

תרגום מכונה: אז איך זה עובד?



רועי אהרוני
המעבדה לעיבוד שפה טבעית
אוניברסיטת בר-אילן

על מה נדבר?

- ❖ תרגום ולמה זה קשה
- ❖ קצת הסטוריה
- ❖ איך מתרגמים חייזרית
- ❖ איך מחשב מתרגם
- ❖ סיכום ושאלות



מה זה תרגום טוב?

- ❖ מעבר בין שפות
- ❖ שמירה על משמעות
- ❖ משפטים תקינים
- ❖ פרופיל כותב
- ❖ ועוד...



למה זה קשה?

“I sat on the bank” ❖

“ראיתי איש קרח” ❖

“ספר עזר לרופא בהוצאת כתר” ❖

“פיתה עם לבנה” ❖

“תה חזק” ❖

“Out of sight, out of mind” - “Invisible idiot” ❖

“The spirit is willing but the flesh is weak” - ❖

“The vodka is good but the meat is rotten”



מה כדאי לדעת?

❖ NLP/בלשנות

❖ סטטיסטיקה והסתברות

❖ Machine Learning

❖ (אלגוריתמים) בעיות חיפוש

❖ אופטימיזציה



היסטוריה וגישות

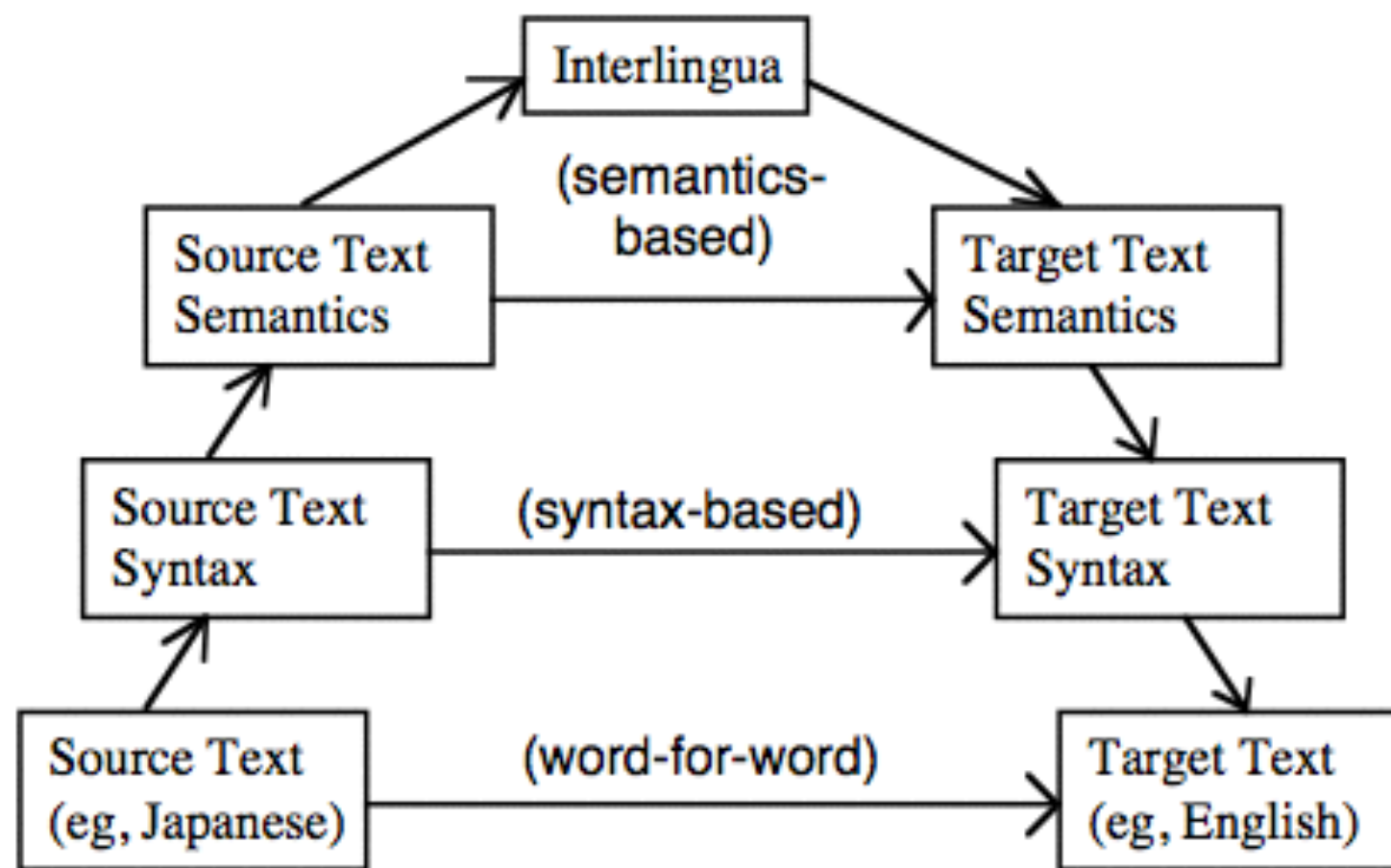


Figure 1: Different strategies for Machine Translation (MT).

דוגמא:

1a. ok-voon ororok sprok .
|

1b. at-voon bichat dat .

2a. ok-drubel ok-voon anak plok sprok .
| |

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

3a. erok sprok izok hihok ghrok .
/

3b. totat dat arrat vat hilat .

4a. ok-voon anak 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 eneas .

8a. lalok brok anak 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 ghrok klok .
/

10b. wat nnat gat mat bat hilat .

דוגמא:

1a. ok-voon ororok sprok .

1b. at-voon bichat dat .

2a. ok-drubel ok-voon anak 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 anak 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 eneas .

8a. lalok brok anak 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 klok .

10b. wat nnat gat mat bat hilat .

דוגמא:

Translation dictionary:

anok - pippat

erok - total

ghirok - hilat

hihok - arrat

izok - vat

ok-drubel - at-drubel

ok-yurp - at-yurp

ok-voon - at-voon

ororok - bichat

plok - rrat

sprok - dat

zanzanok - zanzanat

מודל
תרגום

Word pair counts:

1 . erok

7 . lalok

2 . ok-drubel

2 . ok-voon

3 . wiwok

1 anok drok

1 anok ghirok

1 hihok yorok

1 izok enemok

2 izok hihok

1 izok jok

1 izok kantok

1 izok stok

1 izok vok

מודל
שפה

אז מה היה לנו?

❖ כדי לתרגם השתמשנו ב

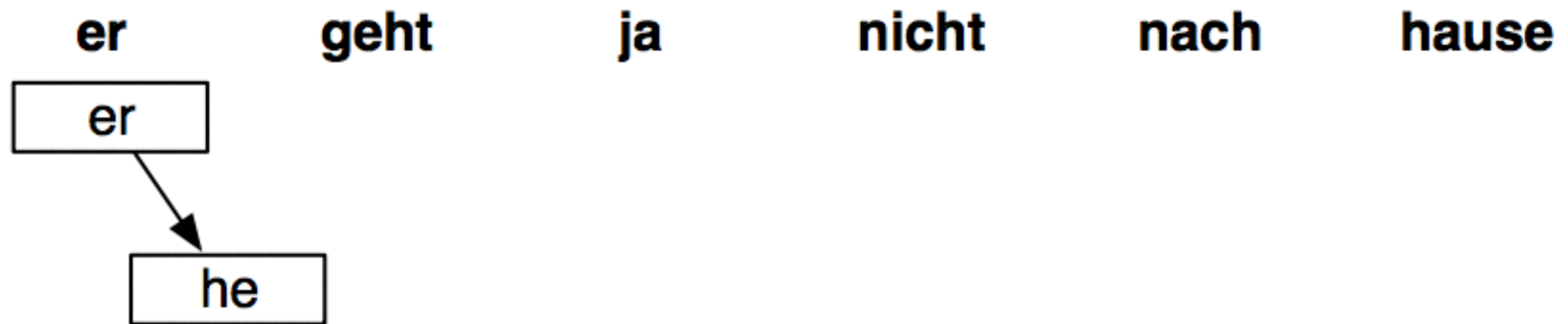
מודל תרגום: קורפוס מקבילי ← העמדה ← טבלת ביטויים

מודל שפה: קורפוס בשפת היעד ← טבלת תדירויות

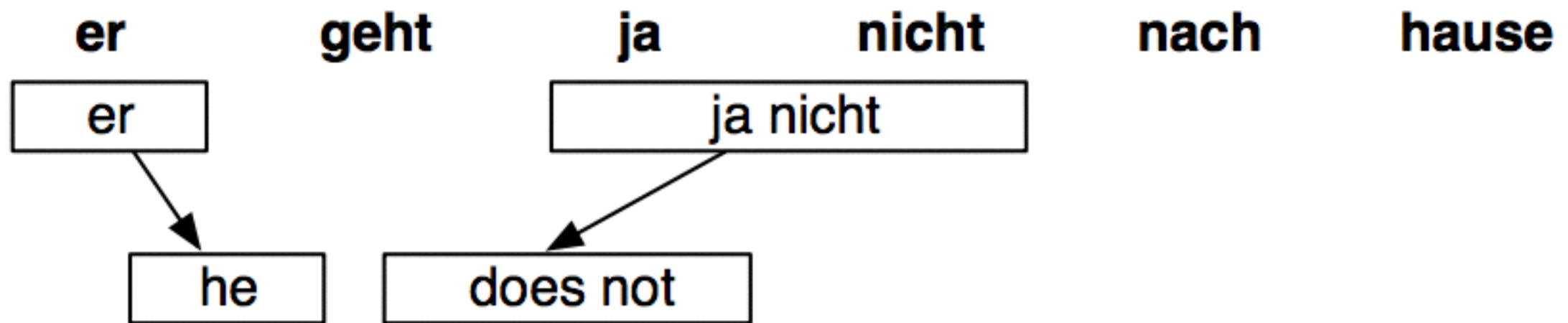
❖ ככה גם המחשב מתרגם



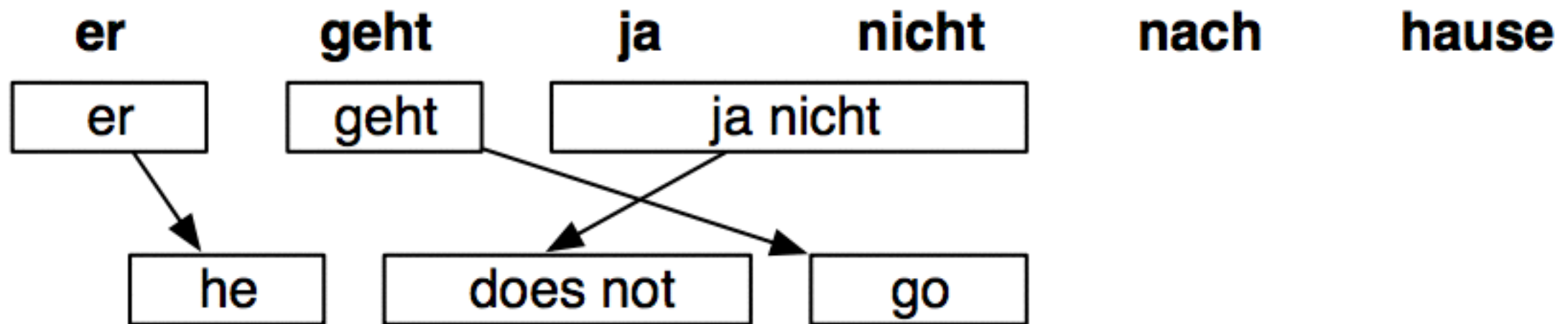
אז איך מוצאים את התרגום הכי טוב?



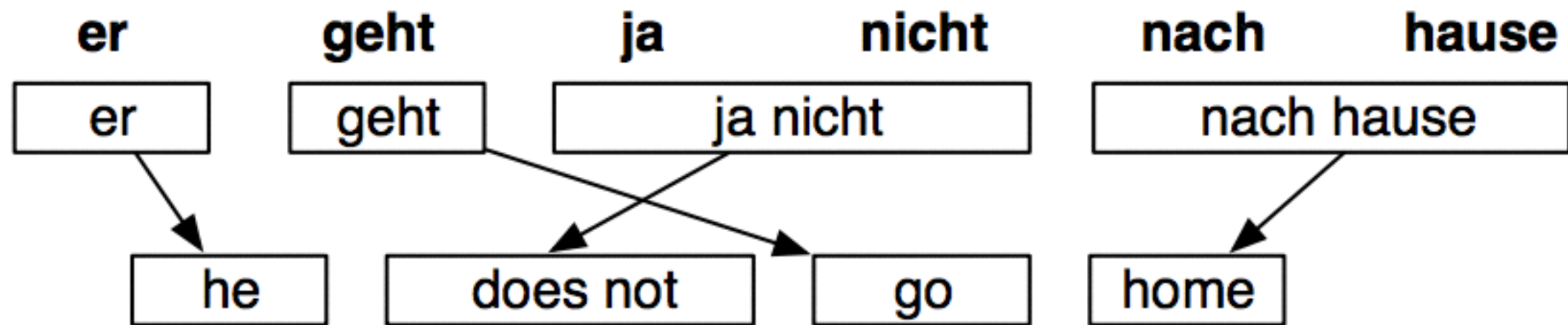
אז איך מוצאים את התרגום הכי טוב?



אז איך מוצאים את התרגום הכי טוב?



אז איך מוצאים את התרגום הכי טוב?

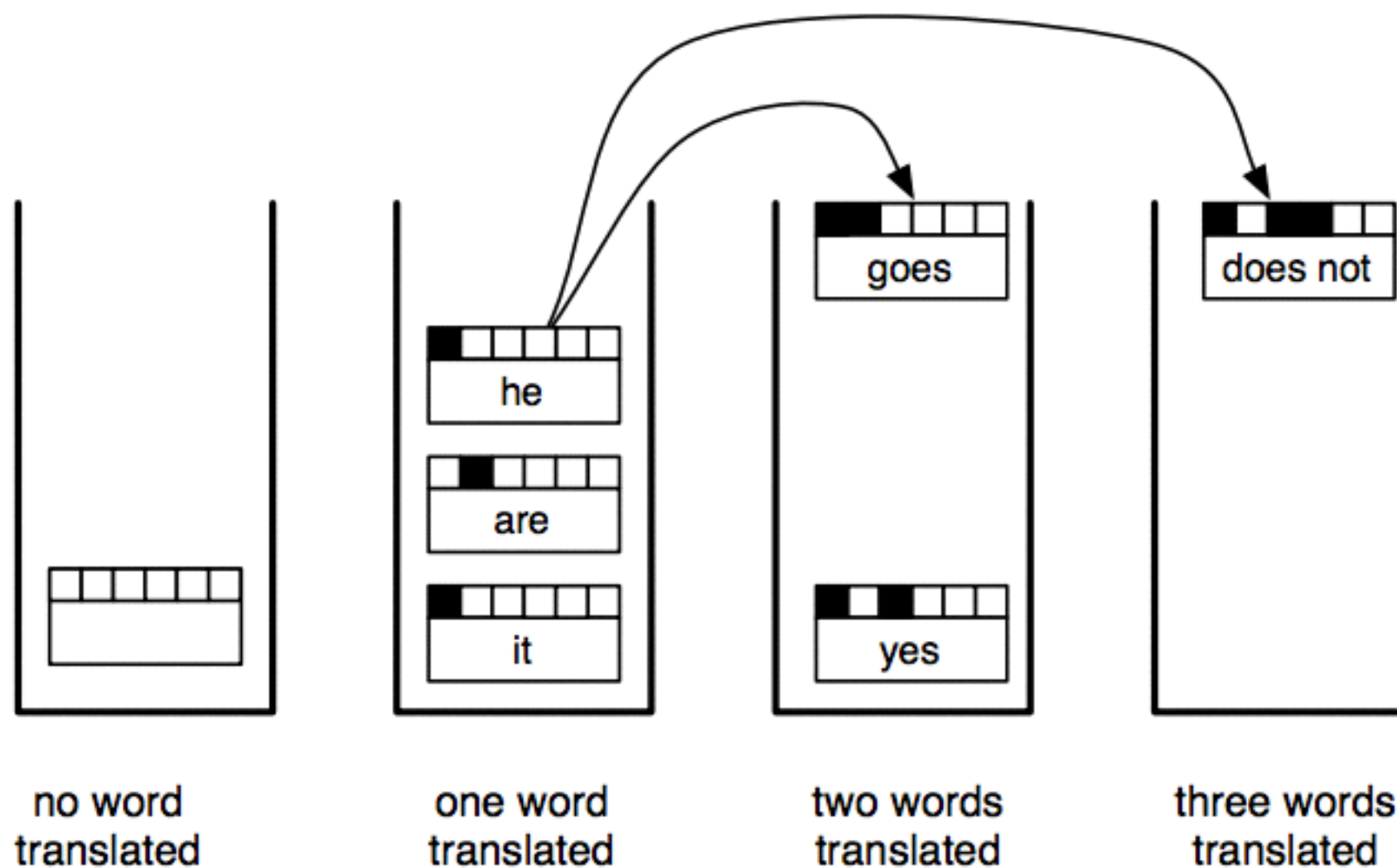


בעיה קטנה...

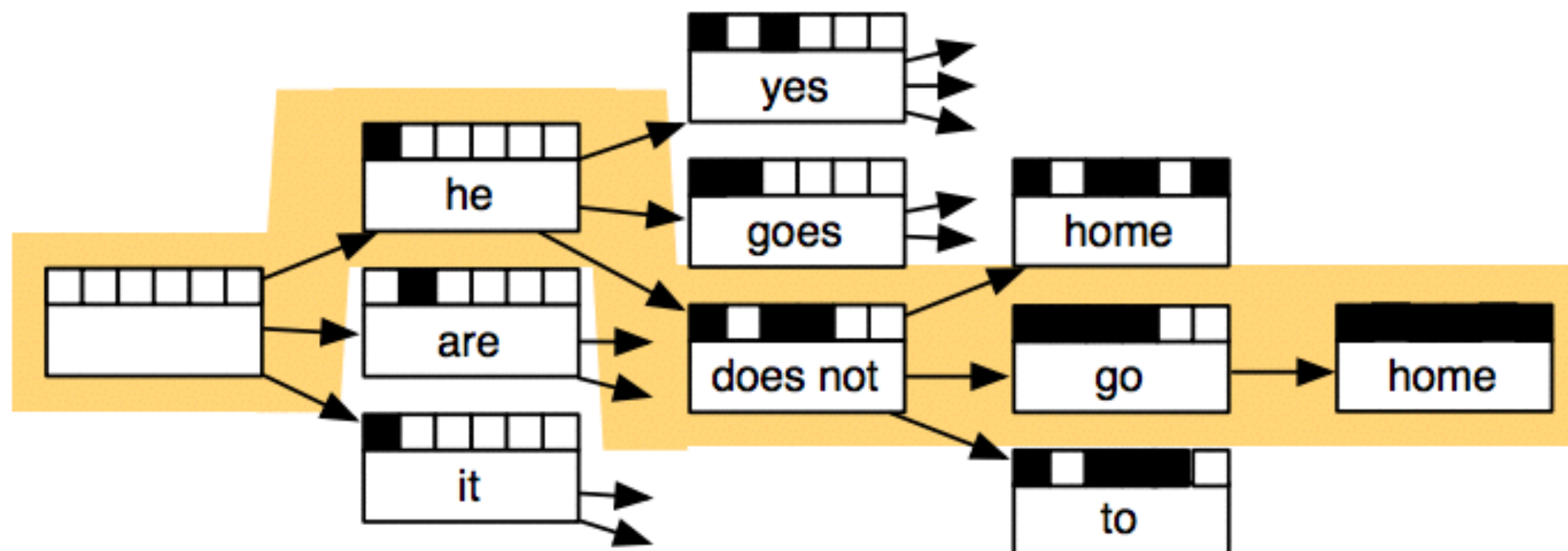
er	geht	ja	nicht	nach	hause
he	is	yes	not	after	house
it	are	is	do not	to	home
, it	goes	, of course	does not	according to	chamber
, he	go		is not	in	at home
it is		not		home	
he will be		is not		under house	
it goes		does not		return home	
he goes		do not		do not	
	is		to		
	are		following		
	is after all		not after		
	does		not to		
	not				
	is not				
	are not				
	is not a				

- Many translation options to choose from
 - in Europarl phrase table: 2727 matching phrase pairs for this sentence
 - by pruning to the top 20 per phrase, 202 translation options remain

הפתרון - Beam Search

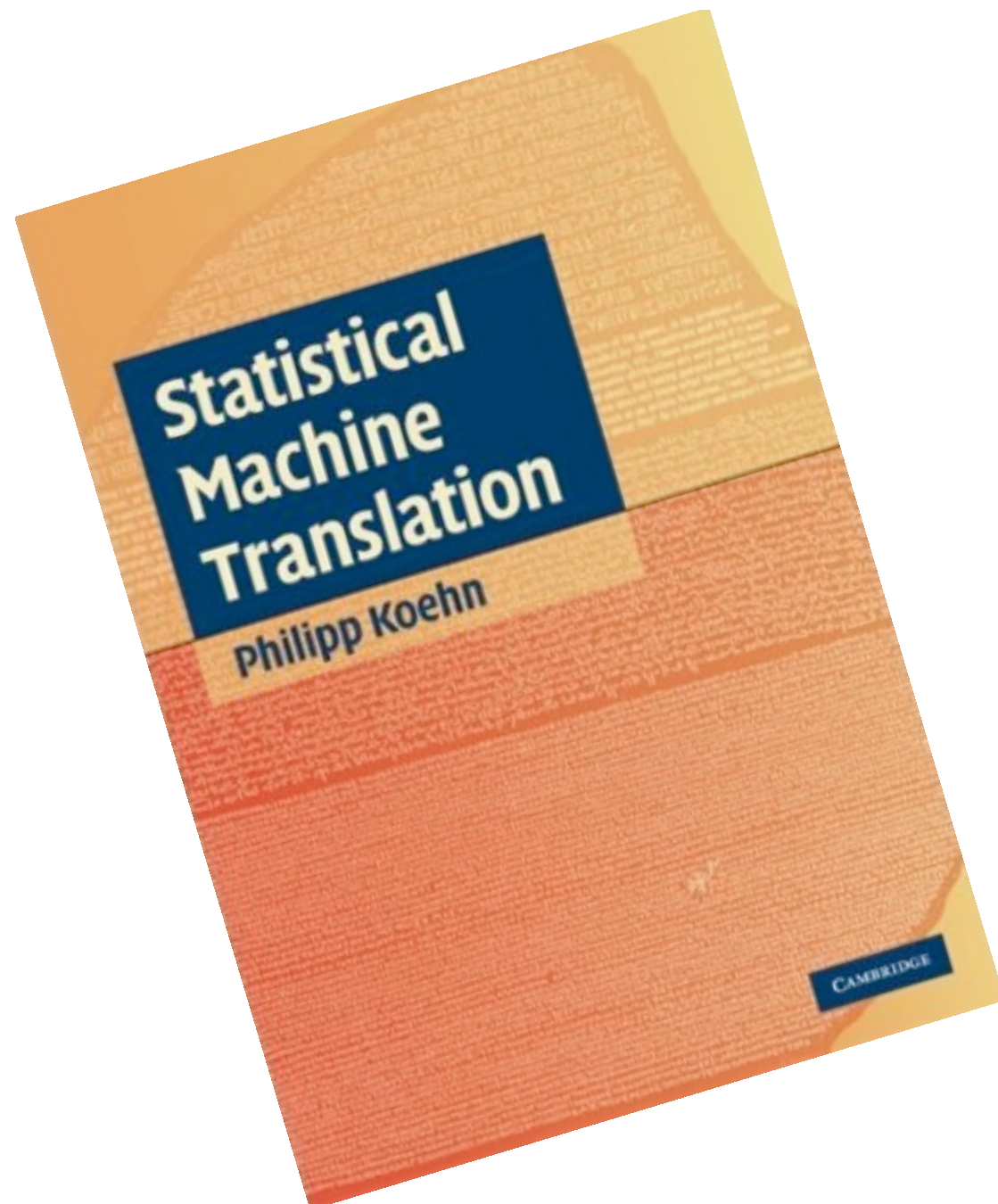


הפתרון - Beam Search



backtrack from highest scoring complete hypothesis

לסיכום



- ❖ ...לא פשוט
- ❖ מודל תרגום
- ❖ מודל שפה
- ❖ חיפוש
- ❖ תחום מחקר פעיל

שאלות?

Google translate for animals

Introducing Translate for Animals (beta): Bridging the gap between animals and humans

