rwth	0.3906
apptek*#	0.3874
lw	0.3741
bbn	0.3690
ntt	0.3680
itcirst	0.3466
cmu-uka	0.3369
umd-jhu	0.3333
edinburgh*#	0.3303
sakhr	0.3296



Chinese-to-English Results

Large Data Track

NIST Subset

Overall BLEU Scores

Site ID	BLEU-4
isi	0.3393
google	0.3316
lw	0.3278
rwth	0.3022
ict	0.2913
edinburgh*#	0.2830
bbn	0.2781
nrc	0.2762
itcirst	0.2749
umd-jhu	0.2704

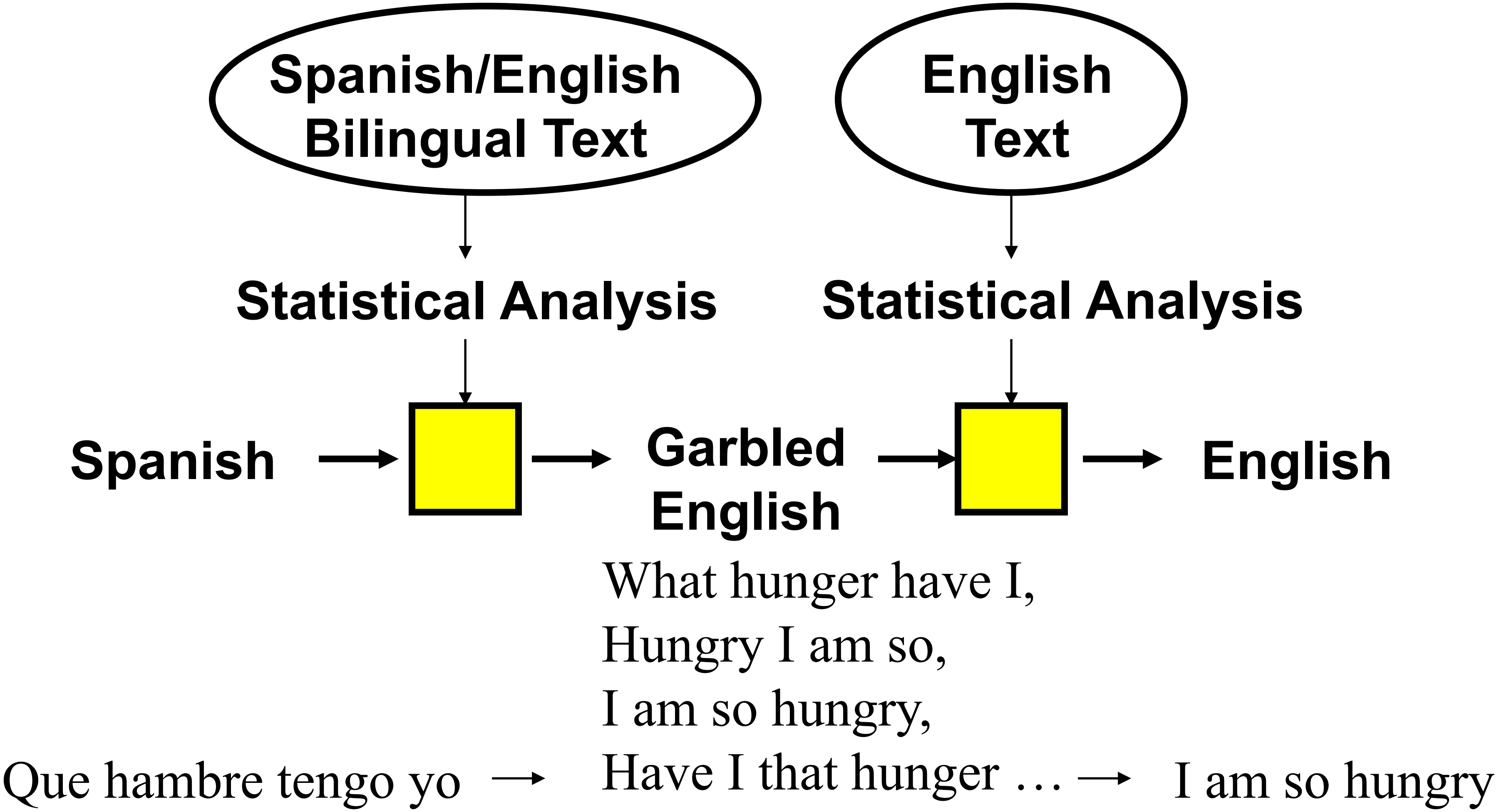
□Machine Translation

□MT Evaluation

□Statistical Machine Translation

□Neural Machine Translation

Statistical MT Systems



Spanish/English Bilingual Text

English Text

Statistical Analysis

Statistical Analysis

Spanish

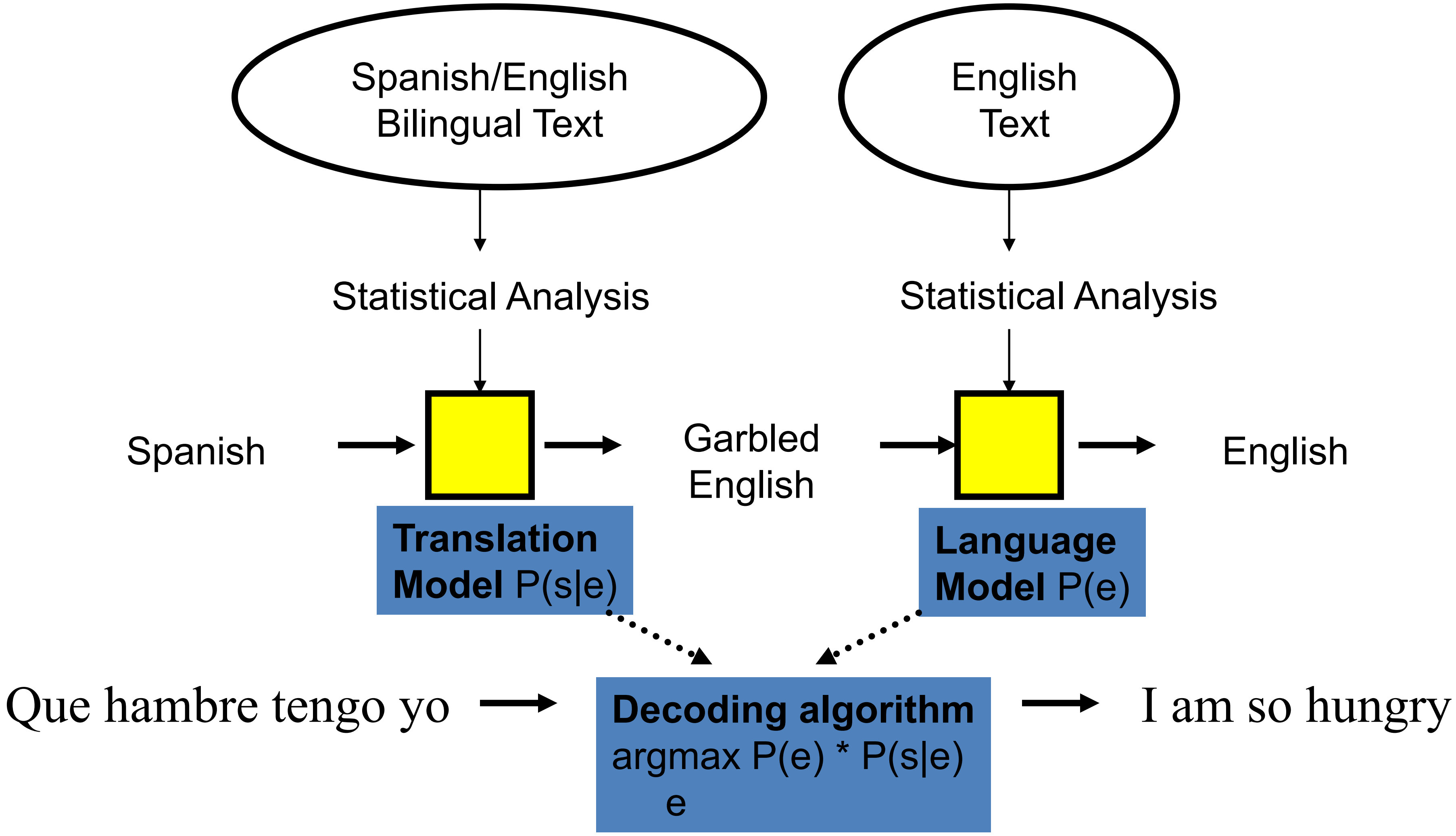
Garbled English

English

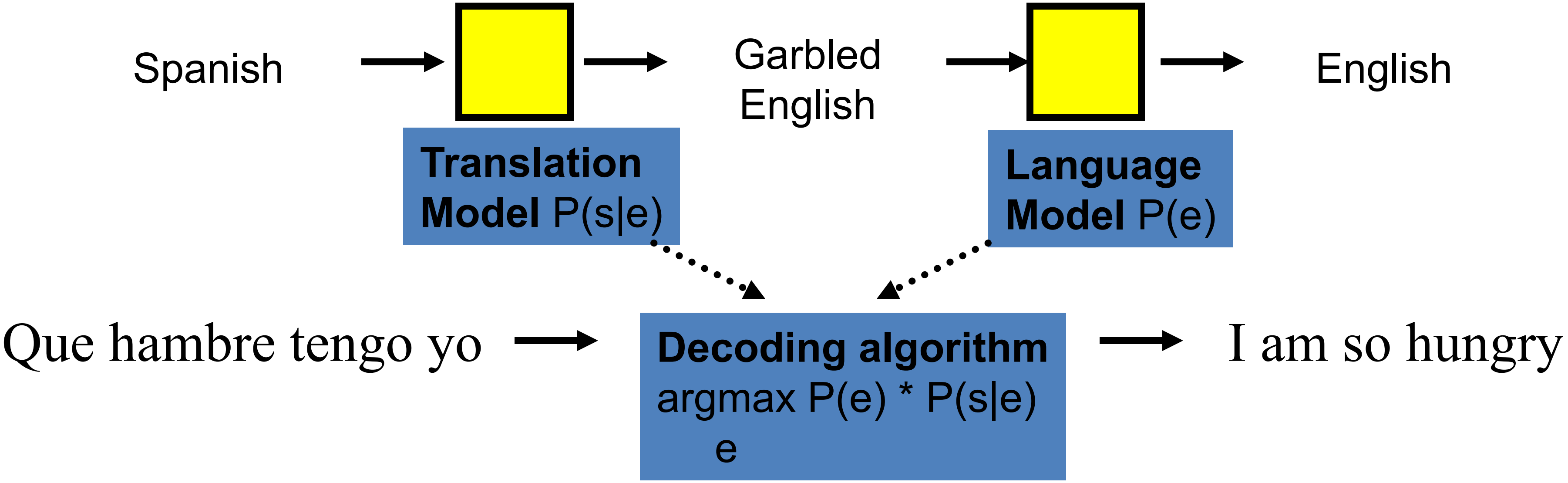
What hunger have I,
Hungry I am so,
I am so hungry,
Have I that hunger …

Que hambre tengo yo → What hunger have I, Hungry I am so, I am so hungry, Have I that hunger … → I am so hungry

Statistical MT Systems



Bayes Rule/Noisy Channel



Given a source sentence s, the decoder should consider many possible translations … and return the target string e that maximizes

P(e | s)

By Bayes Rule, we can also write this as:

P(e) x P(s | e) / P(s)

and maximize that instead. P(s) never changes while we compare different e's, so we can equivalently maximize this:

P(e) x P(s | e)

Three Sub-Problems of Statistical MT

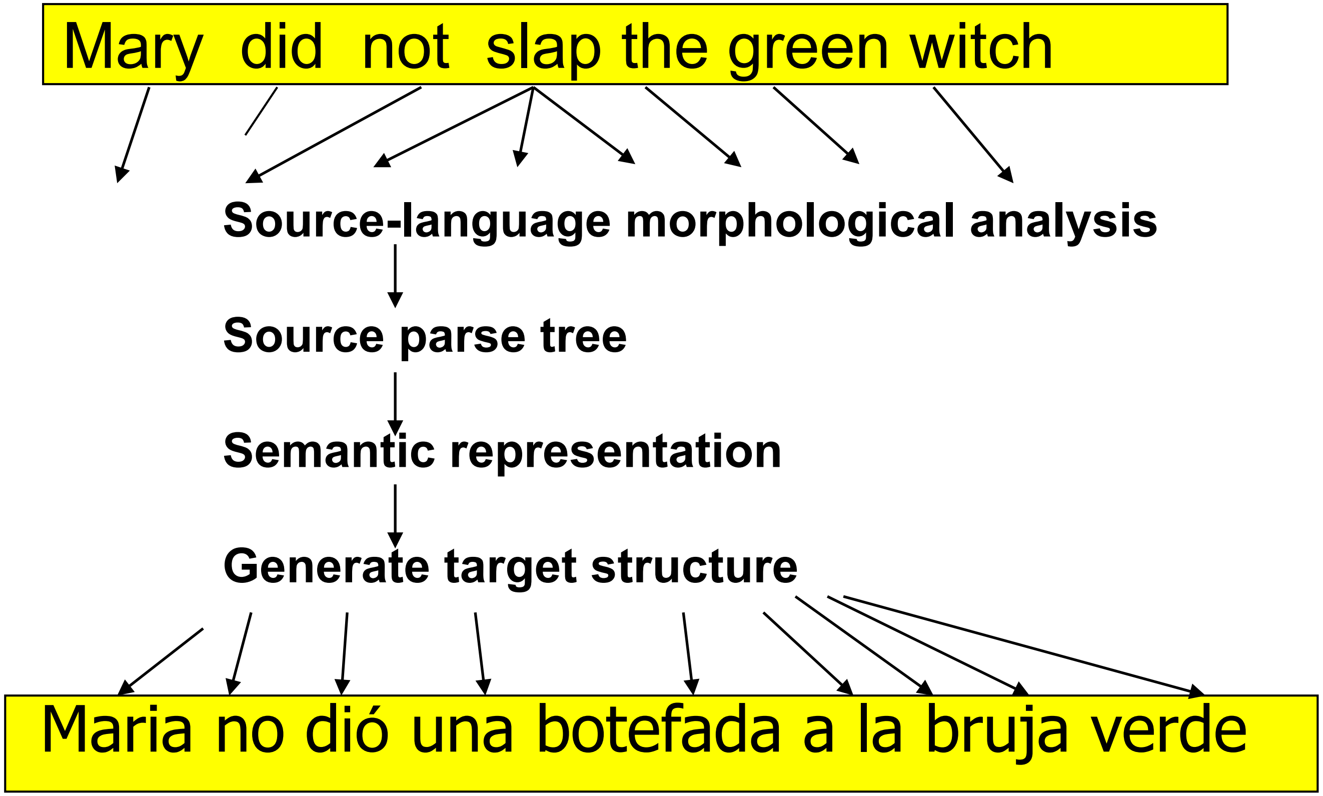
- Language model
  - Given an English string e, assigns P(e) by formula
  - good English string -> high P(e)
  - random word sequence -> low P(e)

- Translation model
  - Given a pair of strings <f,e>, assigns P(f | e) by formula
  - <f,e> look like translations -> high P(f | e)
  - <f,e> don't look like translations -> low P(f | e)

- Decoding algorithm
  - Given a language model, a translation model, and a new sentence f … find translation e maximizing P(e) * P(f | e)

Translation Model

Generative story:



Mary did not slap the green witch

Source-language morphological analysis

Source parse tree

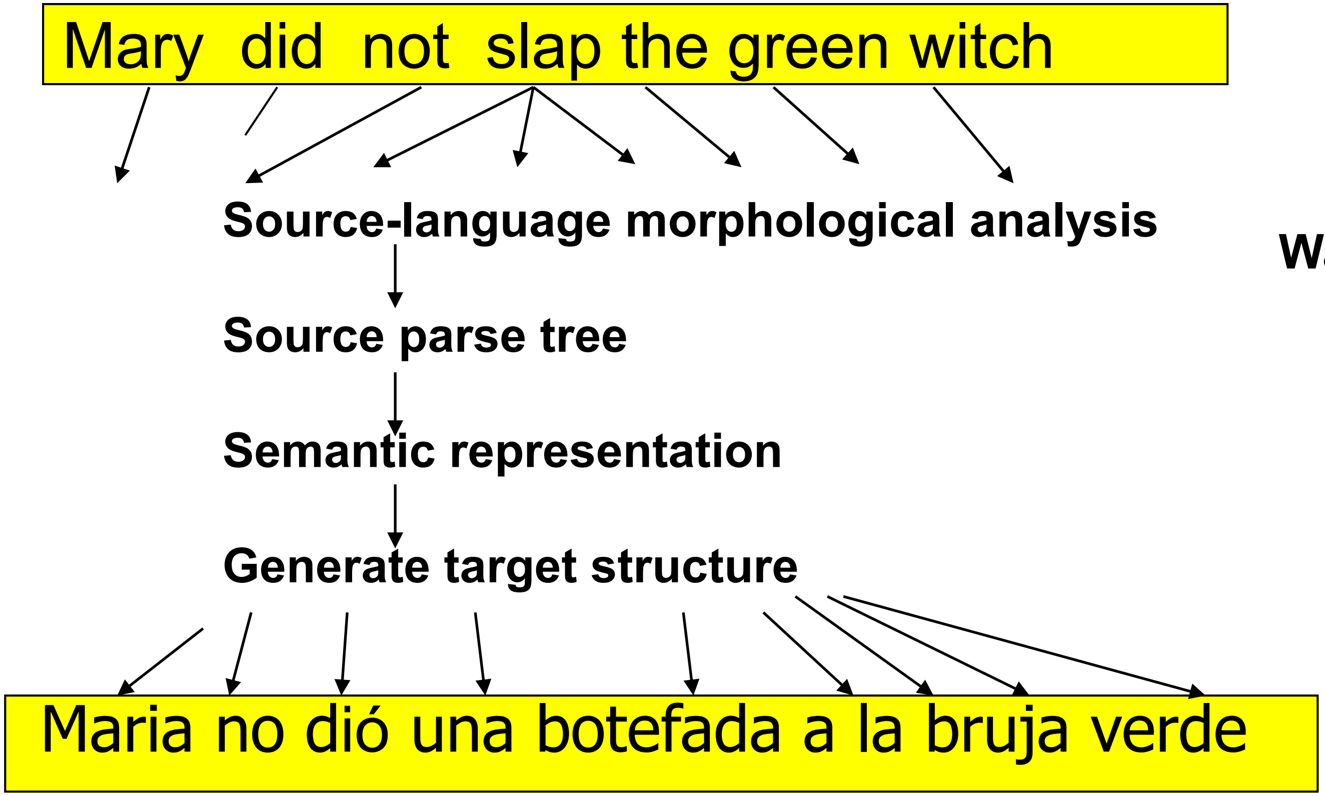
Semantic representation

Generate target structure

Maria no dió una botefada a la bruja verde

Translation Model?

Generative story:



Mary did not slap the green witch

Source-language morphological analysis

Source parse tree

Semantic representation

Generate target structure

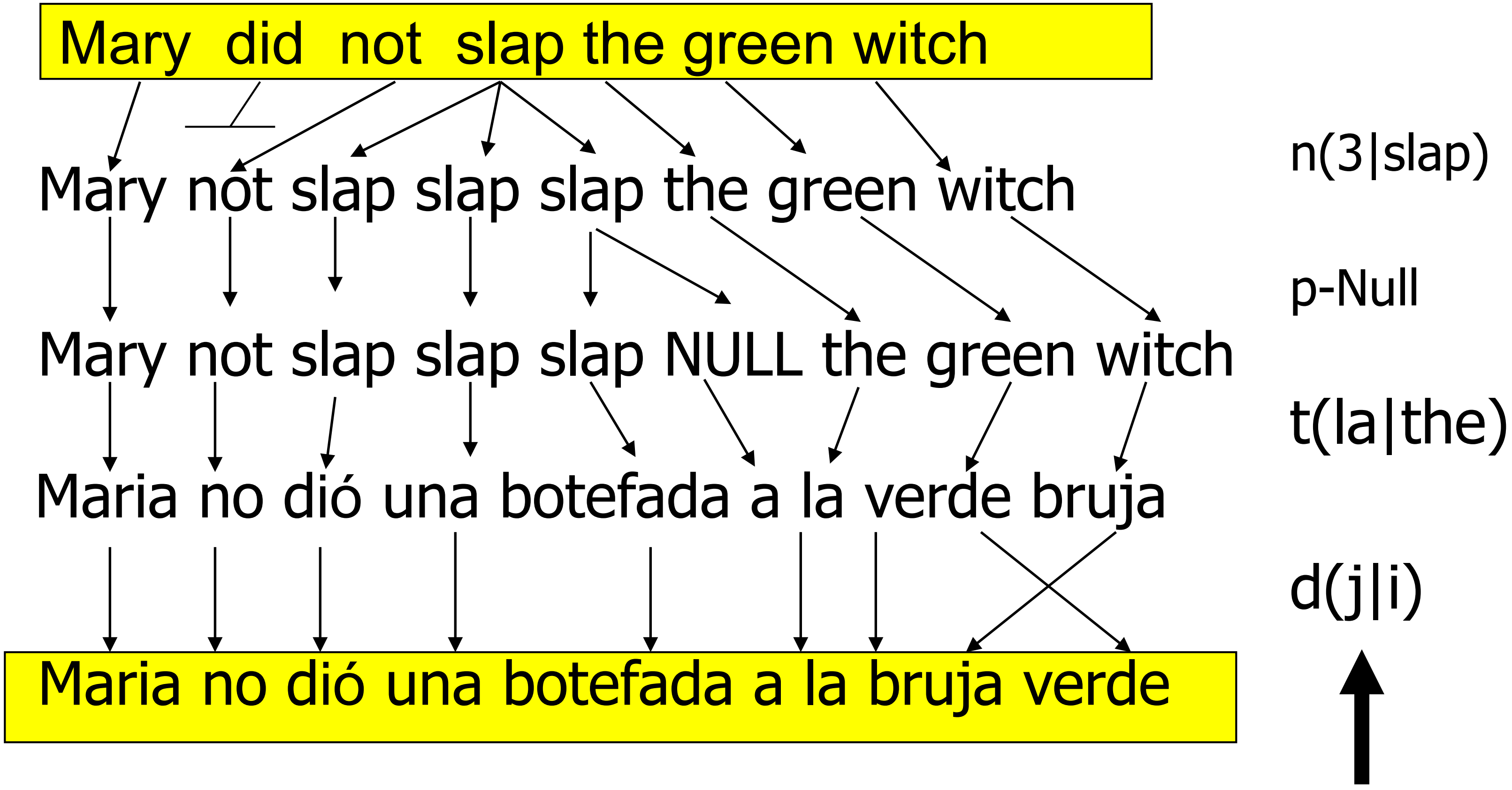
Maria no dió una botefada a la bruja verde

Way too hard.

The Classic Translation Model

Word Substitution/Permutation [IBM Model 3, Brown et al., 1993]

Generative story:



Mary did not slap the green witch

Mary not slap slap slap the green witch

Mary not slap slap slap NULL the green witch

Maria no dió una botefada a la verde bruja

Maria no dió una botefada a la bruja verde

n(3|slap)

p-Null

t(la|the)

d(j|i)

Parts List

- We need probabilities for
  - n (x|y) The probability that word y will yield x outputs in the translation… (fertility)
  - p The probability of a null insertion
  - t The actual word translation probability table
  - d(j|i) the probability that a word at position i will make an appearance at position j in the translation

Parts List

- Every one of these can be learned from a sentence aligned corpus…
  - i.e. A corpus where sentences are paired but nothing else is specified
- And the EM algorithm

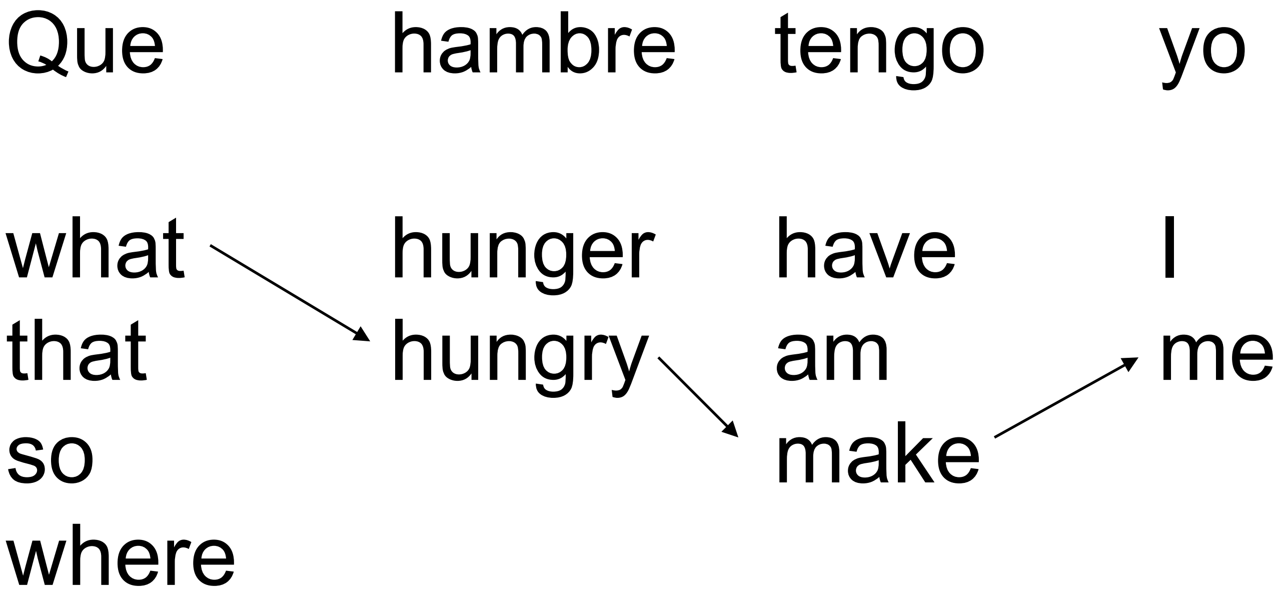
EM/Alignment

- We need some parameters
  - which we don't have
- We can get them from a word-aligned corpus
  - which we don't have
- So we make up some parameters to get the alignment and then use that alignment to get the right numbers.

Decoding

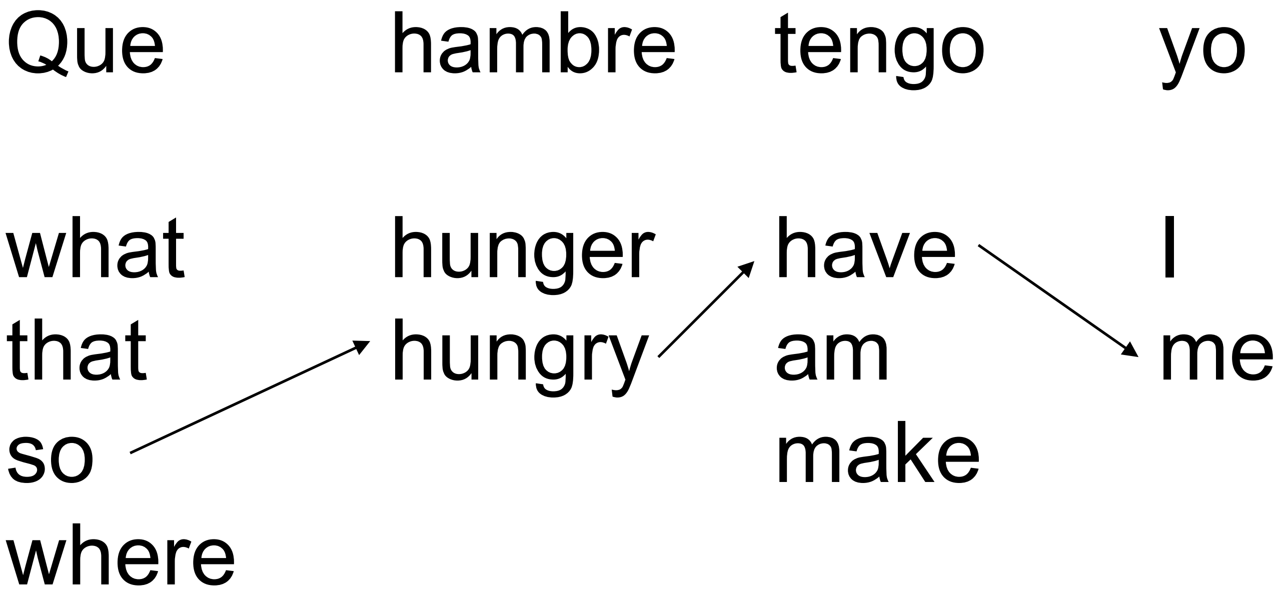
- A Viterbi algorithm
  - Given foreign sentence f, find English sentence e that maximizes P(e) x P(f | e)
  - Space is defined by the model (fertility, distortion, word translation model, etc.)
  - Large space --> efficient decoding is required.

Decoding



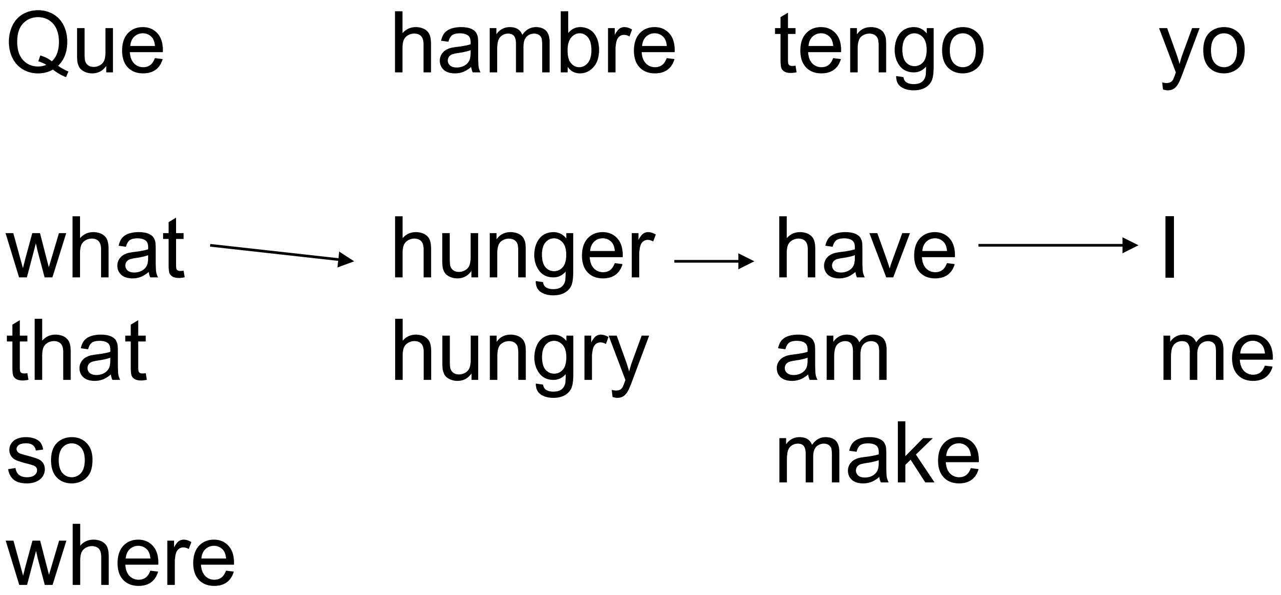
Que	hambre	tengo	yo
what	hunger	have	I
that	hungry	am	me
so		make	
where			

Decoding



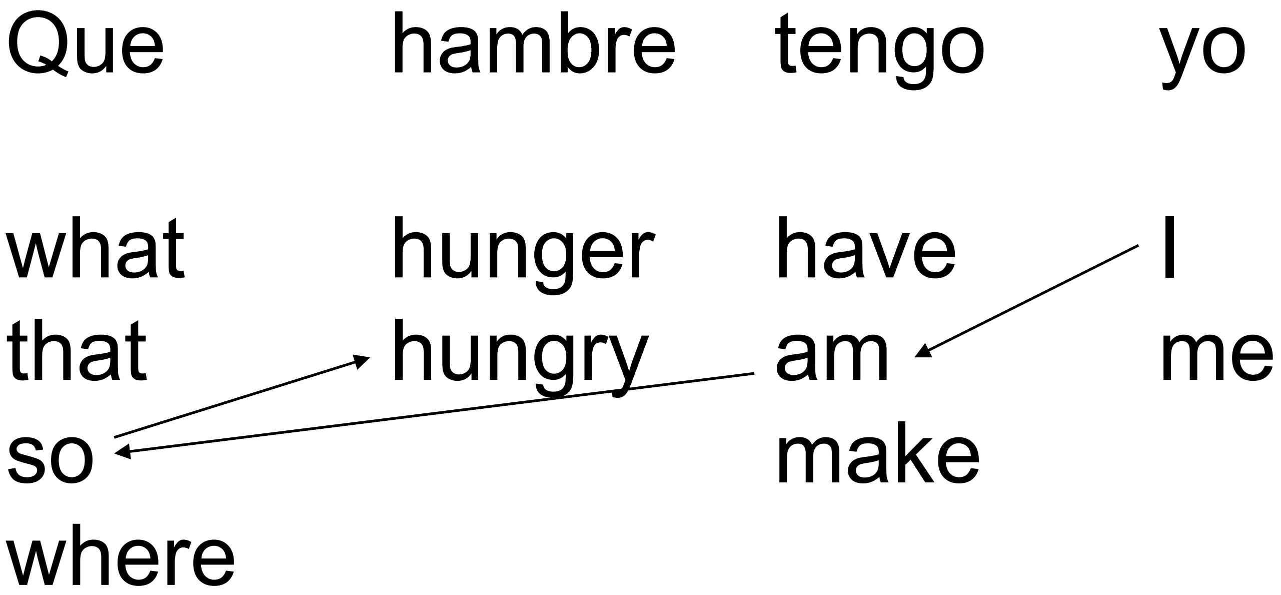
Que	hambre	tengo	yo
what	hunger	have	I
that	hungry	am	me
so		make	
where			

Decoding



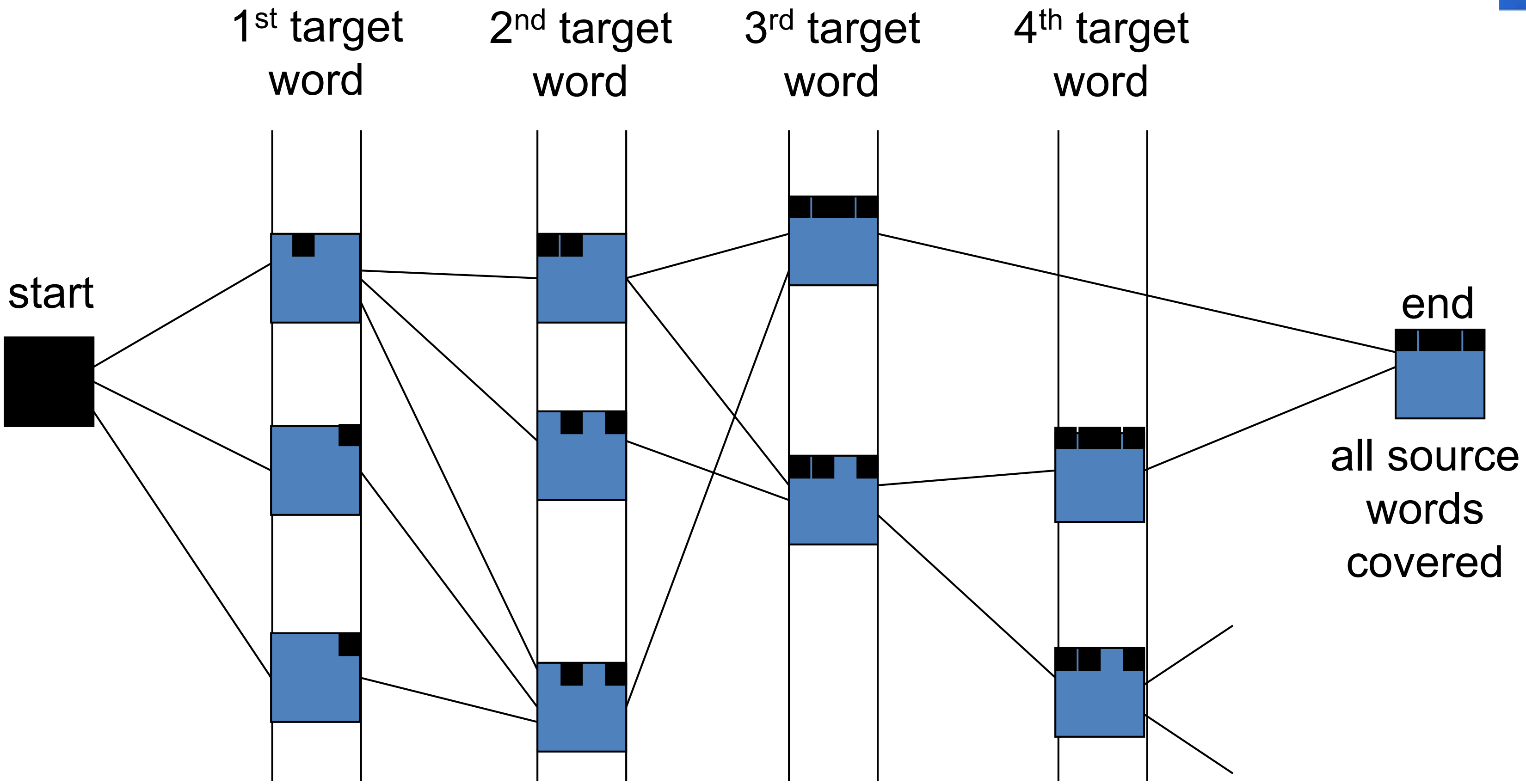
Que	hambre	tengo	yo
what	hunger	have	I
that	hungry	am	me
so		make	
where			

Decoding



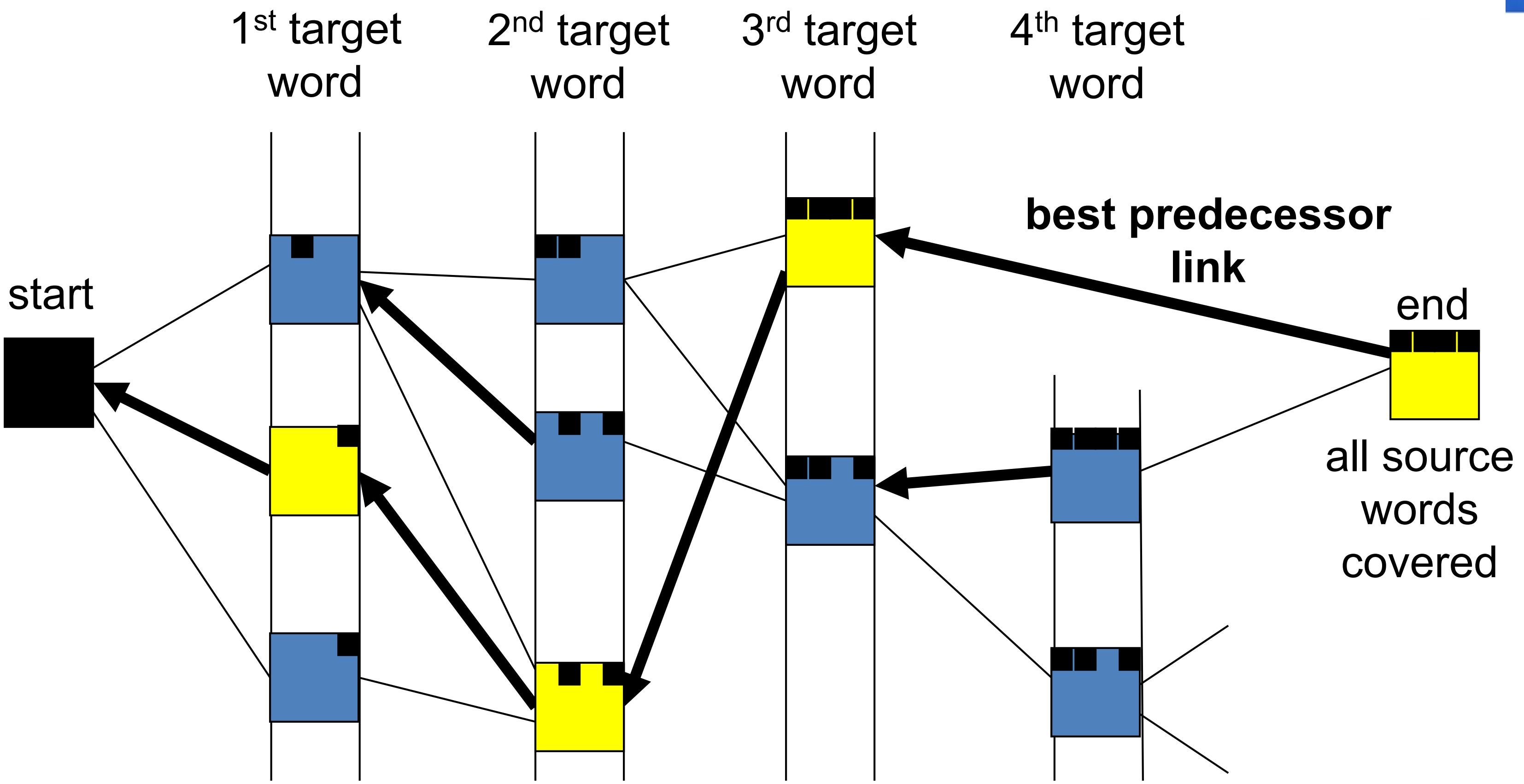
Que	hambre	tengo	yo
what	hunger	have	I
that	hungry	am	me
so		make	
where			

Decoder: Actually Translates New Sentences



Each partial translation hypothesis contains:
- Last English word chosen + source words covered by it
- Next-to-last English word chosen
- Entire coverage vector (so far) of source sentence
- Language model and translation model scores (so far)

Dynamic Programming Beam Search



1st target word

2nd target word

3rd target word

4th target word

start

best predecessor link

end

all source words covered

Each partial translation hypothesis contains:

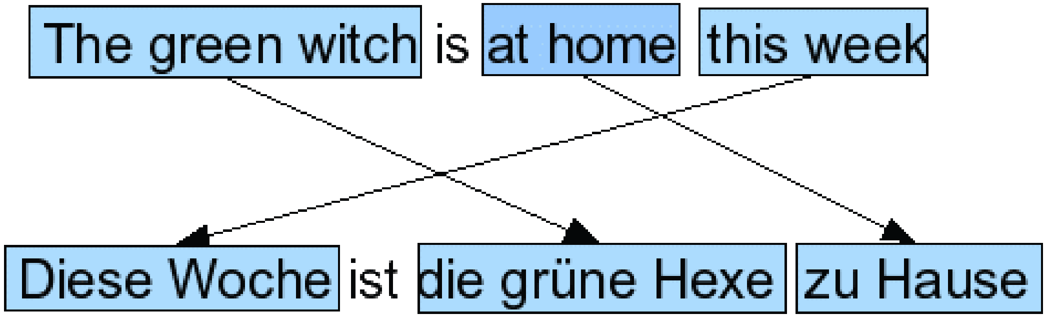
- Last English word chosen + source words covered by it
- Next-to-last English word chosen
- Entire coverage vector (so far) of source sentence ■■ ■
- Language model and translation model scores (so far)

[Jelinek, 1969;
Brown et al, 1996 US Patent;
(Och, Ueffing, and Ney, 2001]

Flaws of Word-Based MT

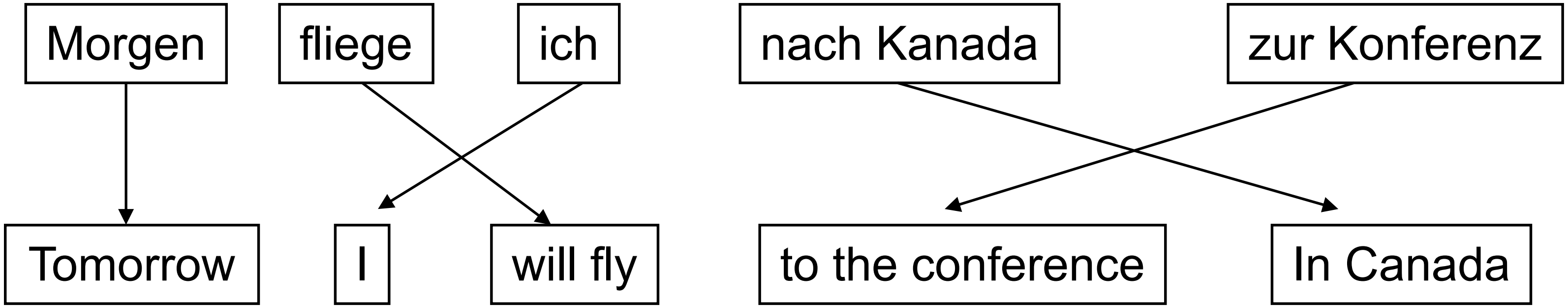
- Multiple English words for one foreign word
  - IBM models can do one-to-many (fertility) but not many-to-one
- Phrasal Translation
  - “real estate”， “note that”， “interest in”
- Syntactic Transformations
  - Verb at the beginning in Arabic
  - Translation model penalizes any proposed re-ordering
  - Language model not strong enough to force the verb to move to the right place

Intuition of phrase-based translation (Koehn et al. 2003)



- Generative story has three steps
  1) Group words into phrases
  2) Translate each phrase
  3) Move the phrases around

Phrase-Based Statistical MT



- Source (foreign) input segmented in to phrases
  - “phrase” is any sequence of words
- Each phrase is probabilistically translated into English
  - P(to the conference | zur Konferenz)
  - P(into the meeting | zur Konferenz)   HUGE TABLE!!
- Phrases are then probabilistically re-ordered

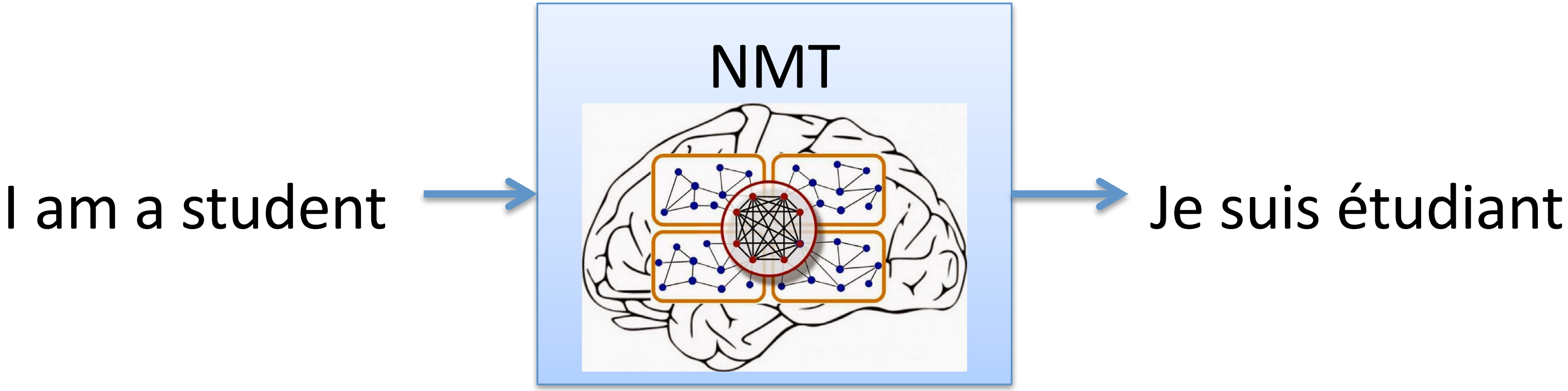
Advantages of Phrase-Based SMT

- Many-to-many mappings can handle non-compositional phrases (e.g., “real estate”)
- Local context is very useful for disambiguating
  - “Interest rate” → …
  - “Interest in” → …
- The more data, the longer the learned phrases
  - Sometimes whole sentences
    - Interesting parallel to concatenative synthesis for TTS

- □Machine Translation
- □MT Evaluation
- □Statistical Machine Translation
- □Neural Machine Translation

(Adapted from Thang Luong's Slides for Stanford CS224d)

Neural Machine Translation



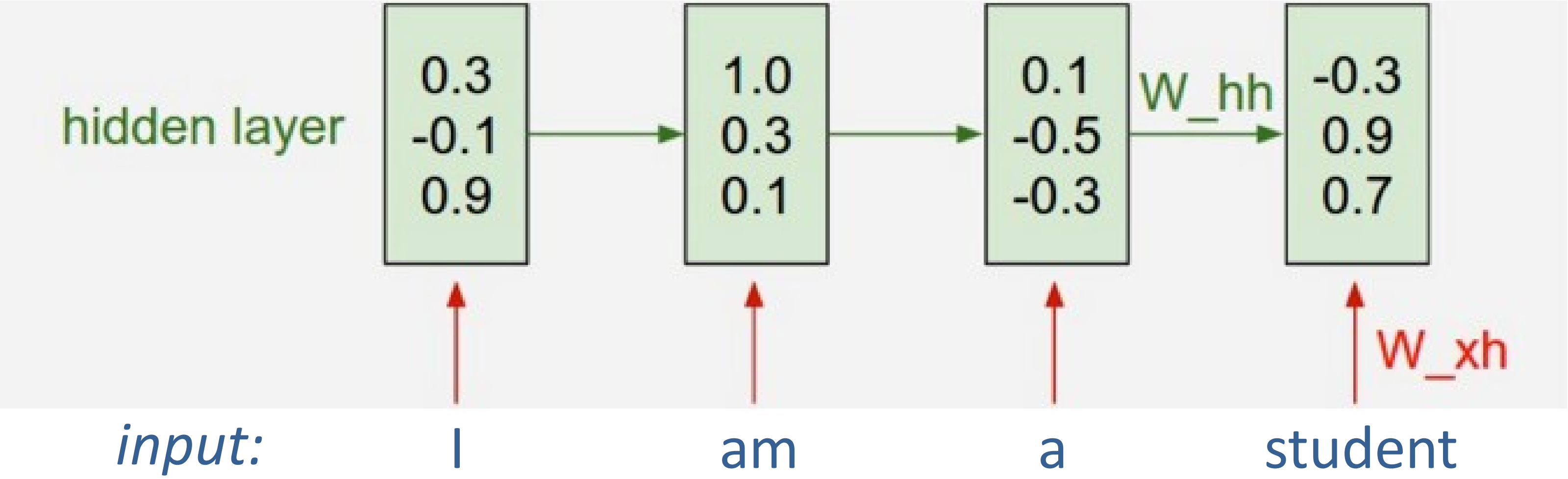
I am a student → NMT → Je suis étudiant

(*Sutskever et al., 2014; Cho et al., 2014)*

- *Sequence-to-sequence*: translate globally.
- *End-to-end*: simple & generalizable.

Let’s find out!

Recurrent Neural Networks (RNNs)



hidden layer

0.3 -0.1 0.9

1.0 0.3 0.1

0.1 -0.5 -0.3

W_hh

-0.3 0.9 0.7

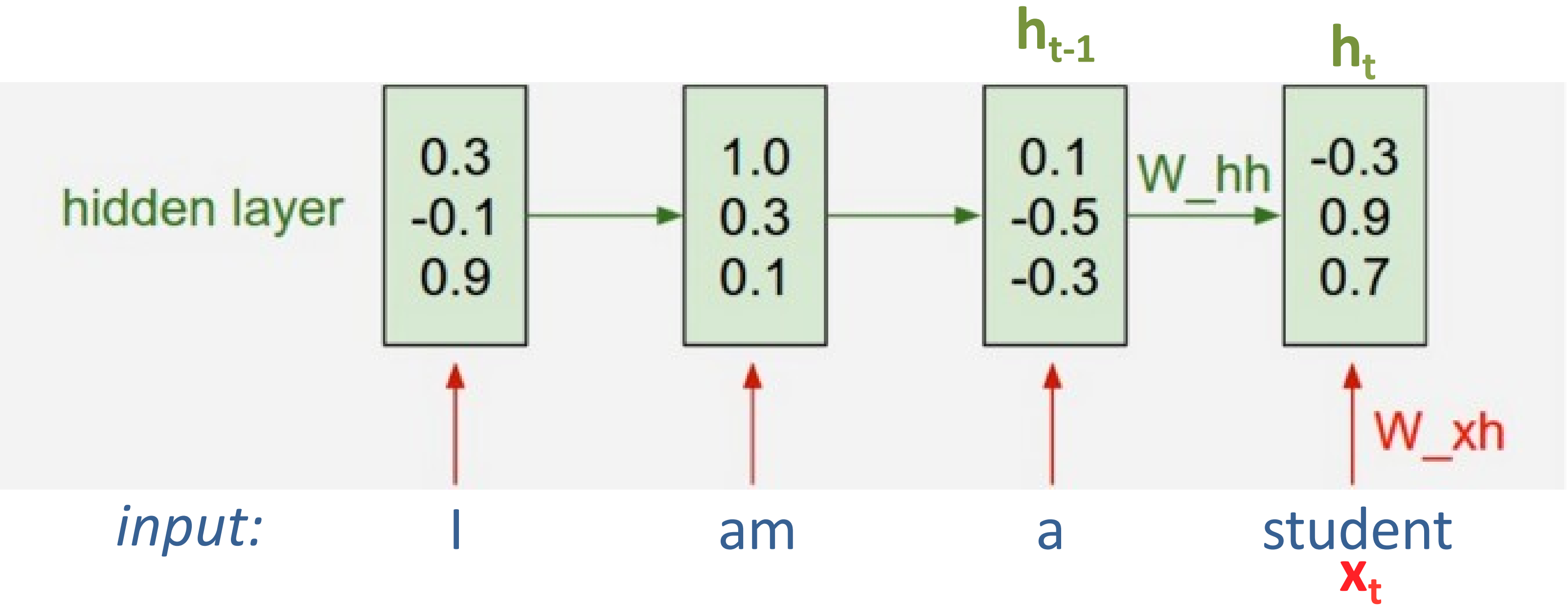
W_xh

input: I am a student

(Picture adapted from Andrej Karparthy)

Recurrent Neural Networks (RNNs)

$$h_t = \sigma\left(W_{xh} x_t + W_{hh} h_{t-1}\right)$$



RNNs to represent sequences!

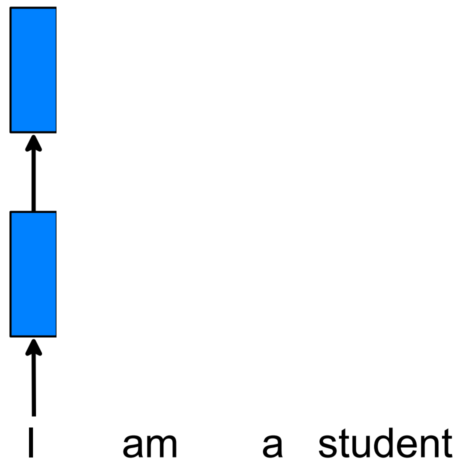
(Picture adapted from Andrej Karparthy)

Neural Machine Translation (NMT)

I am a student Je suis étudiant

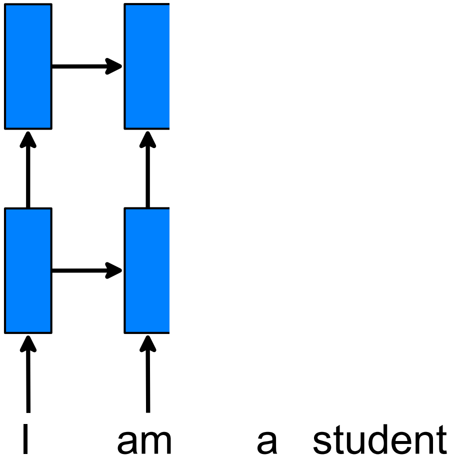
- Recurrent Neural Networks:
  - – Model P(target | source) directly.
  - – Can be trained end-to-end.

Neural Machine Translation (NMT)



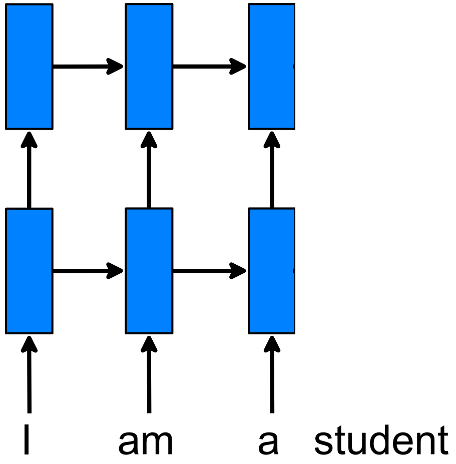
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Neural Machine Translation (NMT)



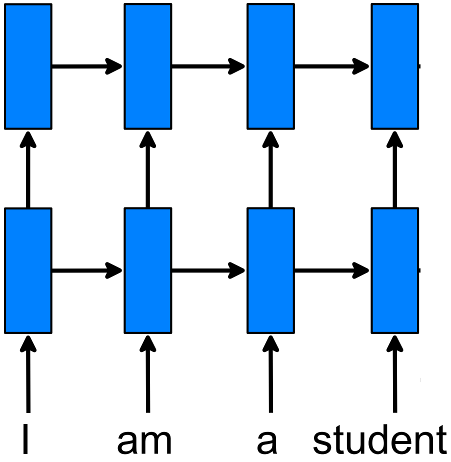
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Neural Machine Translation (NMT)



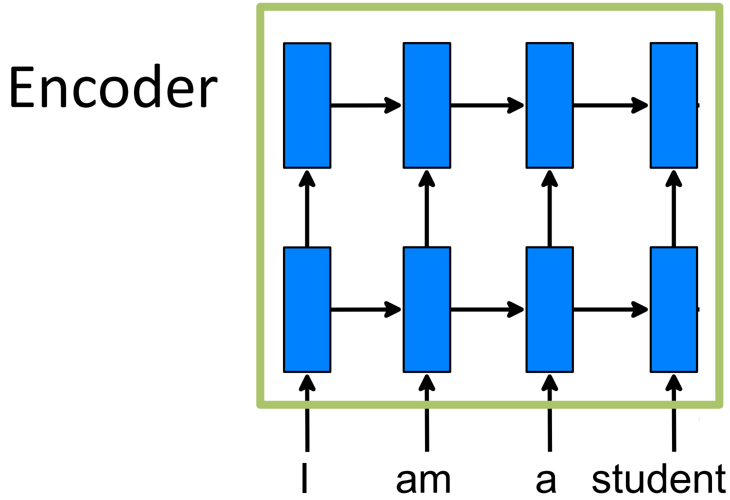
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Neural Machine Translation (NMT)



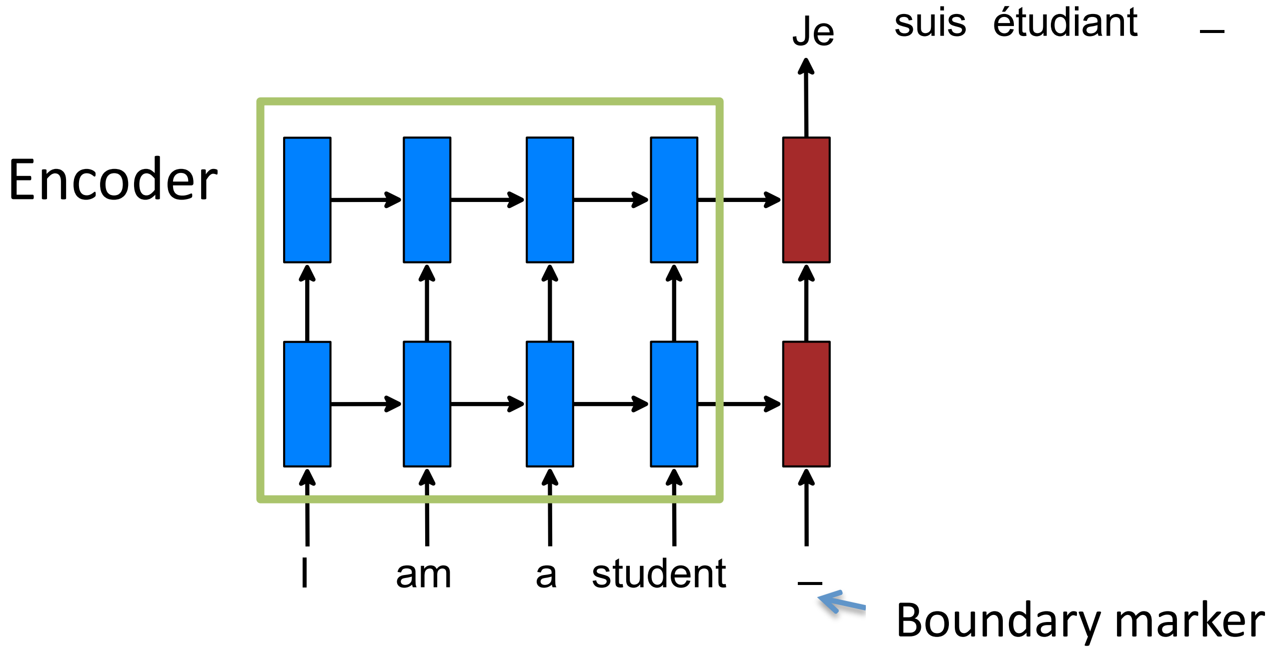
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Neural Machine Translation (NMT)



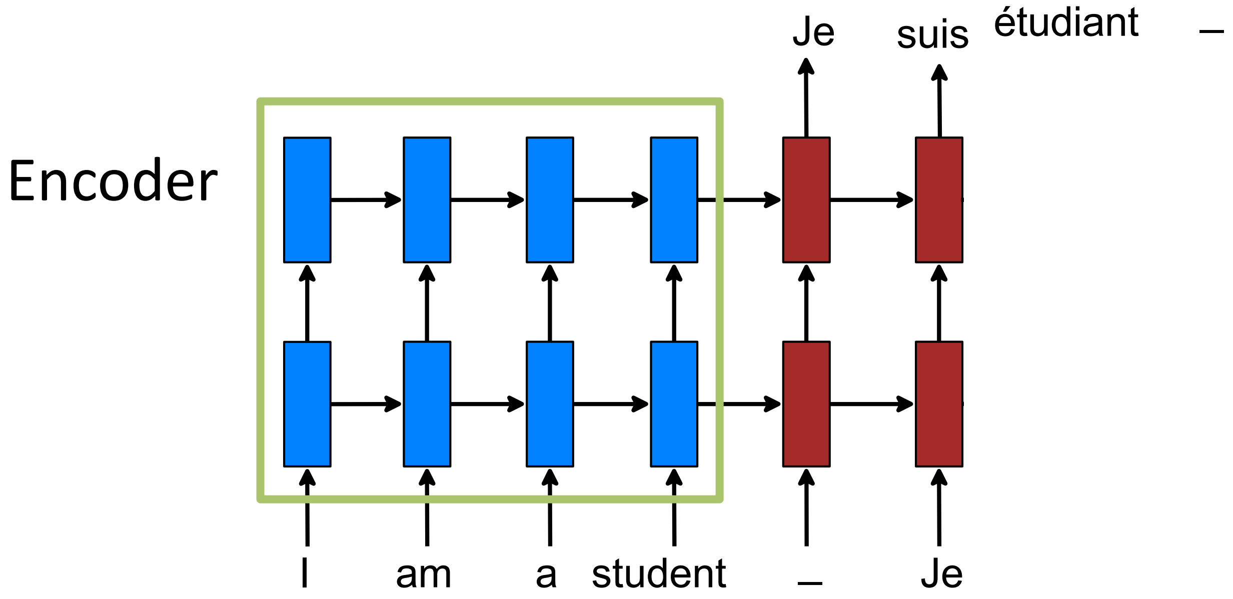
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Neural Machine Translation (NMT)



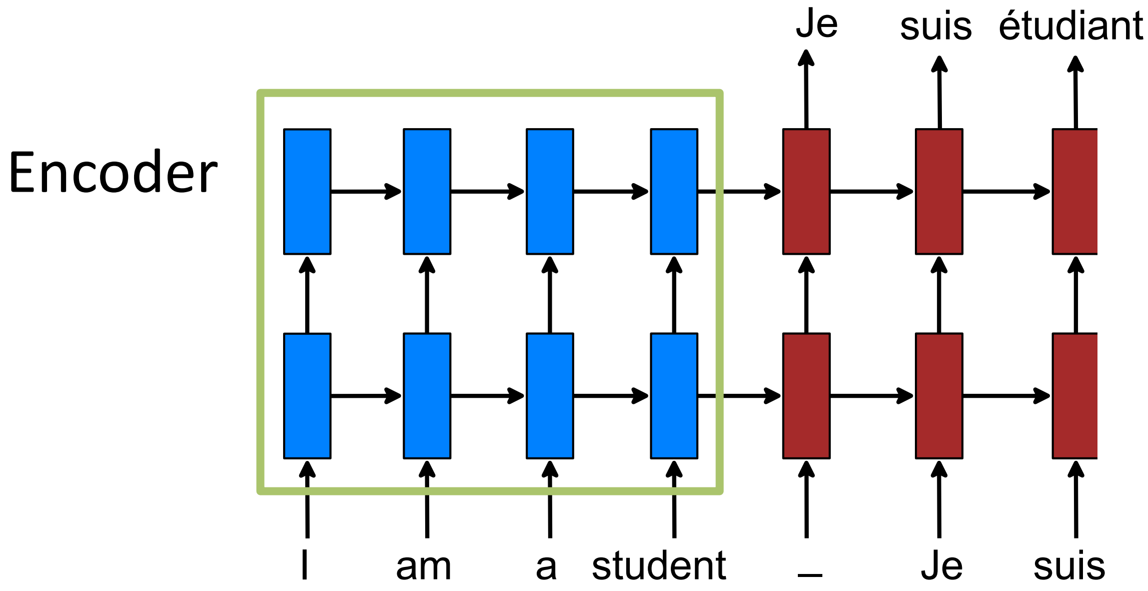
- Recurrent Neural Networks:
  - – Model P(target | source) directly.
  - – Can be trained end-to-end.

Neural Machine Translation (NMT)



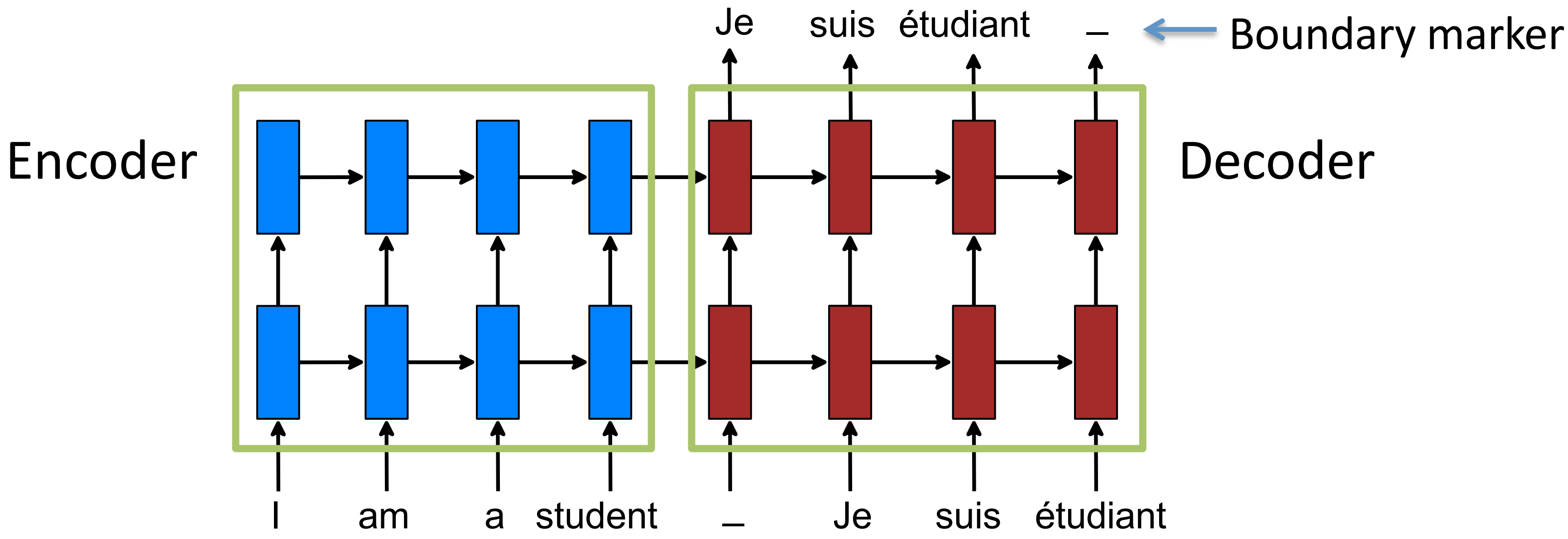
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Neural Machine Translation (NMT)



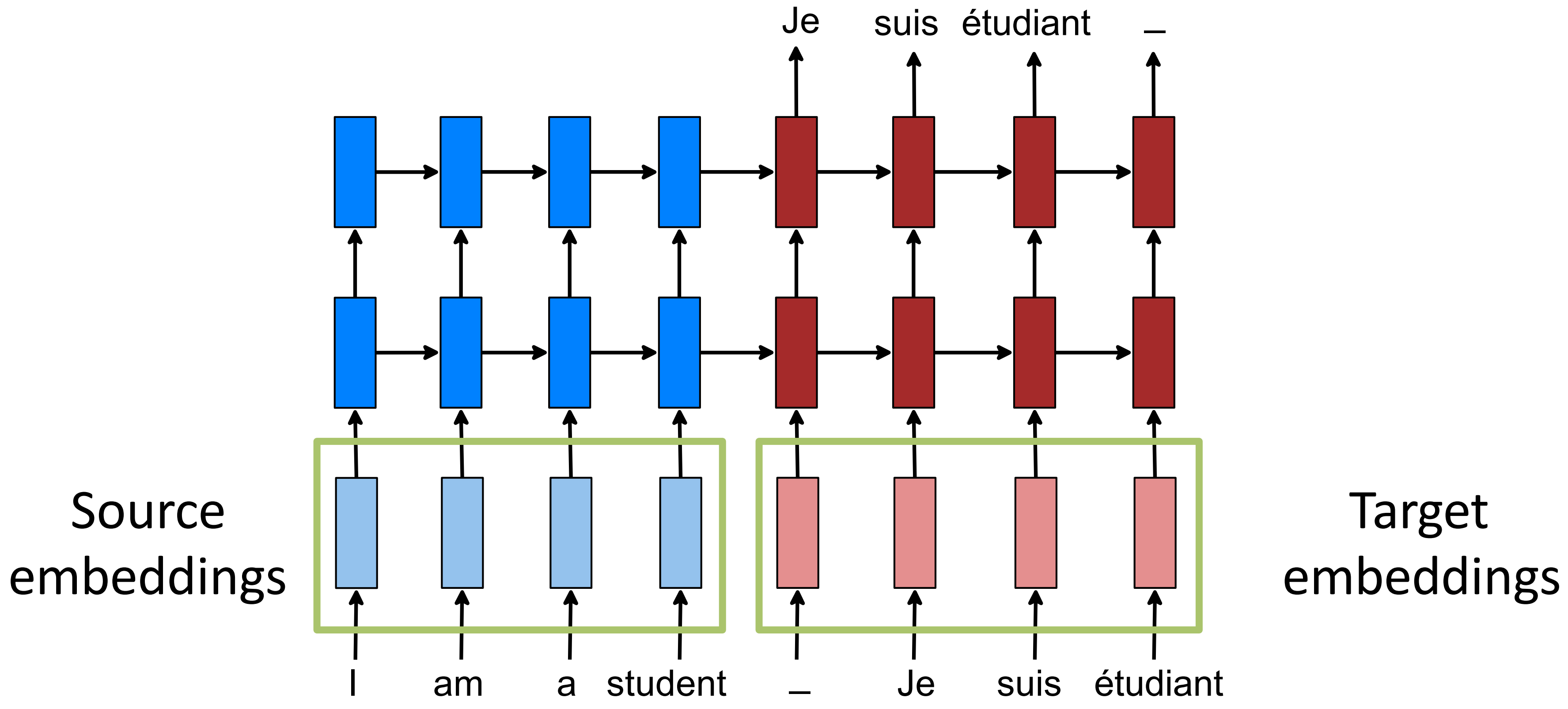
- Recurrent Neural Networks:
  - – Model P(target | source) directly.
  - – Can be trained end-to-end.

Neural Machine Translation (NMT)



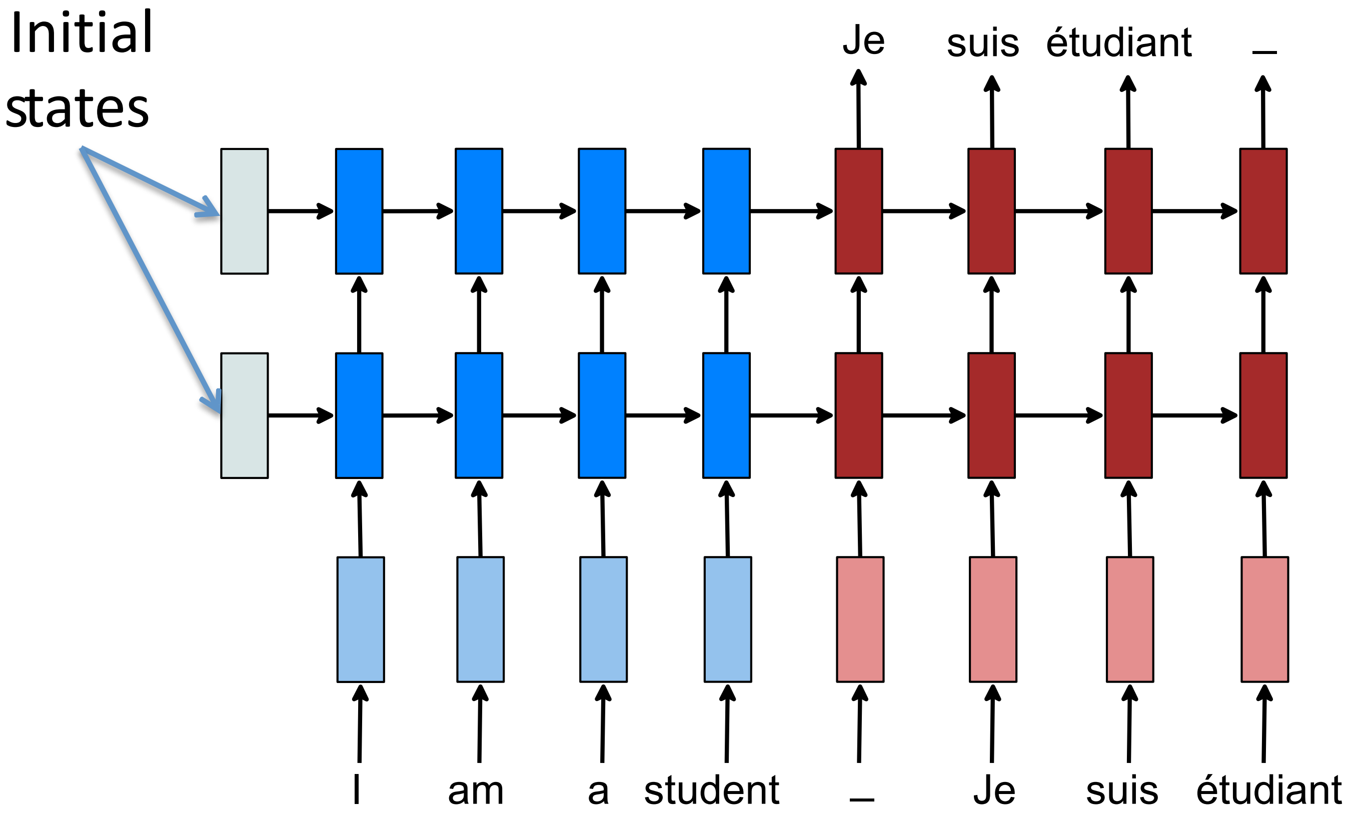
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Word Embedding



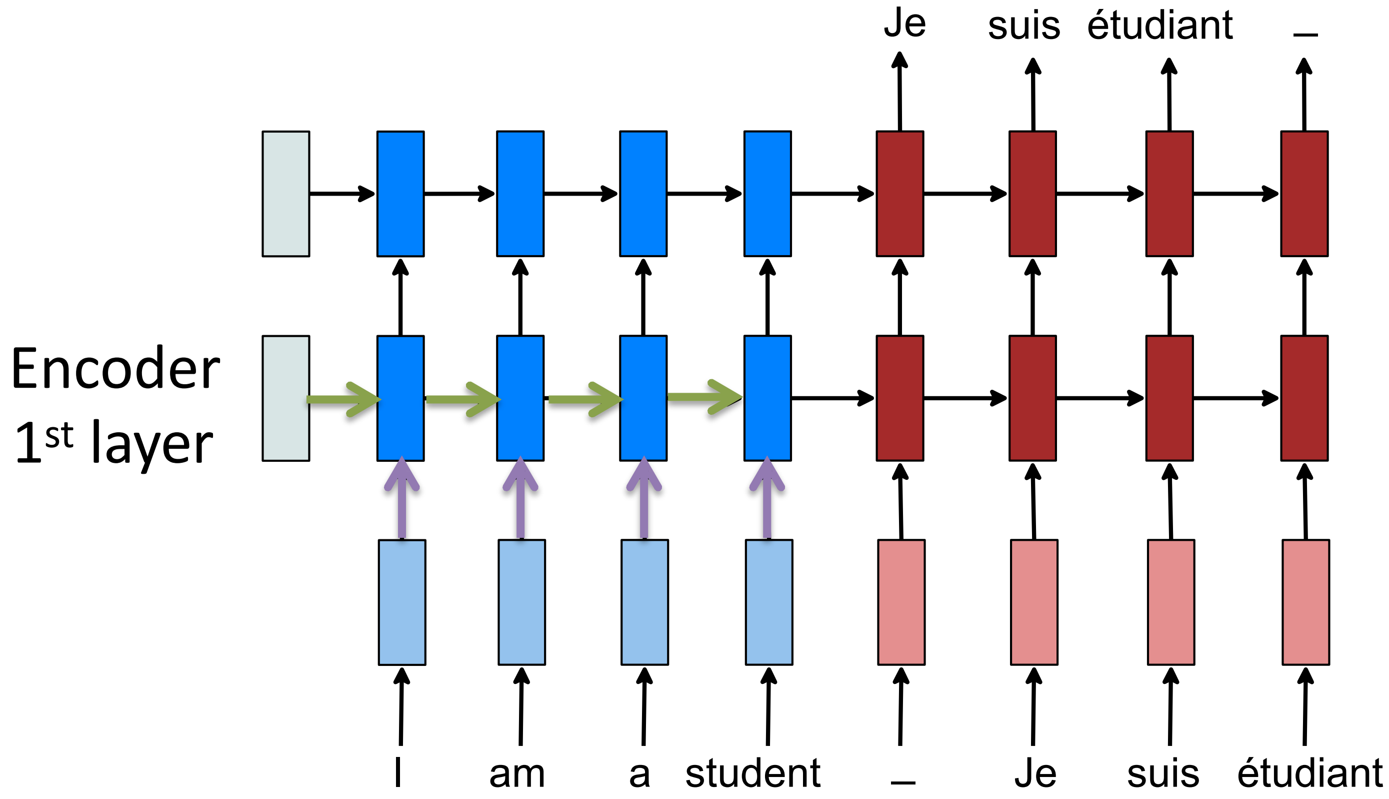
- One for each language: can learn from scratch.

Recurrent Connections



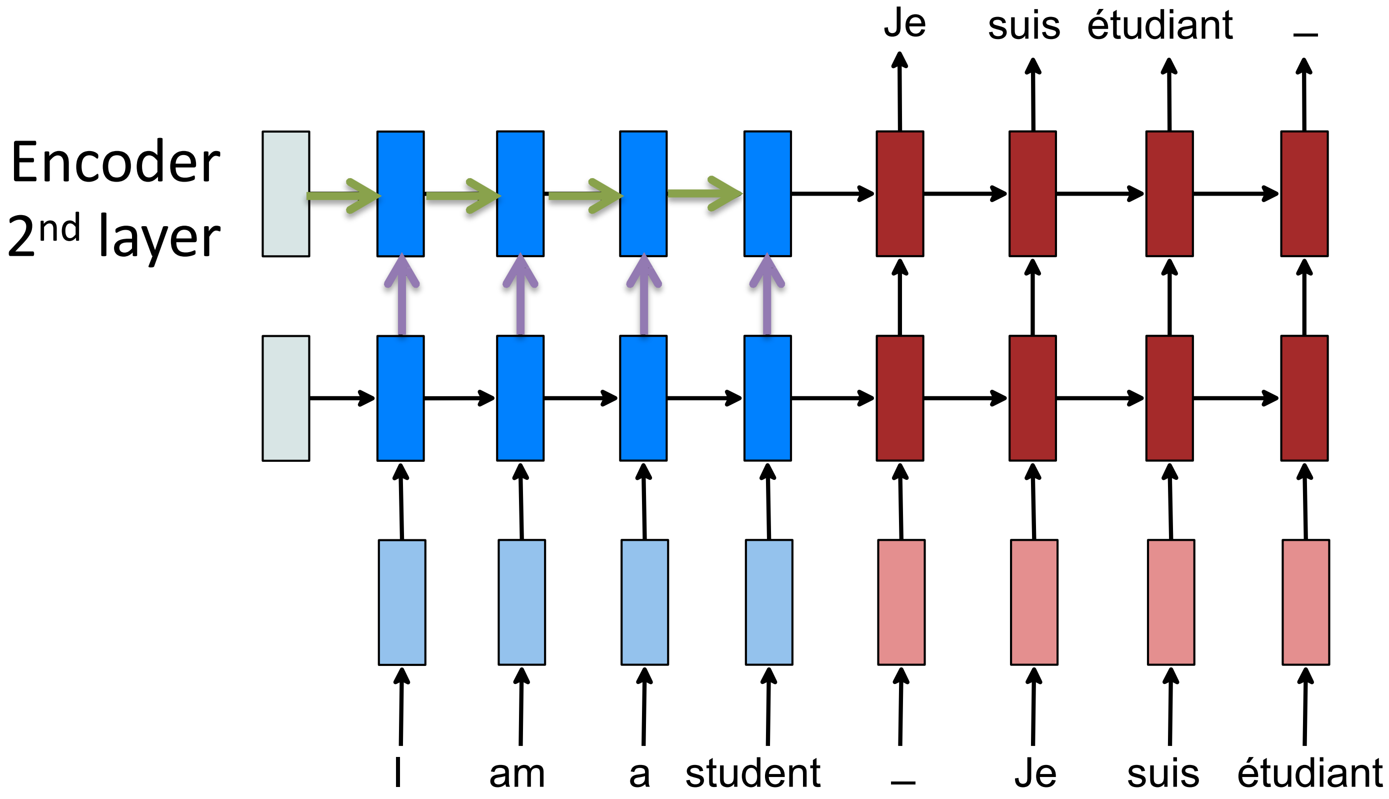
- Often set to 0.

Recurrent Connections



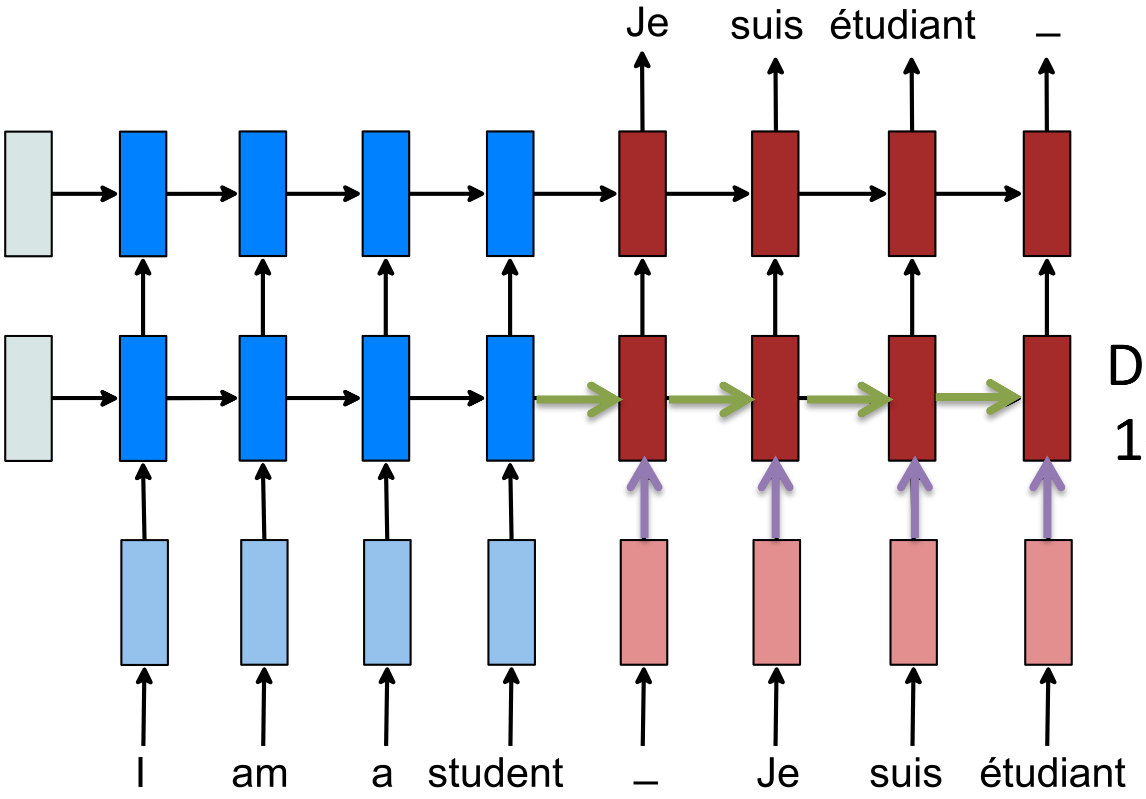
- Different: {1st layer, 2nd layer} x {encoder, decoder}.

Recurrent Connections



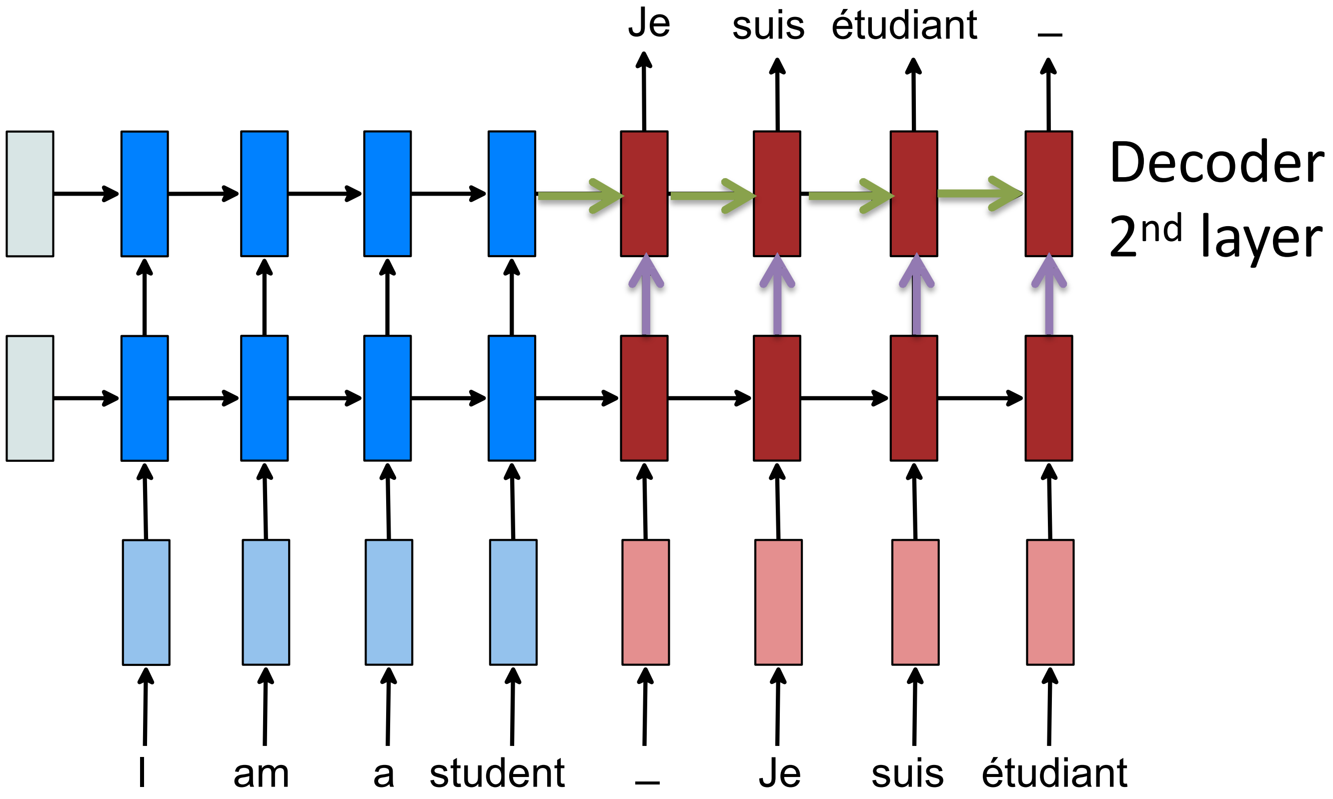
- Different: {1st layer, 2nd layer} x {encoder, decoder}.

Recurrent Connections



- Different: {1st layer, 2nd layer} x {encoder, decoder}.

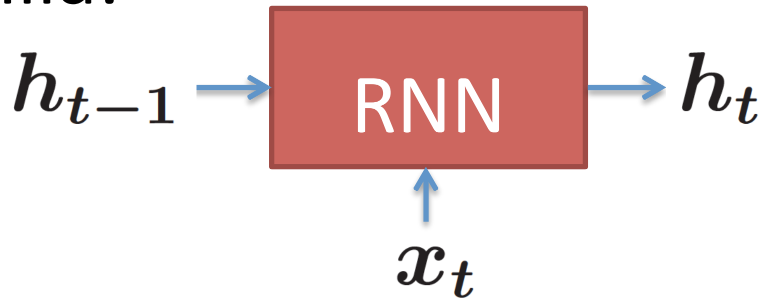
Recurrent Connections



- Different: {1st layer, 2nd layer} x {encoder, decoder}.

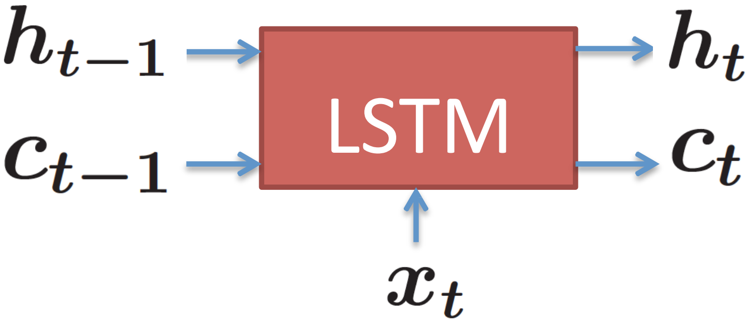
Recurrent Units

- Vanilla:



Vanishing gradient problem!

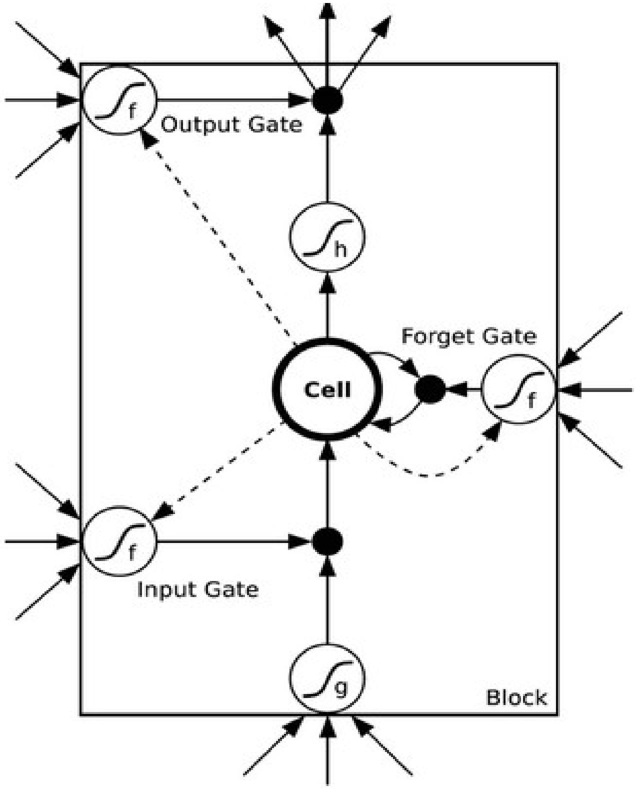
- LSTM:



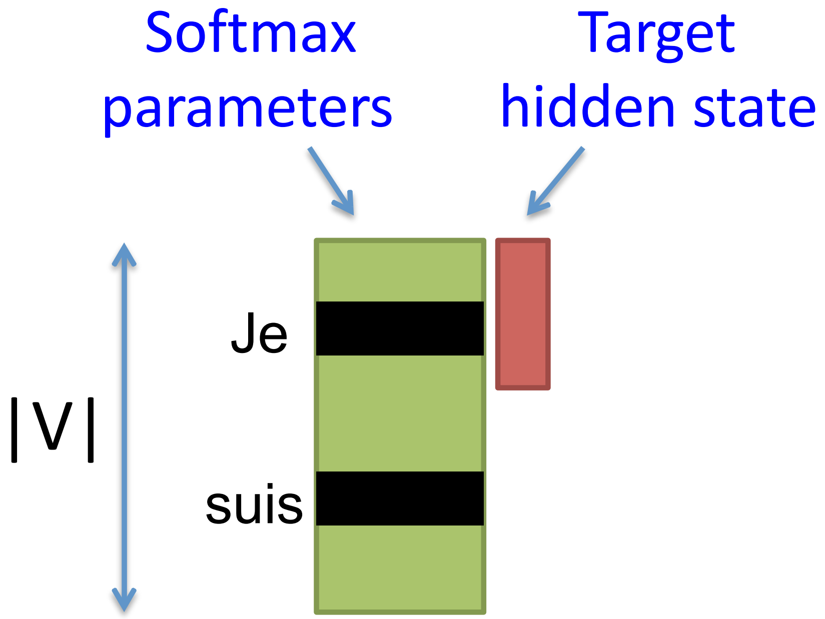
LSTM

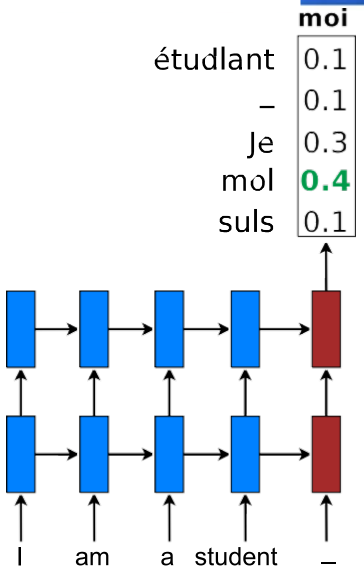
- LSTM Unit
  - a complex hidden unit
  - capable of remembering information over a long span of time steps

- **Unit Components**
  - cell store history information
  - forget gate whether to discard information in the cell
  - input gate whether to update the cell state
  - output gate whether to make an output
  - peephole make gates accept inputs from the cell

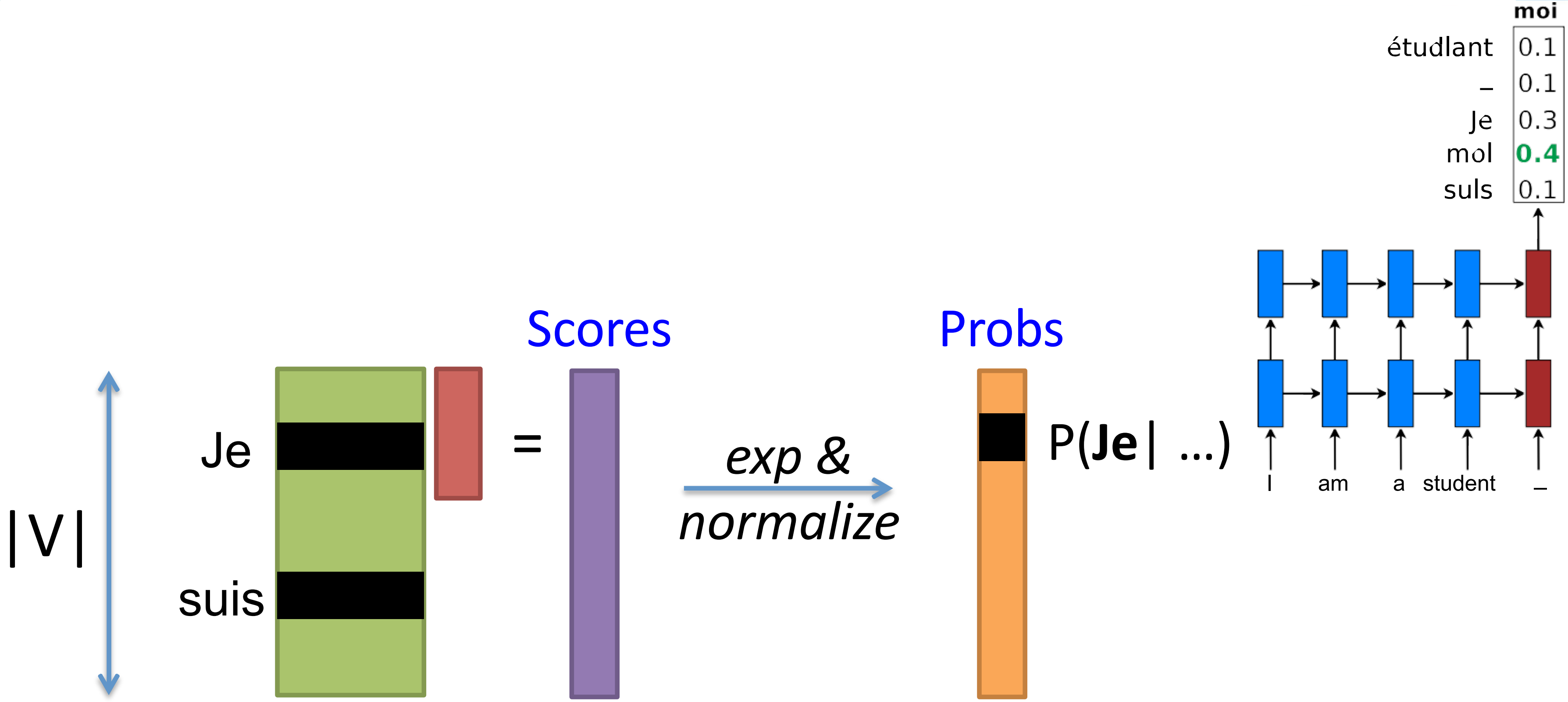


Softmax: *vectors ↦ categories*



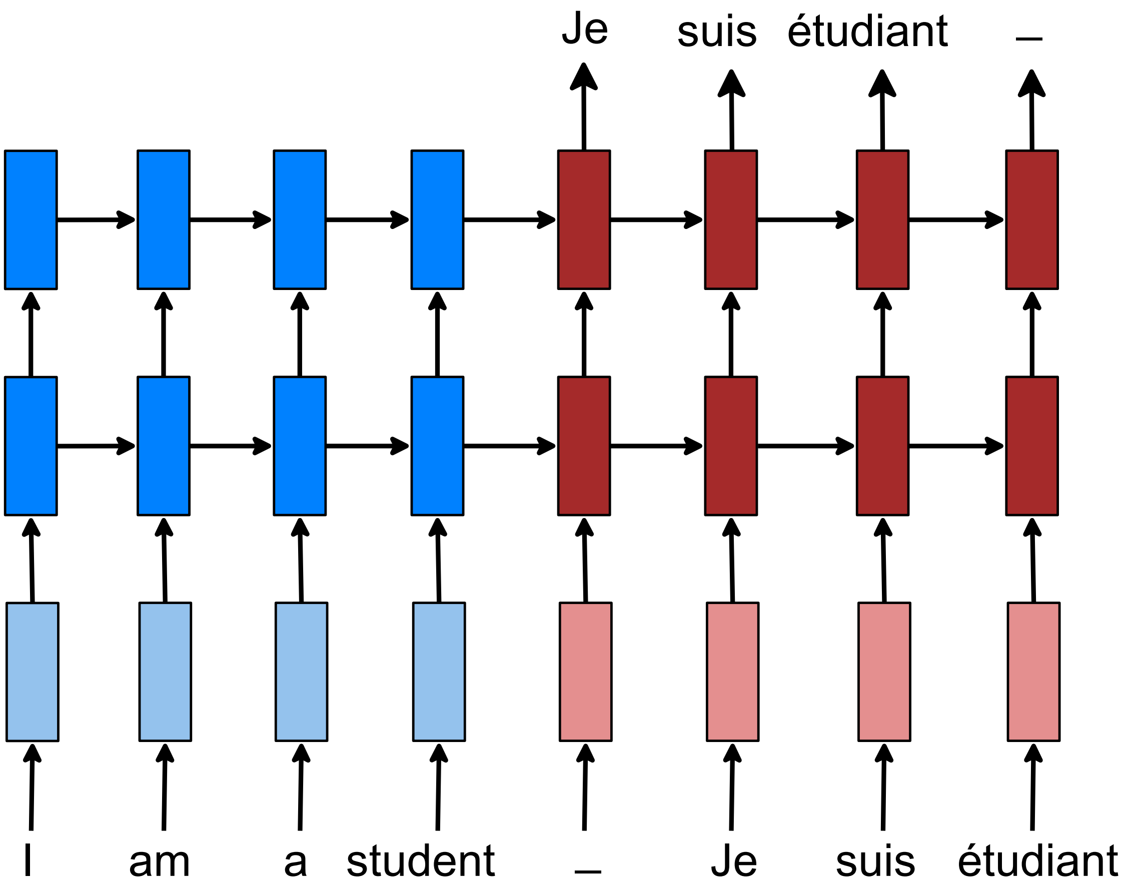


Softmax: *vectors* ↦ *categories*



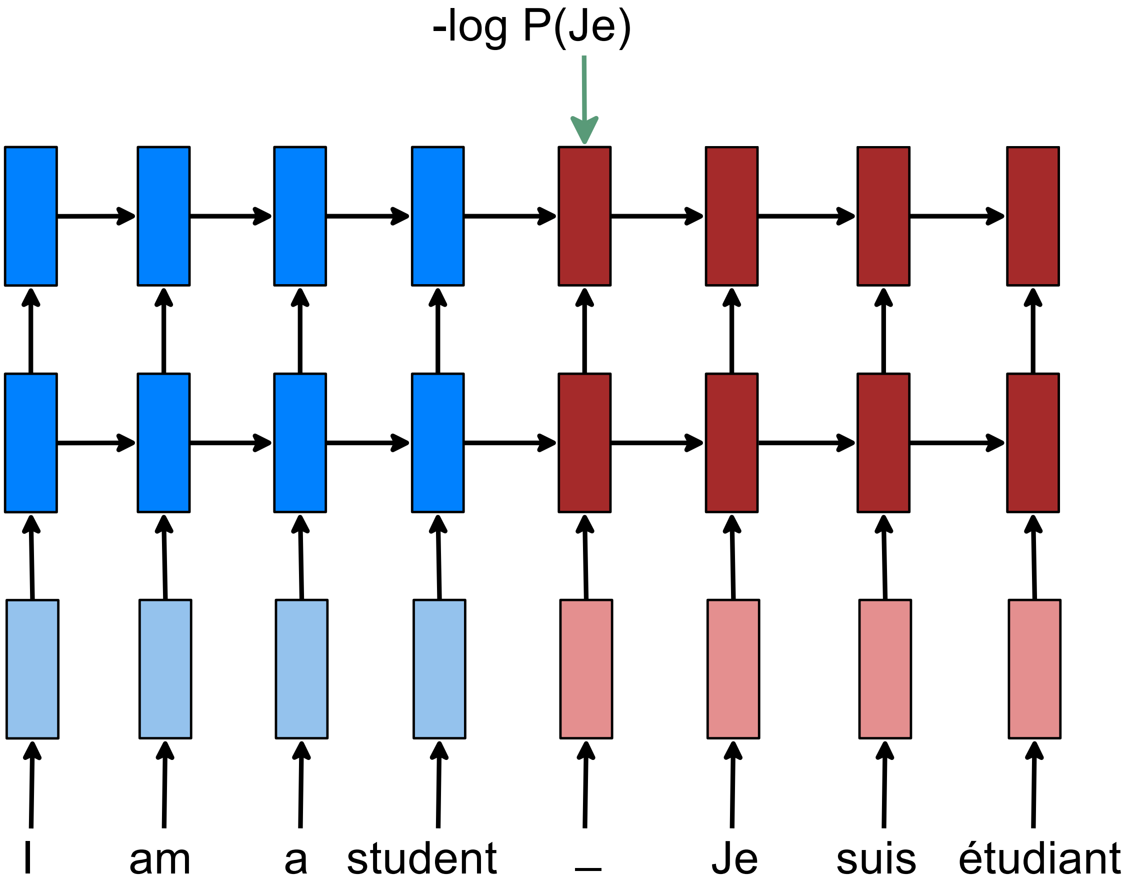
- Hidden states ↦ scores ↦ probabiliHes.

Training Loss



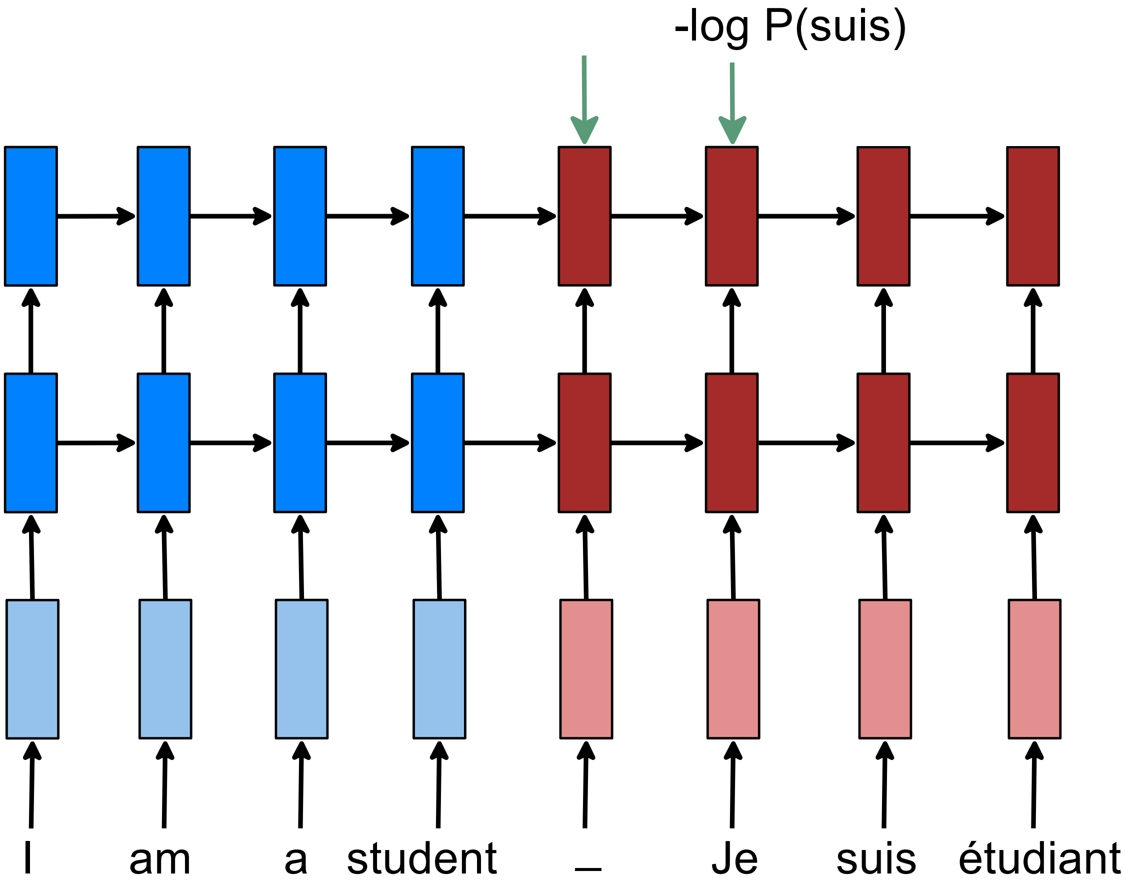
- Maximize P(target | source)

Training Loss



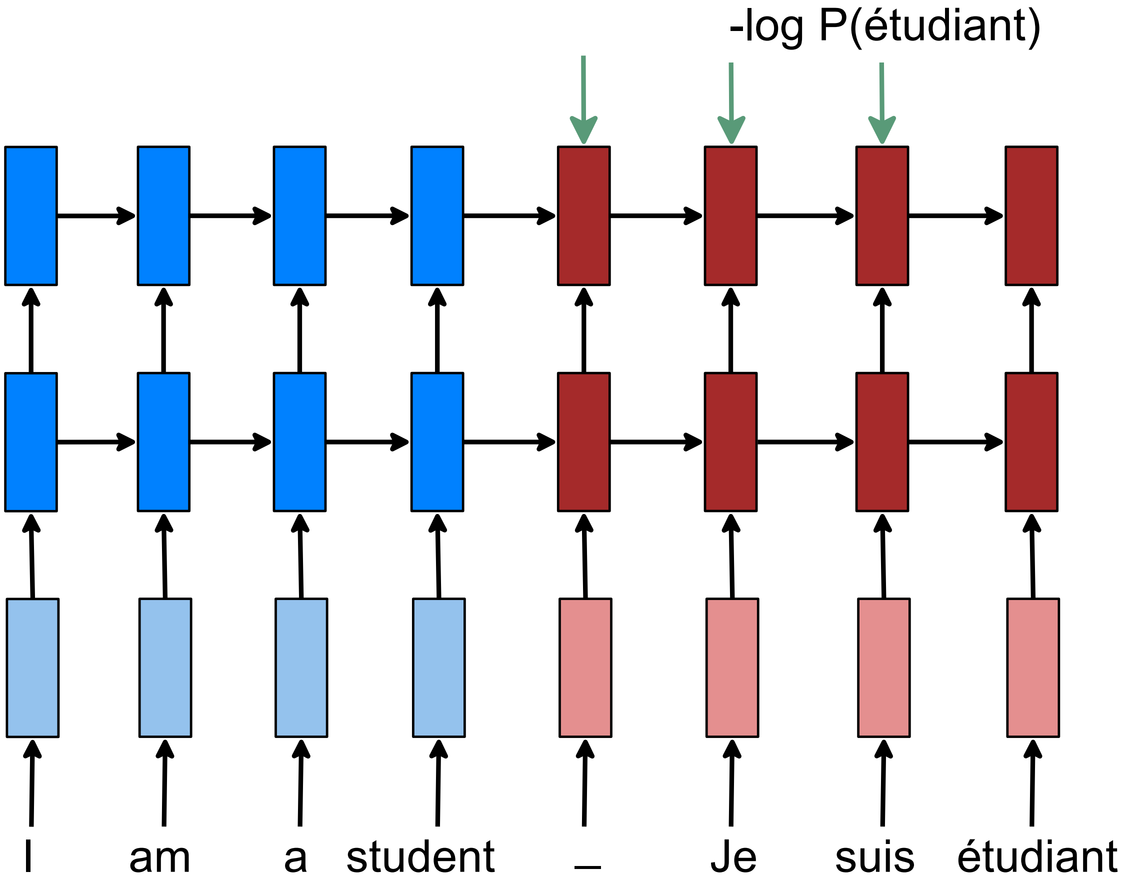
- Sum of all individual losses

Training Loss



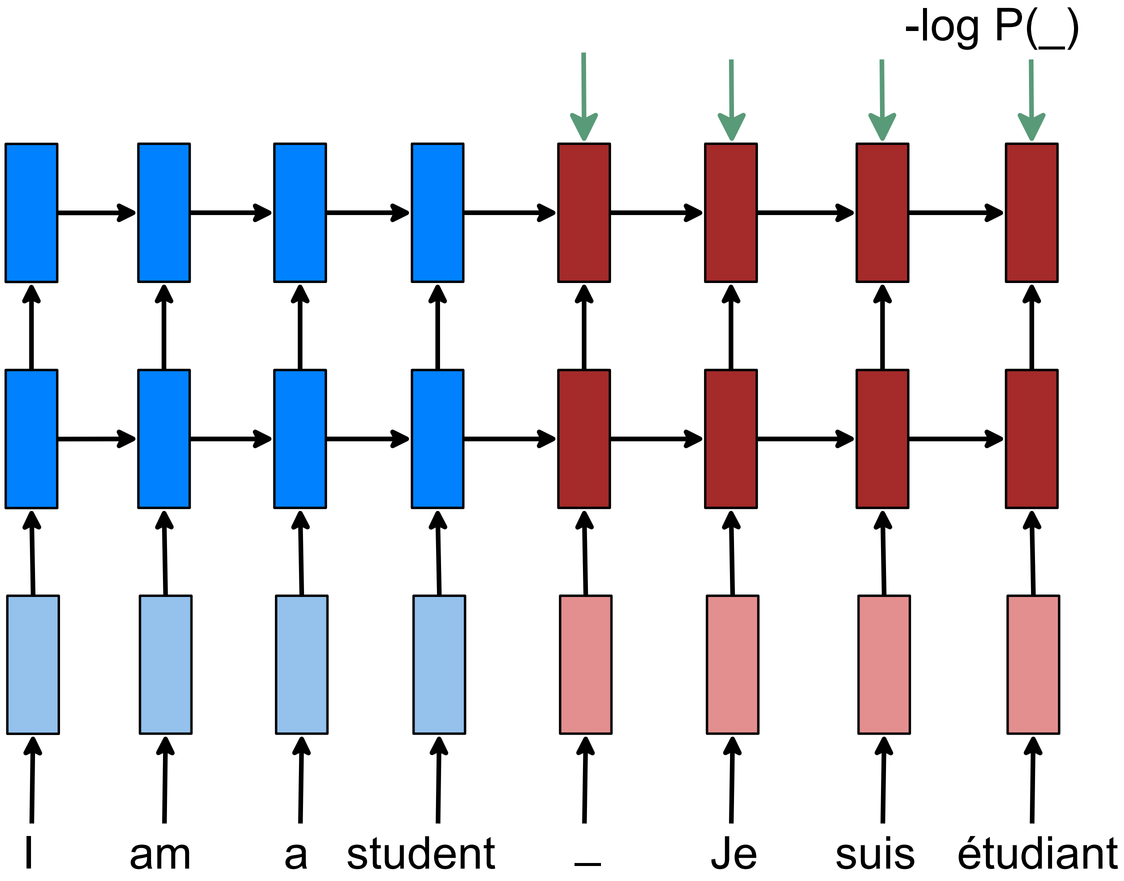
- Sum of all individual losses

Training Loss



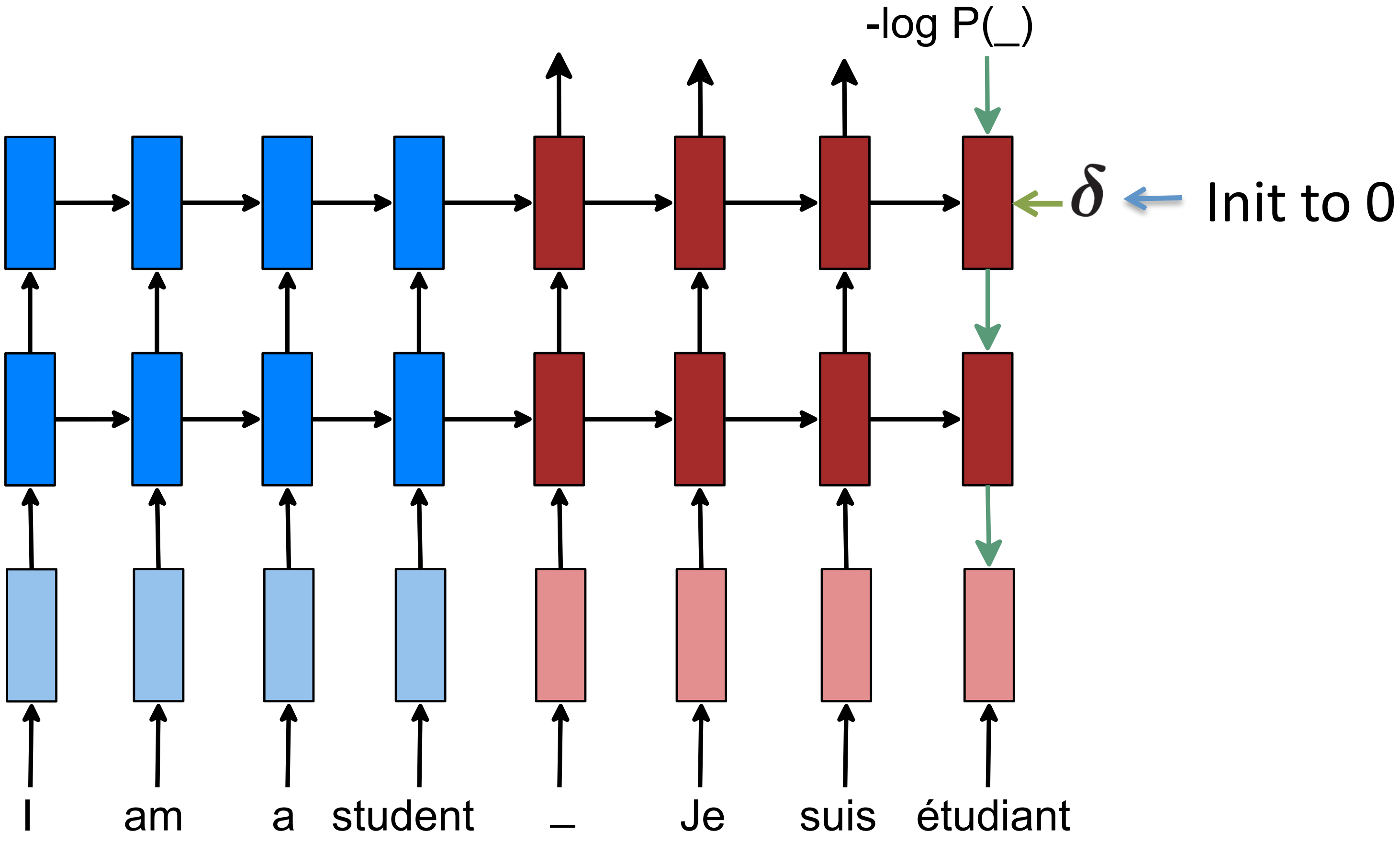
- Sum of all individual losses

Training Loss

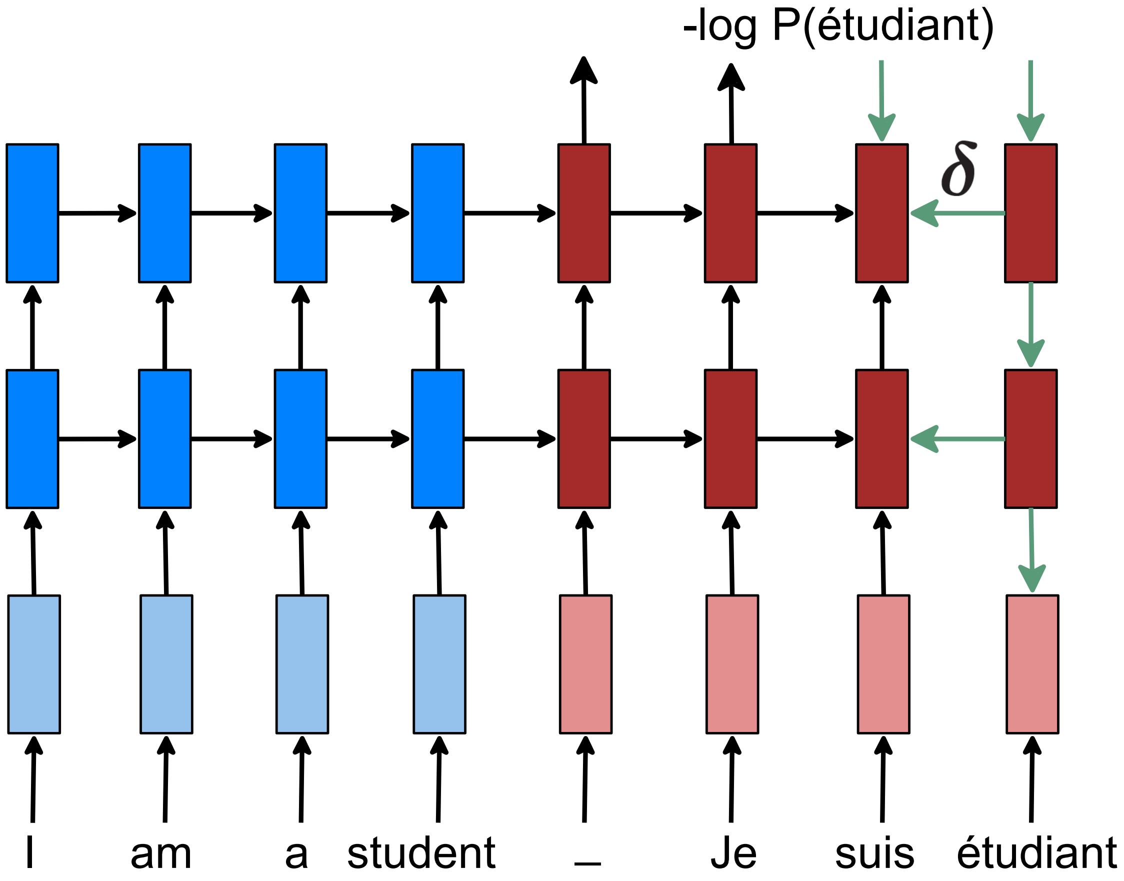


- Sum of all individual losses

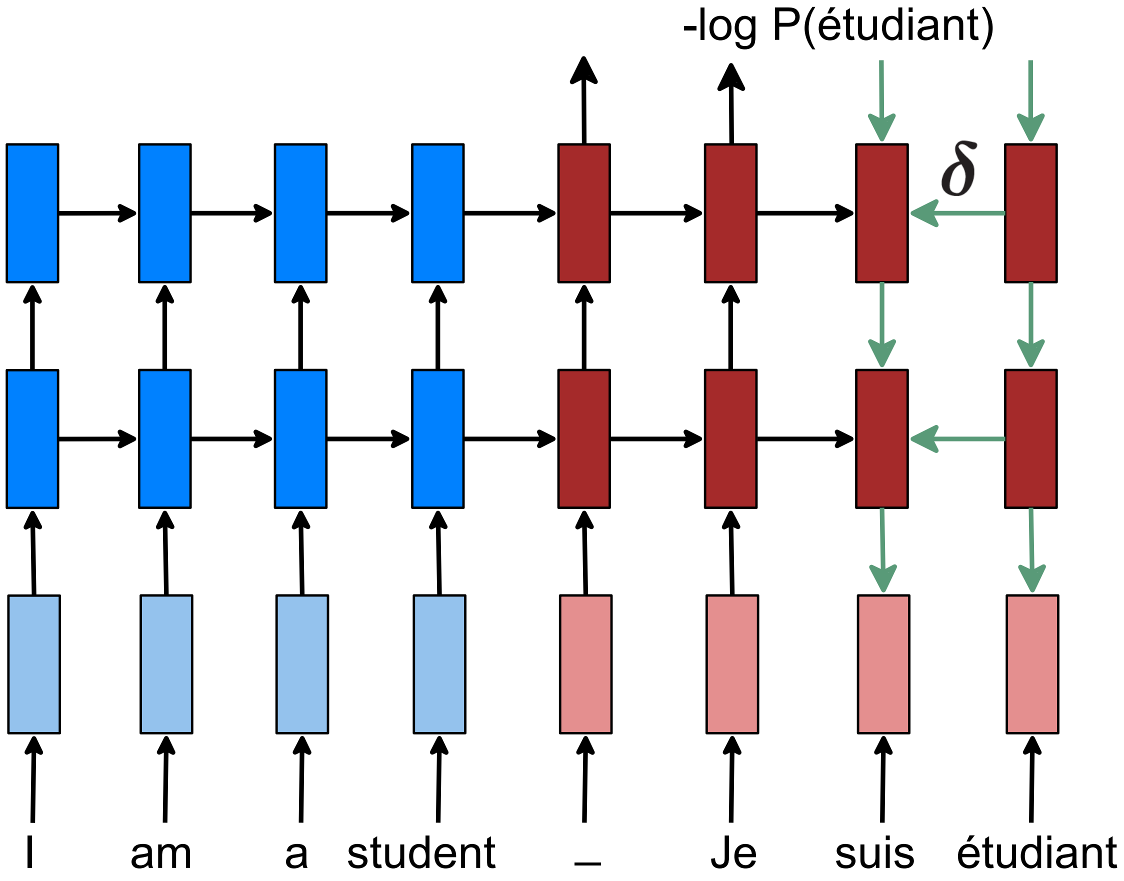
Backpropagation Through Time



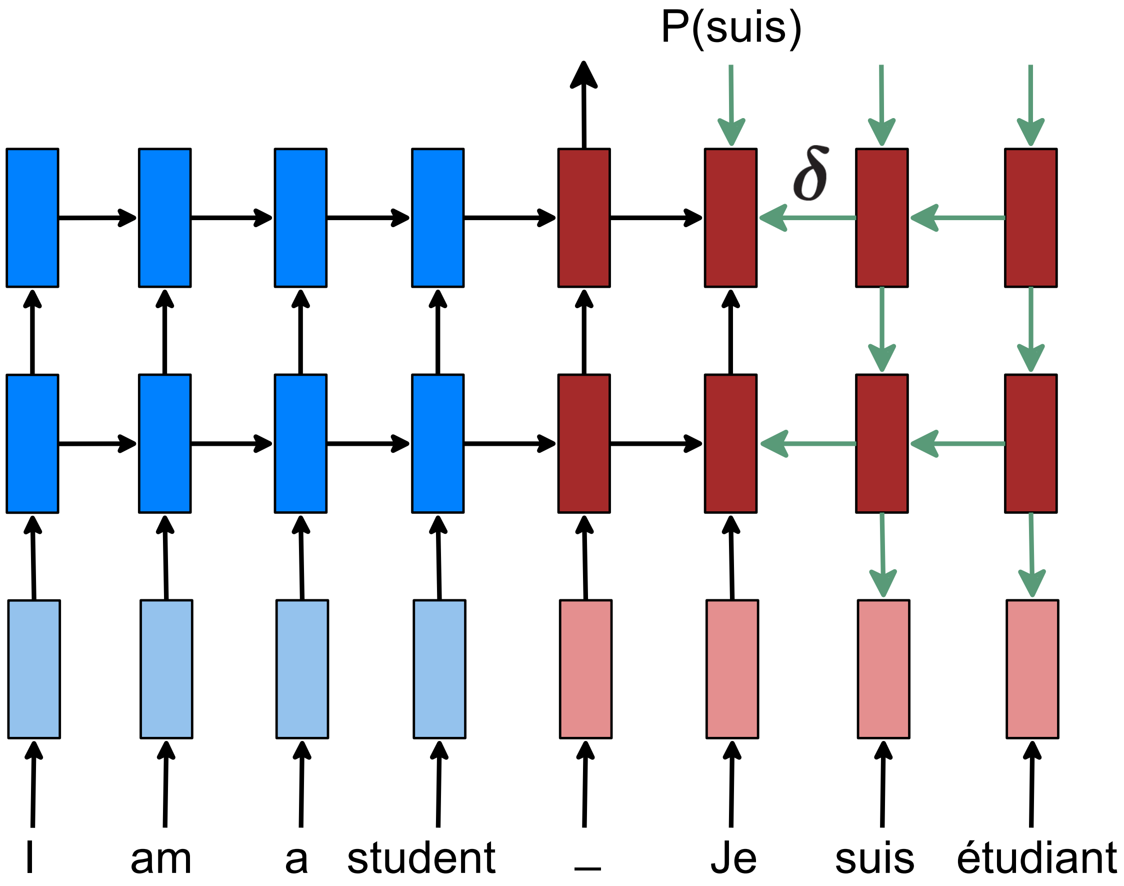
Backpropagation Through Time



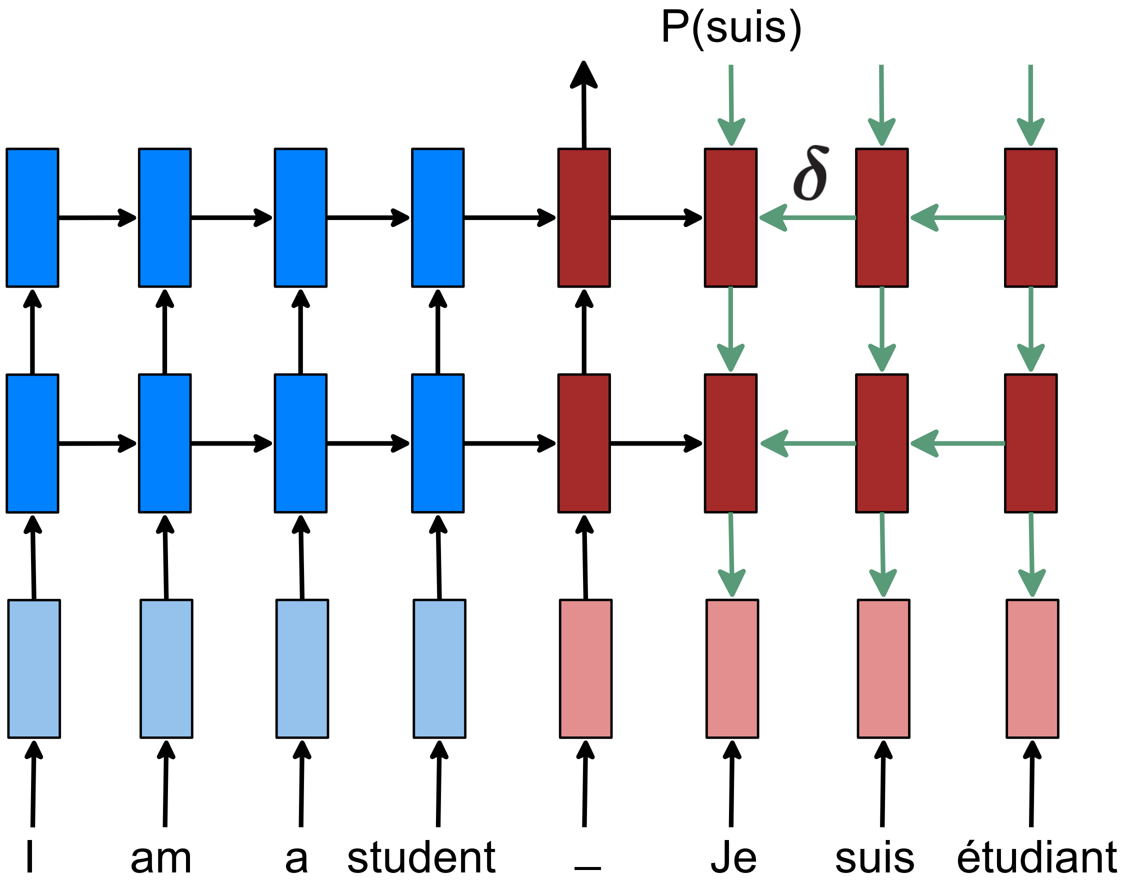
Backpropagation Through Time



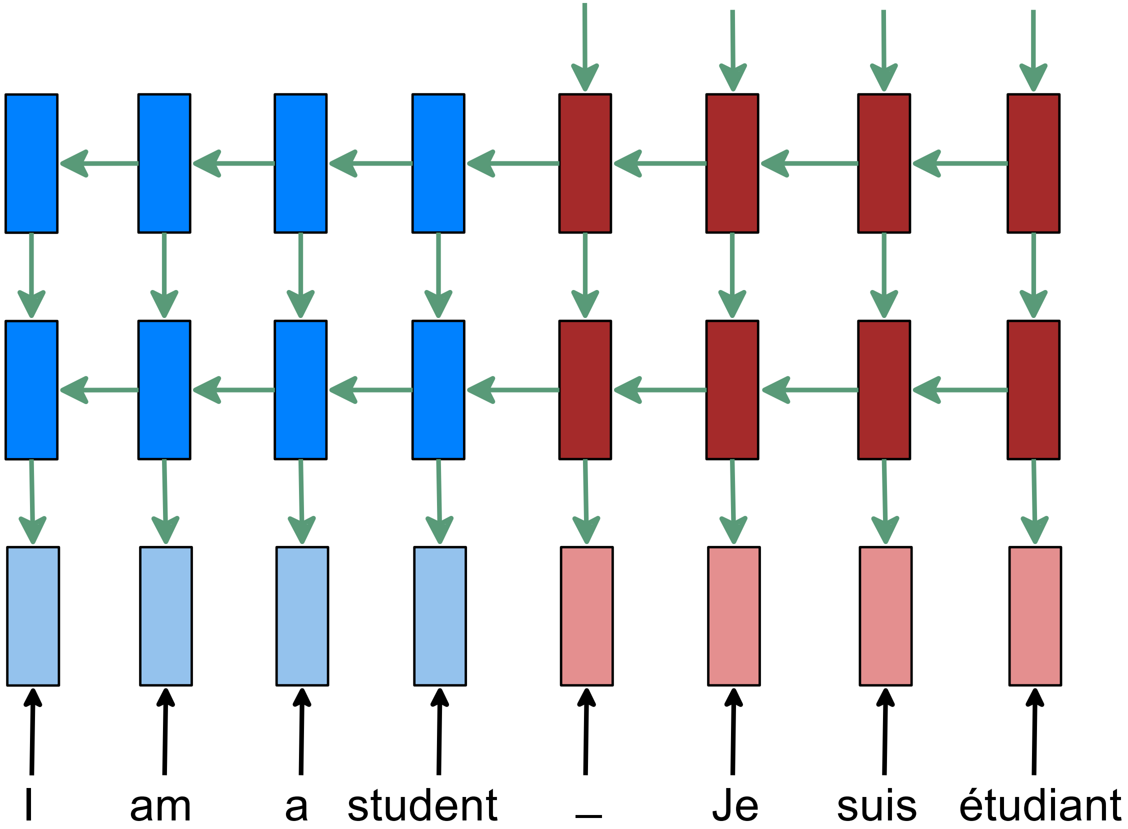
Backpropagation Through Time



Backpropagation Through Time



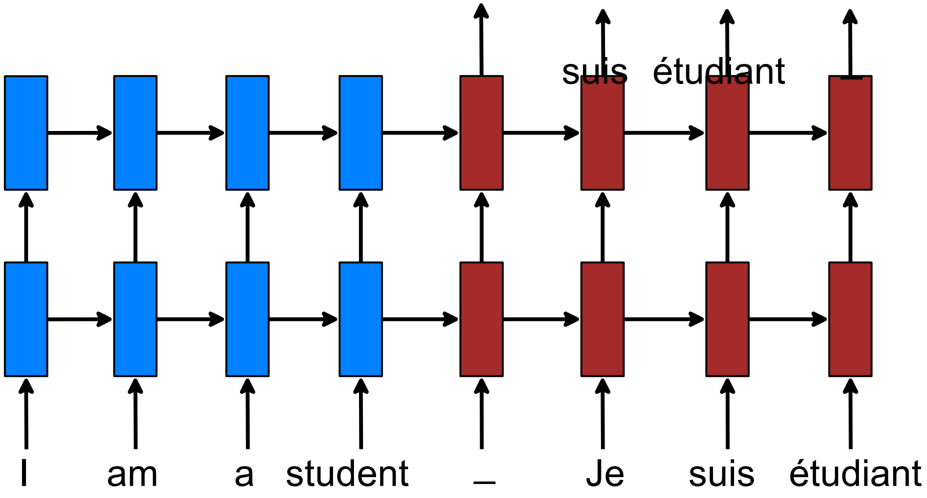
Backpropagation Through Time



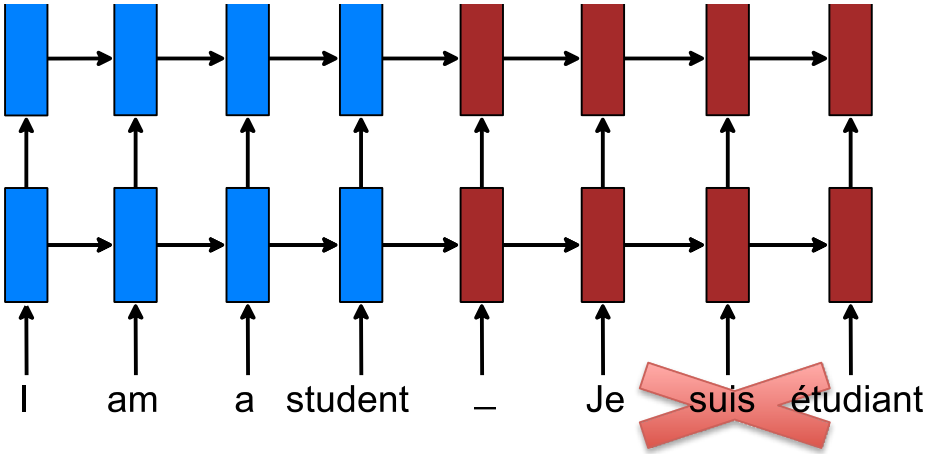
RNN gradients are accumulated.

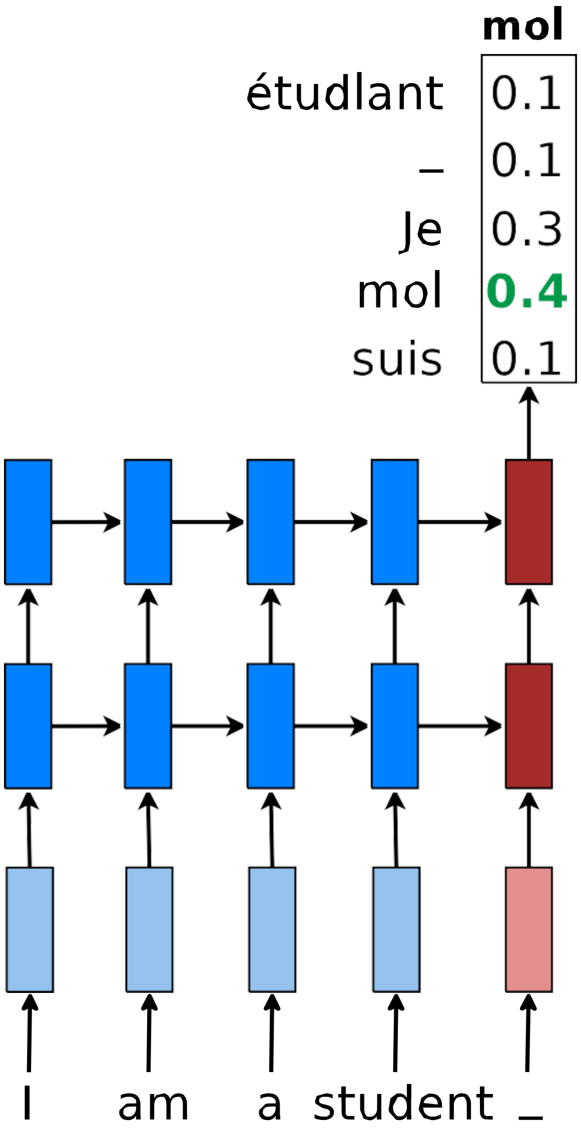
Training & Testing

- *Training*
  - Correct translations are available.

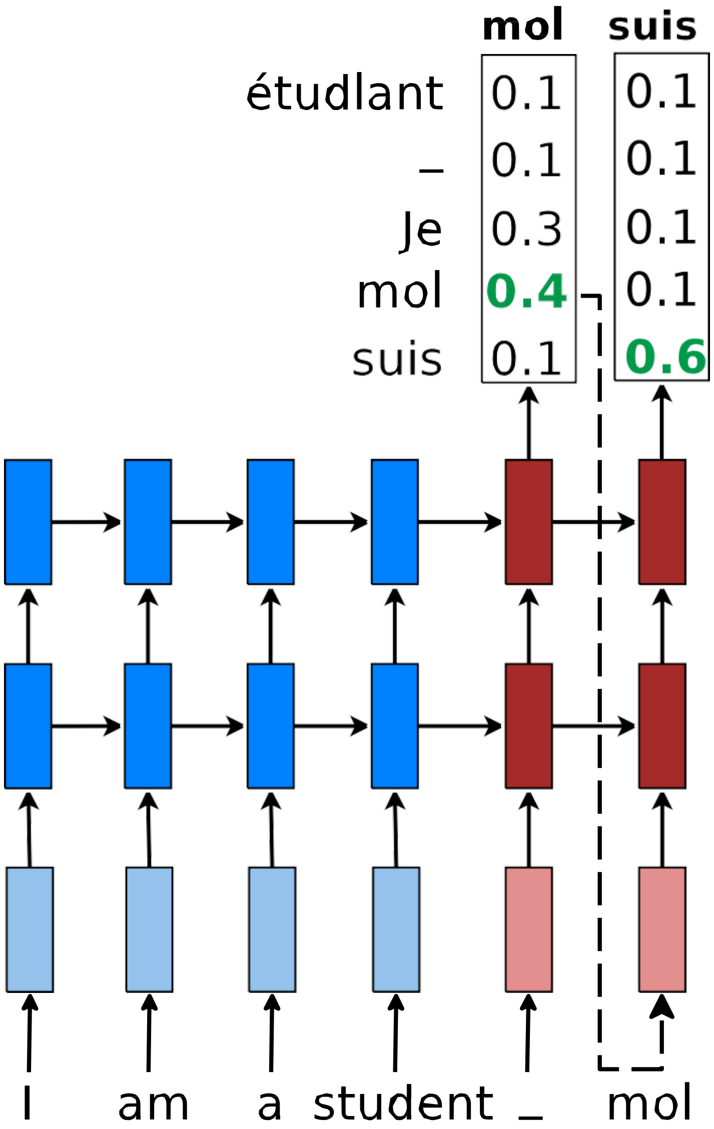


- *Testing*
  - Only source sentences are given.

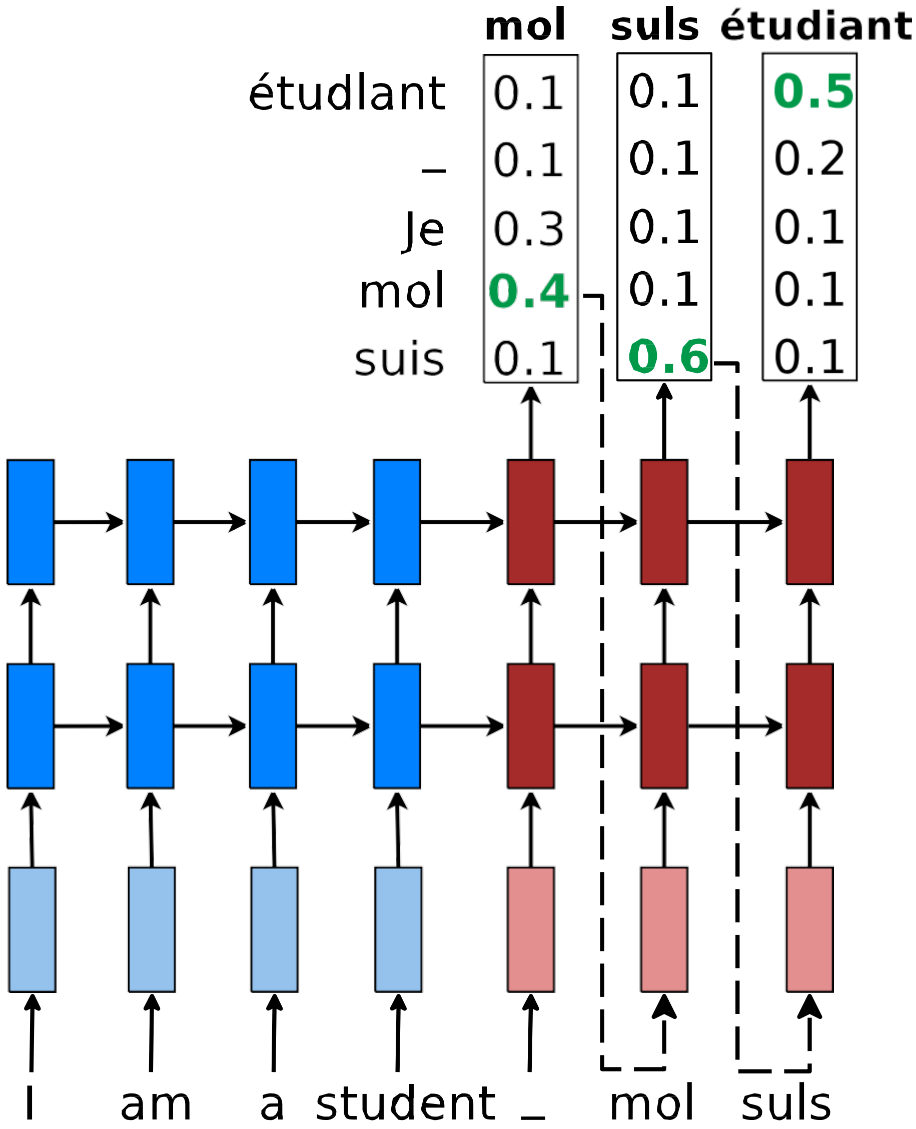




- Feed the most likely word

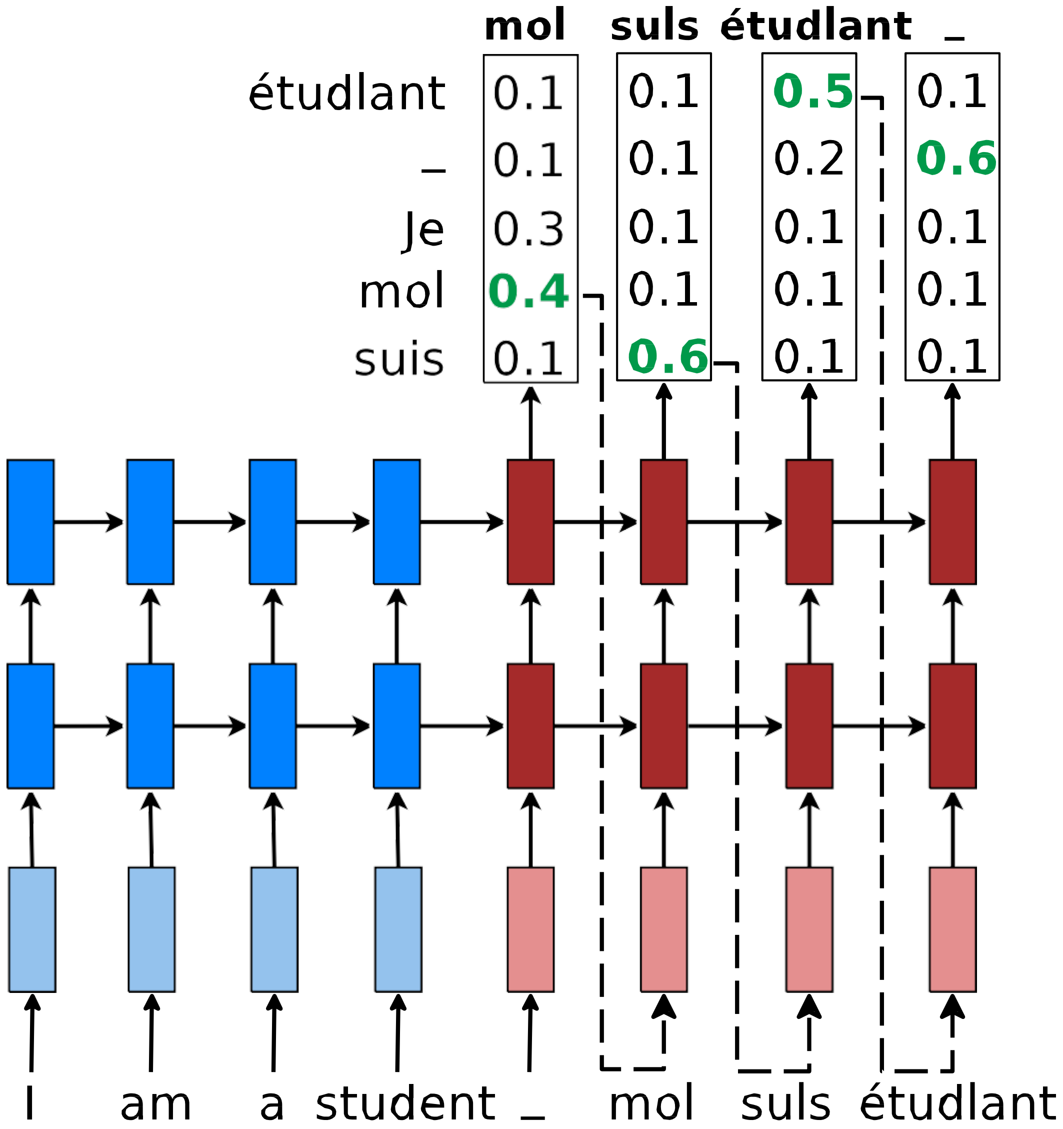


- Feed the most likely word



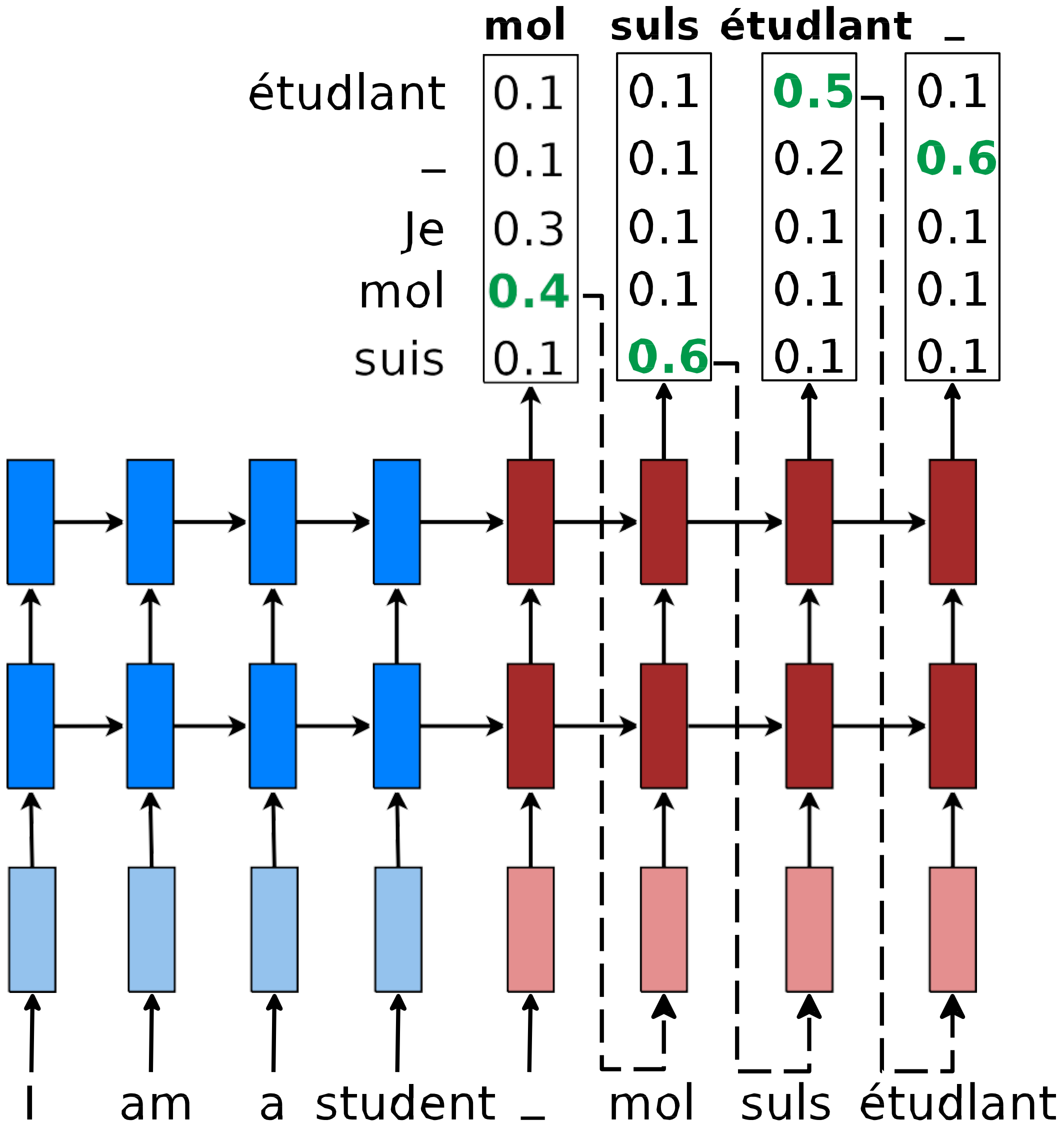
	mol	suls	étudiant
étudlant	0.1	0.1	**0.5**
_	0.1	0.1	0.2
Je	0.3	0.1	0.1
mol	**0.4**	0.1	0.1
suis	0.1	**0.6**	0.1

- Feed the most likely word



	mol	suls	étudlant	_
étudlant	0.1	0.1	**0.5**	0.1
_	0.1	0.1	0.2	**0.6**
Je	0.3	0.1	0.1	0.1
mol	**0.4**	0.1	0.1	0.1
suis	0.1	**0.6**	0.1	0.1

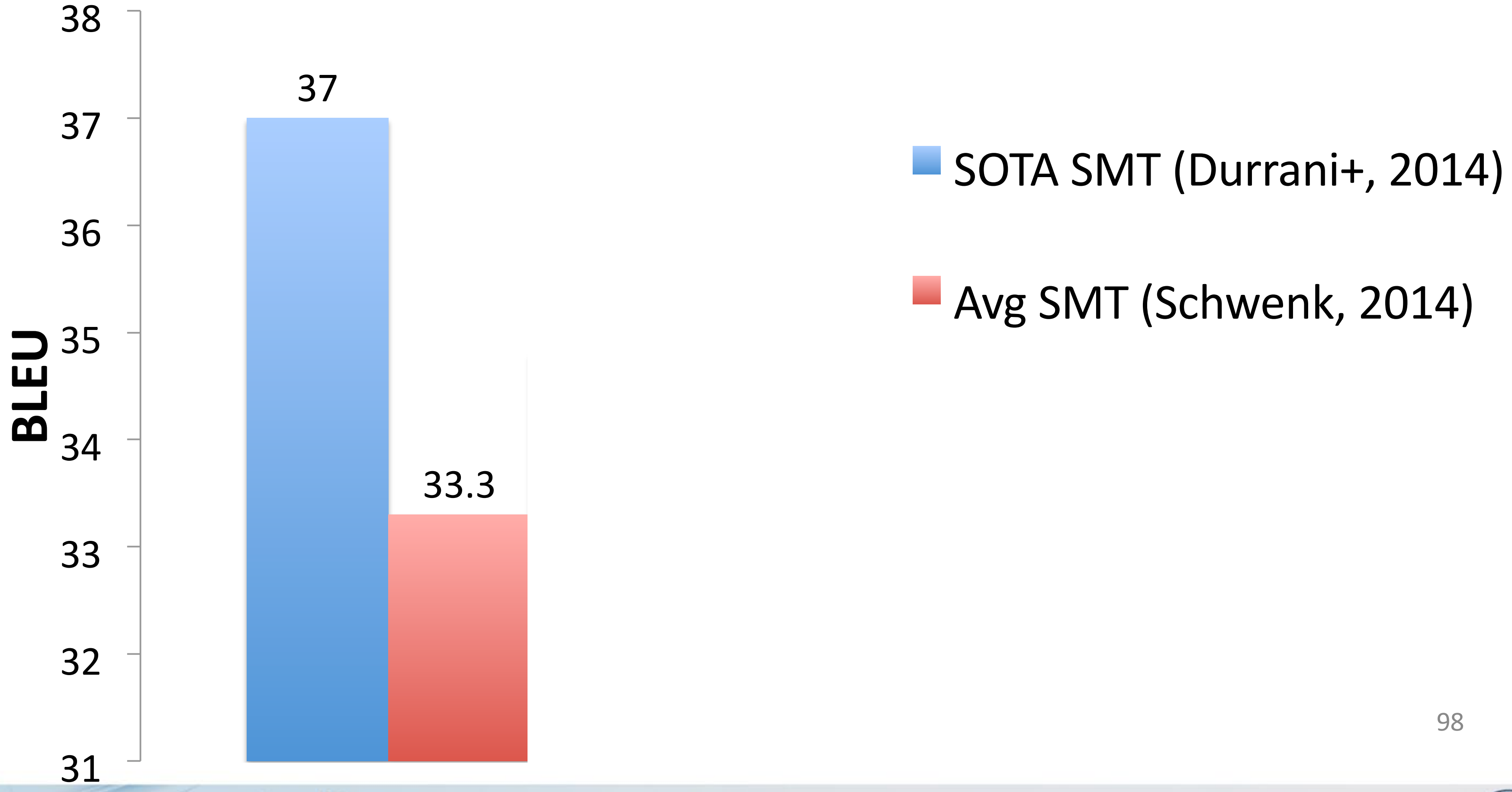
- Feed the most likely word



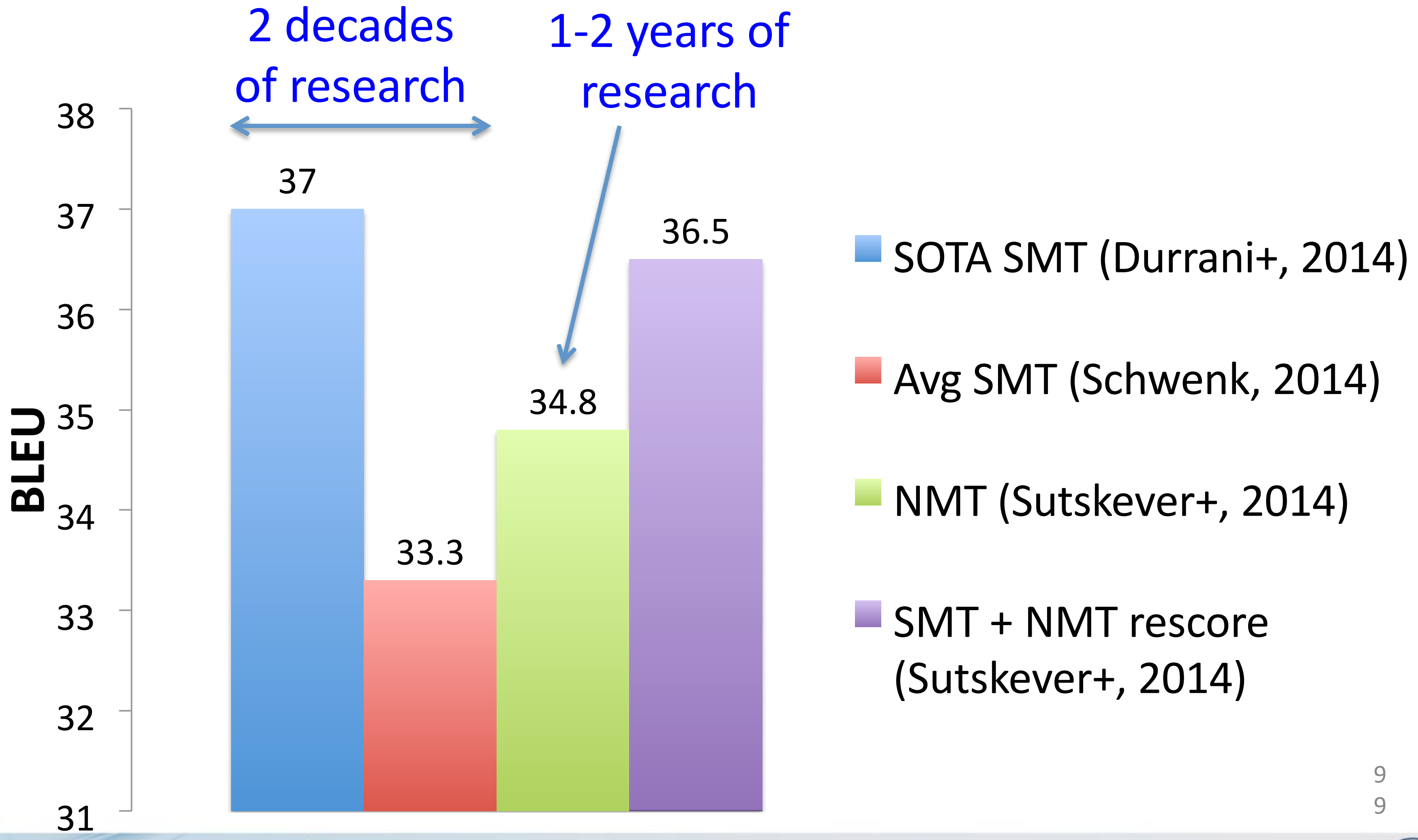
	mol	suls	étudlant	_
étudlant	0.1	0.1	**0.5**	0.1
_	0.1	0.1	0.2	**0.6**
Je	0.3	0.1	0.1	0.1
mol	**0.4**	0.1	0.1	0.1
suis	0.1	**0.6**	0.1	0.1

Simple beam-search decoders!

English-French WMT'14 results



English-French WMT'14 results



2 decades of research

1-2 years of research

Encoder-decoder Variants

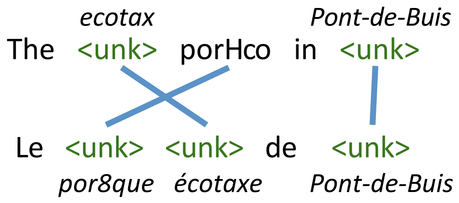
	Encoder	Decoder
(Sutskever et al., 2014)	Deep LSTM	Deep LSTM
(Cho et al., 2014) (Bahdanau et al., 2015) (Jean et al., 2015)	(Bidirectional) GRU	GRU
(Kalchbrenner & Blunsom, 2013)	CNN	(Inverse CNN) RNN

Limitations

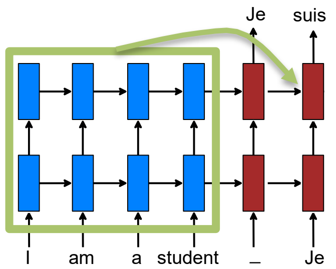
- #1: the ***vocabulary size*** problem
  - *Goal*: extend the vocabulary coverage.
- #2: the ***sentence length*** problem
  - *Goal*: translate long sentences better.
- #3: the ***language complexity*** problem
  - *Goal*: handle more language variations.

Advancing NMT

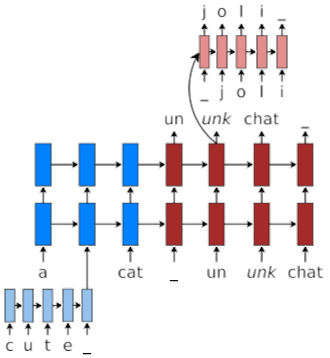
- #1: the ***vocabulary size*** problem
  - *Sol*: "copy" mechanism.



- #2: the ***sentence length*** problem
  - *Sol*: attention mechanism.



- #3: the ***language complexity*** problem
  - *Sol*: character-level translation.



j o l i _
_ j o l i
un *unk* chat _
a cat _ un *unk* chat
c u t e _