

Tłumaczenie maszynowe. Zasady działania

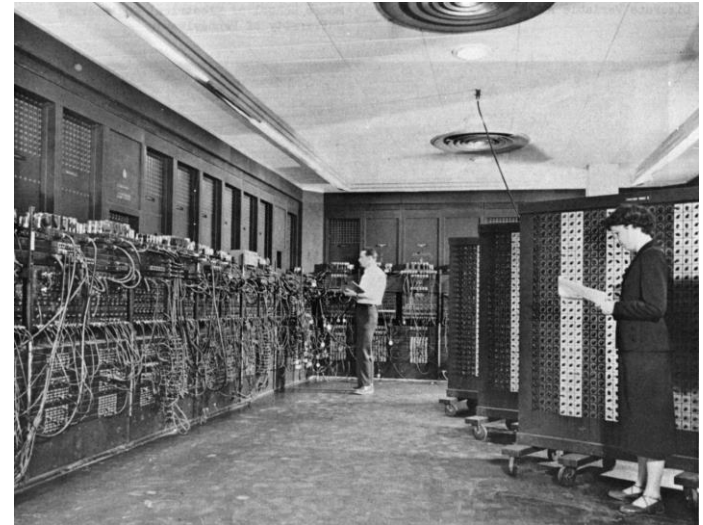
**Autorzy: Josef van Genabith (DFKI),
Krzysztof Łoboda (Uniwersytet Jagielloński)**

W prezentacji wykorzystano materiały pochodzące z warsztatów i konferencji poświęconych technologii tłumaczenia maszynowego

Zarys prezentacji:

- Uzasadnienie dla technologii MT: liczba słów, jakość, cena
- Problemy związane z MT
- Technologia MT + zawodowi tłumacze = jakość
- Działanie współczesnych silników statystycznego MT
- Dane to podstawa
- Rodzaj i jakość danych ma znaczenie!

- Europa = wielojęzyczność
- 24 języki urzędowe – olbrzymia liczba kombinacji i wolumin
- Koszty tłumaczenia
- Czy MT może być pomocne?
- Czy można mówić o jakości?



Źródło: <https://en.wikipedia.org/wiki/ENIAC#/media/File:Eniac.jpg>
License: public domain

- Języki naturalne cechuje:
 - elegancja
 - skuteczność
 - elastyczność
 - złożoność
- Jedno słowo lub zdanie może mieć wiele znaczeń (problemy semantyczne)
- Wiele sposobów wyrażenia tych samych lub podobnych znaczeń (synonimia)
- Znaczenie zakotwiczone kontekstowo
- Znaczenie dosłowne i poetyckie (metafory)
- Język a kultura (różne sposoby konceptualizacji desygnatów)
- Szyk zdania
- Morfologia
- ...

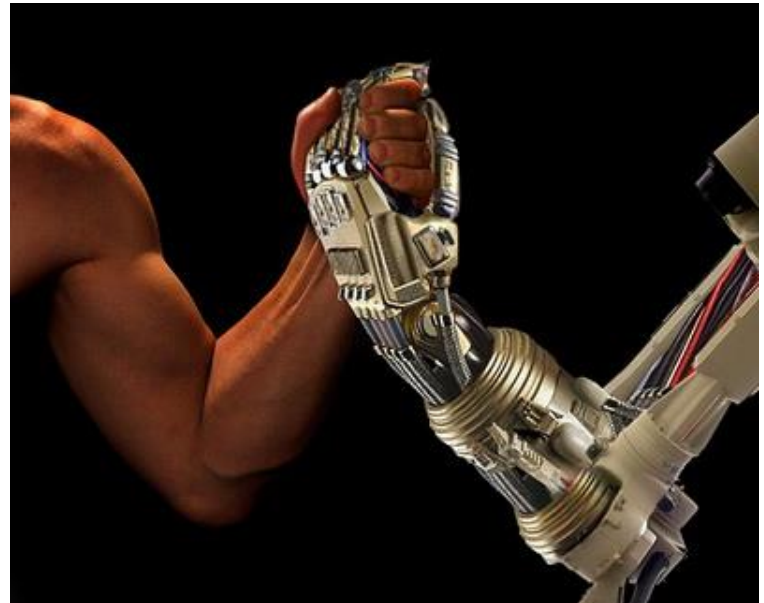
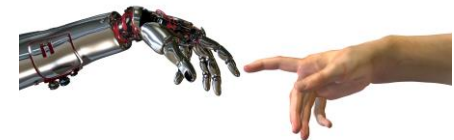
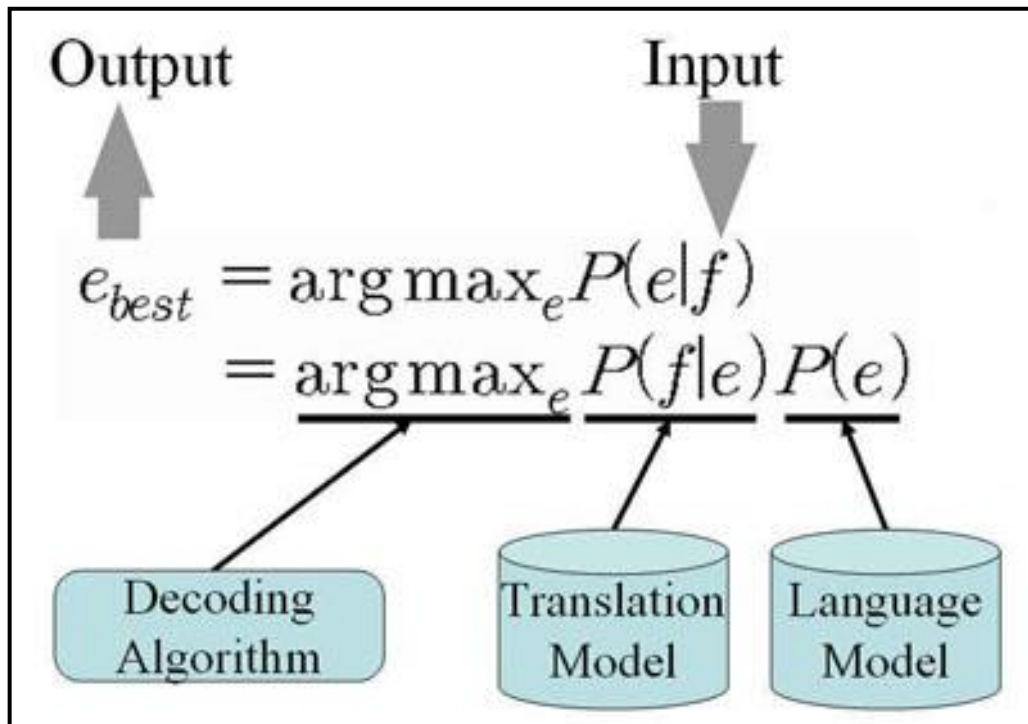


Image: <http://workingtropes.lmc.gatech.edu/wiki/index.php/File:Man-vs-machine.jpg>
License: CC BY-NC-SA 3.0

- Język jest złożony (tłumaczenie także)
- Niemożność stworzenia dokładnego modelu matematycznego
- Próbowano podejścia klasycznego, opartego na regułach (RBMT); za mała skuteczność
- Dalsze działania?
- Systemy uczące się (machine learning)
 - Uczenie się z **danych** \Rightarrow dane są najważniejsze
 - przybliżone (zgrubne) rozwiązania \Rightarrow potrzebna pomoc
 - tłumacze zawodowi
 - postędyca
 - tłumaczenie maszynowe \neq automatyczne





- Zamiast matematyki
- Obrazowe wyjaśnienie statystycznego MT...
- **Dane** to podstawa

Silnik statystycznego MT jest doskonałony dzięki 2 rodzajom danych:

- Przekłady tłumaczy zawodowych
- Teksty w językach docelowych
- Im więcej danych, tym lepiej.
- Potrzebne także **odpowiednie dane**

GERMAN

Einleitung

I. Von dem Unterschiede der reinen und empirischen Erkenntnis

Daß alle unsere Erkenntnis mit der Erfahrung anfangt, daran ist gar kein Zweifel; denn wodurch sollte das Erkenntnisvermögen sonst zur Ausübung erweckt werden, geschähe es nicht durch Gegenstände, die unsere Sinne rühren und teils von selbst Vorstellungen bewirken, teils unsere Verstandstätigkeit in Bewegung bringen, diese zu vergleichen, sie zu verknüpfen oder zu trennen, und so den rohen Stoff sinnlicher Eindrücke zu einer Erkenntnis der Gegenstände zu verarbeiten, die Erfahrung heißt? Der Zeit nach geht also keine Erkenntnis in uns vor der Erfahrung vorher, und mit dieser fängt alle an.

ENGLISH

Introduction

I. Of the difference between Pure and Empirical Knowledge

That all our knowledge begins with experience there can be no doubt. For how is it possible that the faculty of cognition should be awakened into exercise otherwise than by means of objects which affect our senses, and partly of themselves produce representations, partly rouse our powers of understanding into activity, to compare these, and so to convert the raw material of our sensuous impressions into a knowledge of objects, which is called experience? In respect of time, therefore, no knowledge of ours is antecedent to experience, but begins with it.

FRENCH

Introduction

I. De la différence de la connaissance pure et de la connaissance empirique.

Que toute notre connaissance commence avec l'expérience, cela ne soulève aucun doute. En effet, par quoi notre pouvoir de connaître pourrait-il être éveillé et mis en action, si ce n'est par des objets qui frappent nos sens et qui, d'une part, produisent par eux-mêmes des représentations et, d'autre part, mettent en mouvement notre faculté intellectuelle, afin qu'elle compare, lie ou sépare ces représentations, et travaille ainsi la matière brute des impressions sensibles pour en tirer une connaissance des objets, celle qu'on nomme l'expérience? Ainsi, chronologiquement, aucune connaissance ne précède en nous l'expérience et c'est avec elle que toutes commencent.

- Odpowiadające są sobie zdania: **dopasowanie segmentów/zdań (alignment)**
- Odpowiadające są sobie słowa: **dopasowanie słów + prawdopodobieństwo tłumaczenia**
- Jaki jest język docelowy: **model języka**

GERMAN

Einleitung

I. Von dem Unterschiede der reinen und empirischen Erkenntnis

Daß alle unsere Erkenntnis mit der Erfahrung anfangt, daran ist gar kein Zweifel; denn wodurch sollte das Erkenntnisvermögen sonst zur Ausübung erweckt werden, geschähe es nicht durch Gegenstände, die unsere Sinne rühren und teils von selbst Vorstellungen bewirken, teils unsere Verstandstätigkeit in Bewegung bringen, diese zu vergleichen, sie zu verknüpfen oder zu trennen, und so den rohen Stoff sinnlicher Eindrücke zu einer Erkenntnis der Gegenstände zu verarbeiten, die Erfahrung heißt? Der Zeit nach geht also keine Erkenntnis in uns vor der Erfahrung vorher, und mit dieser fängt alle an.

ENGLISH

Introduction

I. Of the difference between Pure and Empirical Knowledge

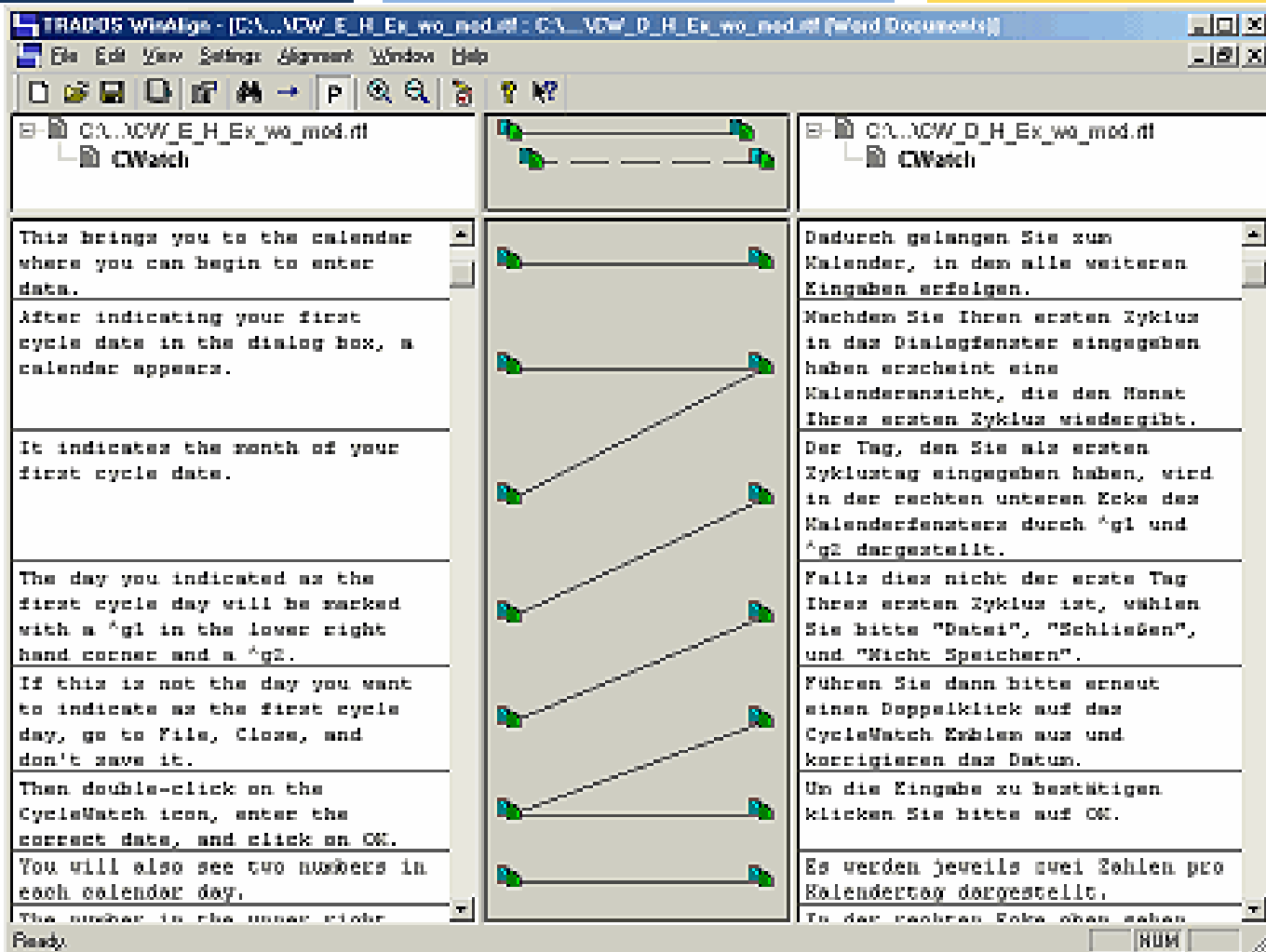
That all our knowledge begins with experience there can be no doubt. For how is it possible that the faculty of cognition should be awakened into exercise otherwise than by means of objects which affect our senses, and partly of themselves produce representations, partly rouse our powers of understanding into activity, to compare to connect, or to separate these, and so to convert the raw material of our sensuous impressions into a knowledge of objects, which is called experience? In respect of time, therefore, no knowledge of ours is antecedent to experience, but begins with it.

FRENCH

Introduction

I. De la différence de la connaissance pure et de la connaissance empirique.

Que toute notre connaissance commence avec l'expérience, cela ne soulève aucun doute. En effet, par quoi notre pouvoir de connaître pourrait-il être éveillé et mis en action, si ce n'est par des objets qui frappent nos sens et qui, d'une part, produisent par eux-mêmes des représentations et, d'autre part, mettent en mouvement notre faculté intellectuelle, afin qu'elle compare, lie ou sépare ces représentations, et travaille ainsi la matière brute des impressions sensibles pour en tirer une connaissance des objets, celle qu'on nomme l'expérience? Ainsi, chronologiquement, aucune connaissance ne précède en nous l'expérience et c'est avec elle que toutes commencent.



TRADOS WinAlign - [C:\...NOW_E_H_Ex_wo_mod.rtf : C:\...NOW_D_H_Ex_wo_mod.rtf (Inford Documents)]

File Edit View Settings Segment Window Help

C:\...NOW_E_H_Ex_wo_mod.rtf
 CycleMatch

This brings you to the calendar where you can begin to enter data.
 After indicating your first cycle date in the dialog box, a calendar appears.
 It indicates the month of your first cycle date.
 The day you indicated as the first cycle day will be marked with a ^g1 in the lower right hand corner and a ^g2.
 If this is not the day you want to indicate as the first cycle day, go to File, Close, and don't save it.
 Then double-click on the CycleMatch icon, enter the correct date, and click on OK.
 You will also see two numbers in each calendar day.
 The number in the upper right

C:\...NOW_D_H_Ex_wo_mod.rtf
 CycleMatch

Dadurch gelangen Sie zum Kalender, in dem alle weiteren Eingaben erfolgen.
 Nachdem Sie Ihren ersten Zyklus in das Dialogfenster eingegeben haben erscheint eine Kalenderansicht, die den Monat Ihres ersten Zyklus wiedergibt.
 Der Tag, den Sie als ersten Zyklustag eingegeben haben, wird in der rechten unteren Ecke des Kalenderfensters durch ^g1 und ^g2 dargestellt.
 Falls dies nicht der erste Tag Ihres ersten Zyklus ist, wählen Sie bitte "Datei", "Schließen", und "Nicht Speichern".
 Führen Sie dann bitte erneut einen Doppelklick auf das CycleMatch Icon aus und korrigieren das Datum.
 Um die Eingabe zu bestätigen klicken Sie bitte auf OK.
 Es werden jeweils zwei Zahlen pro Kalendertag dargestellt.
 In der rechten Ecke oben sehen

Ready.



		CLASSIC SOUPS		Sm.	Lg.
清 燉 雞	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75	
雞 飯	58.	Chicken Rice Soup	1.85	3.25	
雞 麵	59.	Chicken Noodle Soup	1.85	3.25	
廣 東 雲 吞	60.	Cantonese Wonton Soup.....	1.50	2.75	
蕃 茄 蛋	61.	Tomato Clear Egg Drop Soup	1.65	2.95	
雲 吞	62.	Regular Wonton Soup	1.10	2.10	
酸 辣	63.	Hot & Sour Soup	1.10	2.10	
蛋	64.	Egg Drop Soup.....	1.10	2.10	
雲 吞	65.	Egg Drop Wonton Mix.....	1.10	2.10	
豆 腐 菜	66.	Tofu Vegetable Soup	NA	3.50	
雞 玉 米	67.	Chicken Corn Cream Soup	NA	3.50	
蟹 肉 玉 米	68.	Crab Meat Corn Cream Soup.....	NA	3.50	
海 鮮	69.	Seafood Soup.....	NA	3.50	



		CLASSIC SOUPS		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	61.	Tomato Clear Egg Drop Soup	1.65	2.95		
雲吞	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		
雲吞	吞	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		



- Moduł dopasowywania na poziomie słów wie sporo na temat zupek chińskich
- Ale niewiele więcej...

- Wie tylko tyle, co znalazł w danych treningowych
- Trochę jak u ludzi...
- „Takie widzi świata koło” ...

- Czy na podstawie danych o tłumaczeniu na poziomie słów można stworzyć słownik tłumaczeń?
- Tak, to możliwe...



I love the boy.
J'aime le garçon.
I love the dog.
J'aime le chien.
They love the dog.
Ils aiment le chien.
They talk to the girl.
Ils parlent à la fille.
They talk to the dog.
Ils parlent au chien.
I talk to the mother.
Je parle à la mère.

Aligned Data



I love the boy.
J'aime le garçon.

I love the dog.
J'aime le chien.

They love the dog.
Ils aiment le chien.

They talk to the girl.
Ils parlent à la fille.

They talk to the dog.
Ils parlent au chien.

I talk to the mother.
Je parle à la mère.

Aligned Data



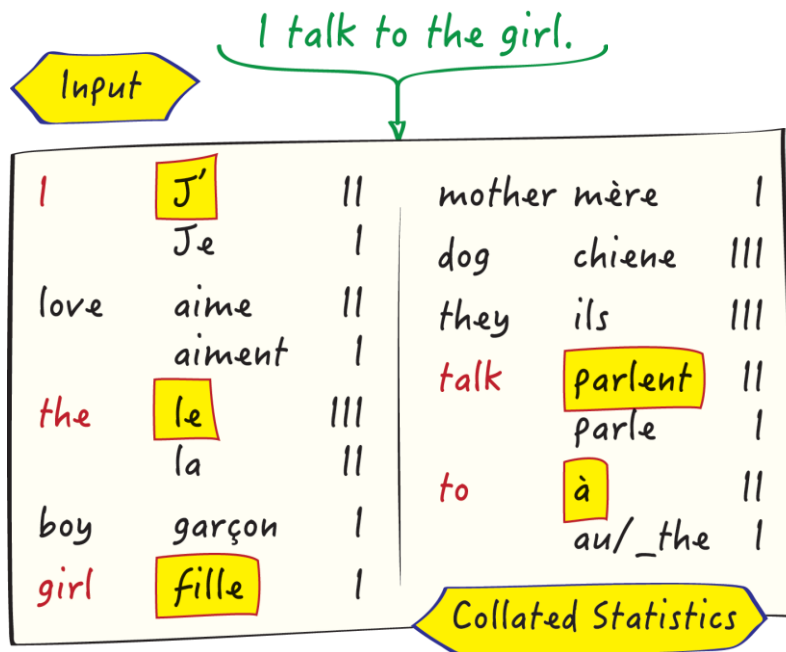
I	J'		mother	mère	
	Je		dog	chiene	
love	aime		they	ils	
	aiment		talk	parlent	
the	le			parle	
	la		to	à	
boy	garçon			au/_the	
girl	fille				

Collated Statistics



I love the boy.
 J'aime le garçon.
 I love the dog.
 J'aime le chien.
 They love the dog.
 Ils aiment le chien.
 They talk to the girl.
 Ils parlent à la fille.
 They talk to the dog.
 Ils parlent au chien.
 I talk to the mother.
 Je parle à la mère.

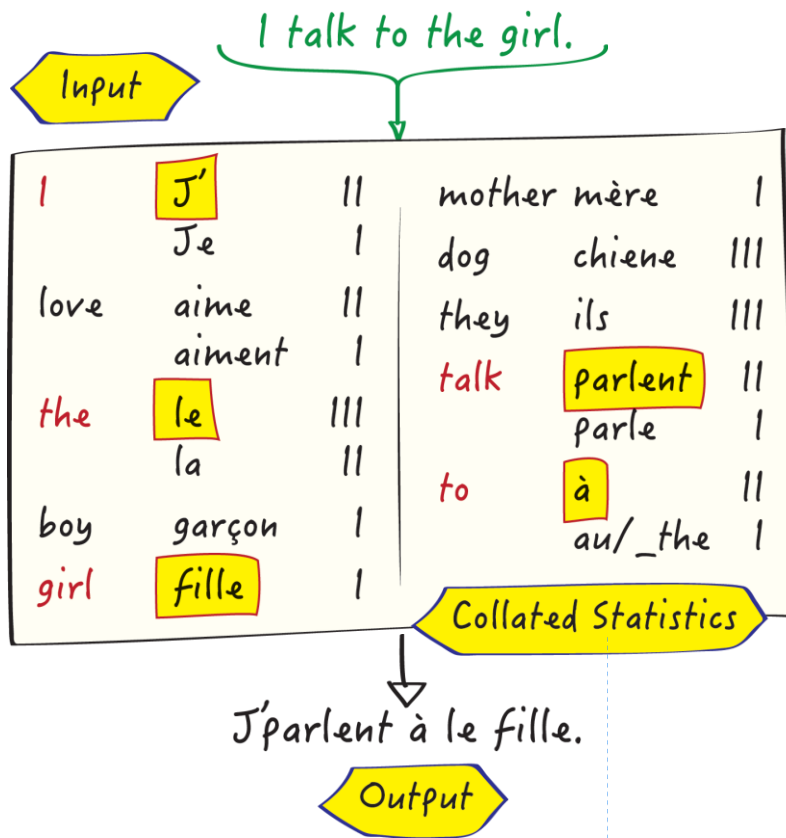
Aligned Data





I love the boy.
J'aime le garçon.
I love the dog.
J'aime le chien.
They love the dog.
Ils aiment le chien.
They talk to the girl.
Ils parlent à la fille.
They talk to the dog.
Ils parlent au chien.
I talk to the mother.
Je parle à la mère.

Aligned Data





I love the boy.
 J'aime le garçon.
 I love the dog.
 J'aime le chien.
 They love the dog.
 Ils aiment le chien.
 They talk to the girl.
 Ils parlent à la fille.
 They talk to the dog.
 Ils parlent au chien.
 I talk to the mother.
 Je parle à la mère.



Aligned Data

I	talk	to	the	girl
J'	parlent	au	le	fille
2/3	2/3	2/3	3/5	1/1
Je	parle	à	la	fille
1/3	1/3	1/3	2/5	1/1

Jak wybrać?



I love the boy.
J'aime le garçon.
I love the dog.
J'aime le chien.
They love the dog.
Ils aiment le chien.
They talk to the girl.
Ils parlent à la fille.
They talk to the dog.
Ils parlent au chien.
I talk to the mother.
Je parle à la mère.



Aligned Data

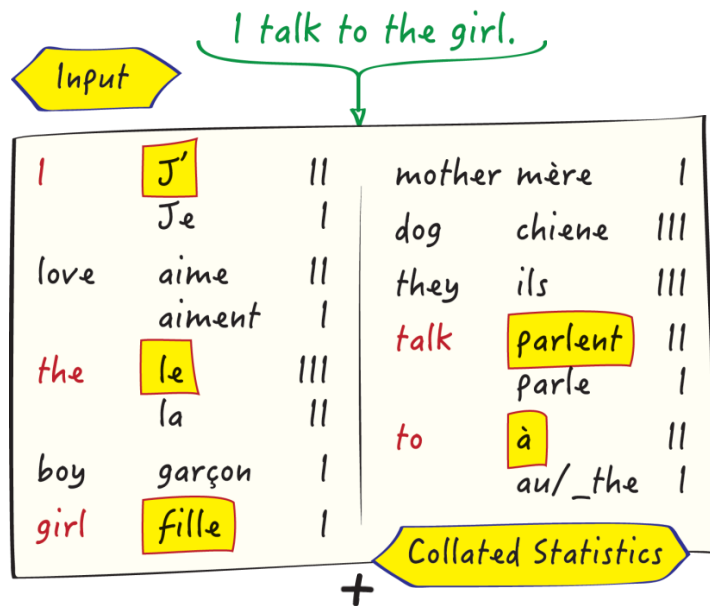
Model:

- Co jest poprawne w jęz. docelowym?
- Możliwy i niemożliwy szyk zdania... gramatyka
- Zdobywana poprzez uczenie się z danych...
 - *Je parle* - poprawnie...
 - *J' parlent* - niepoprawnie...
 - *la fille* - poprawnie...
 - *le fille* - niepoprawnie...
- *Je parle à la fille >> J' parlent à le fille*



I love the boy.
J'aime le garçon.
I love the dog.
J'aime le chien.
They love the dog.
Ils aiment le chien.
They talk to the girl.
Ils parlent à la fille.
They talk to the dog.
Ils parlent au chien.
I talk to the mother.
Je parle à la mère.

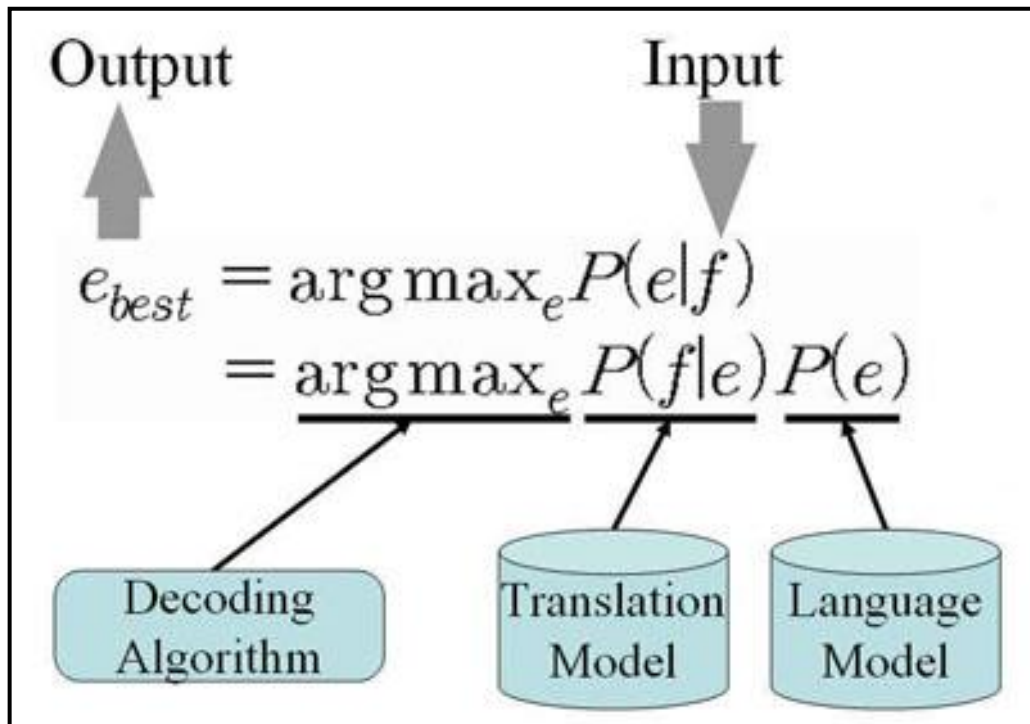
Aligned Data



Language Model

J'parlent à le fille.

Output



- Zamiast matematyki
- Obrazowe wyjaśnienie statystycznego MT...
- **Dane** to podstawa



- Dotychczas tłumaczyliśmy pojedyncze słowa
- Utrata kontekstu, np. uzgodnienie rodzajnika (*le fille* ...)
- Do pewnego stopnia “korekta” poprzez model języka

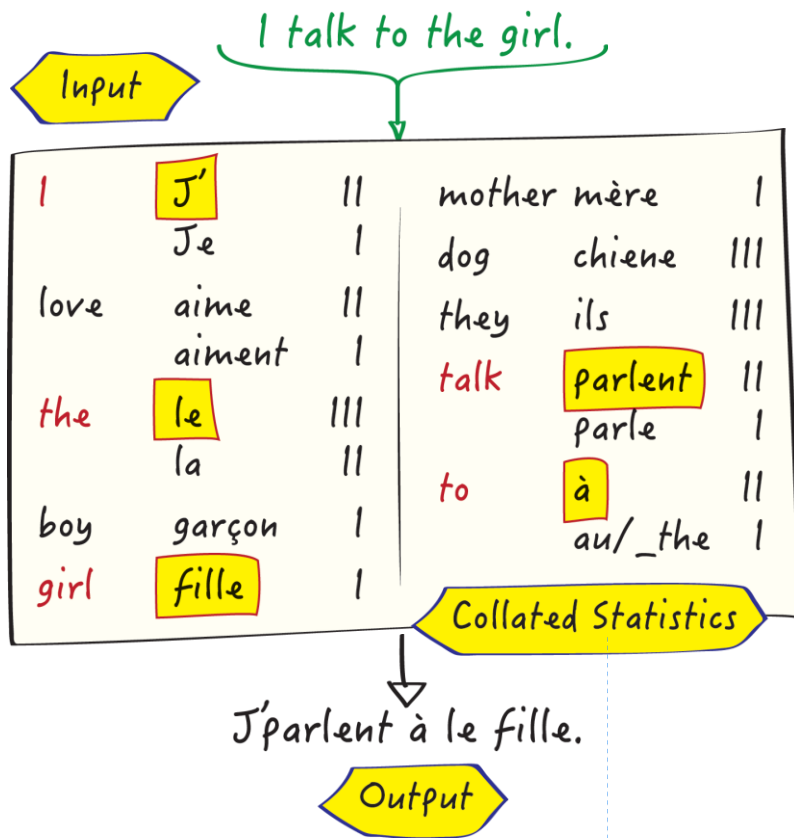
- Lepszy model:
- Wyjście poza tłumaczenie pojedynczych słów
- Tłumaczenie całych wyrażień:

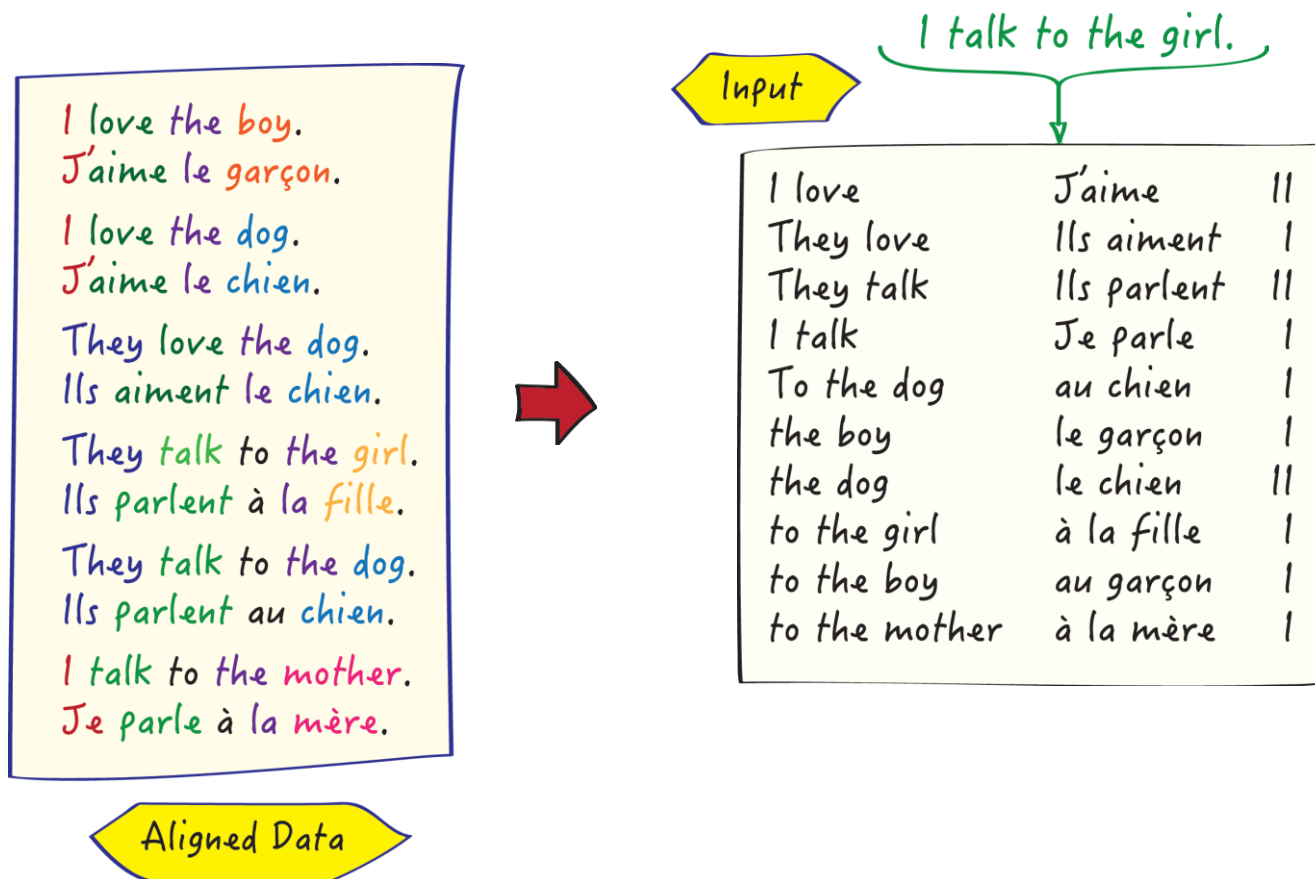
– *the girl* : *la fille*
– *to the girl* : *a la fille*
– *I talk* : *Je parle*

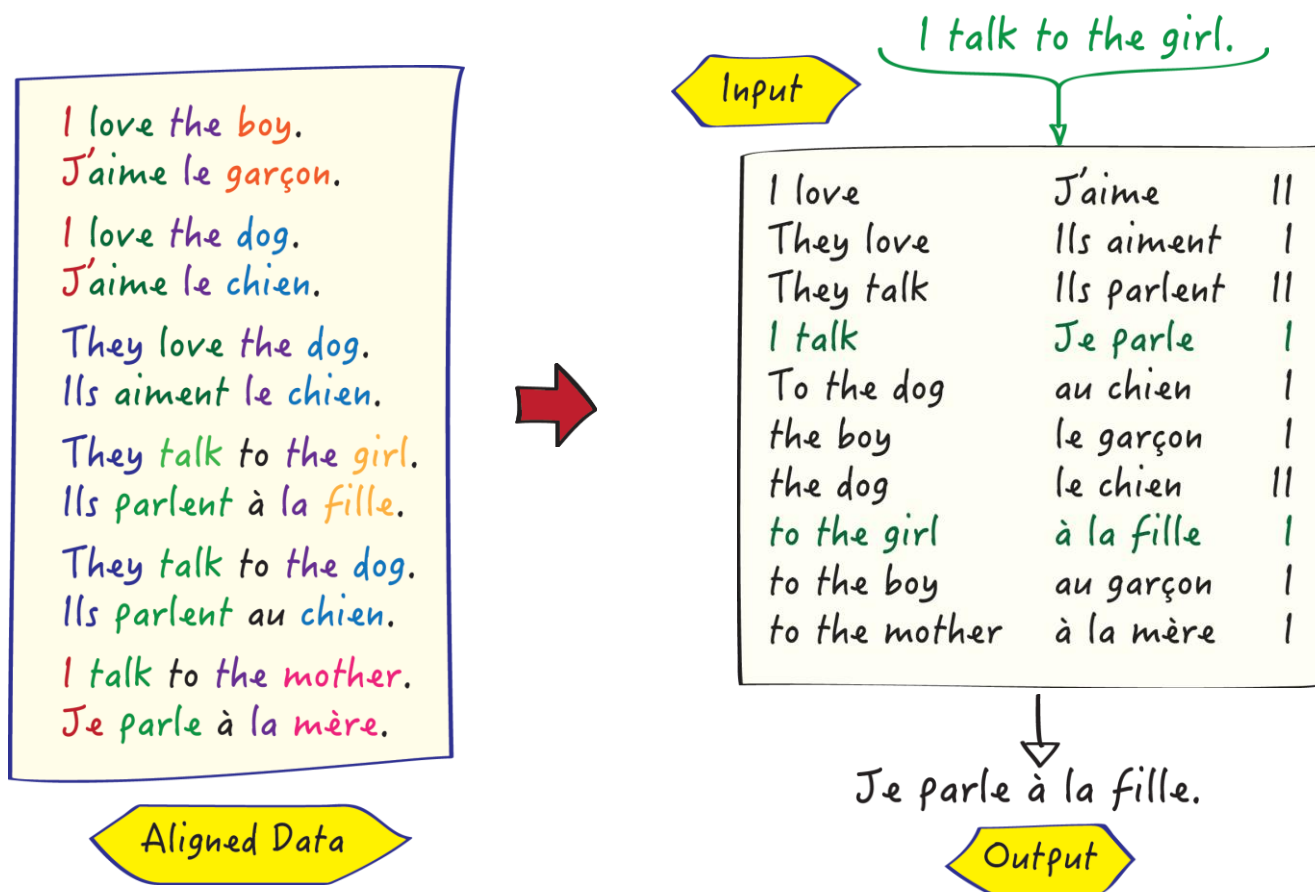


I love the boy.
J'aime le garçon.
I love the dog.
J'aime le chien.
They love the dog.
Ils aiment le chien.
They talk to the girl.
Ils parlent à la fille.
They talk to the dog.
Ils parlent au chien.
I talk to the mother.
Je parle à la mère.

Aligned Data



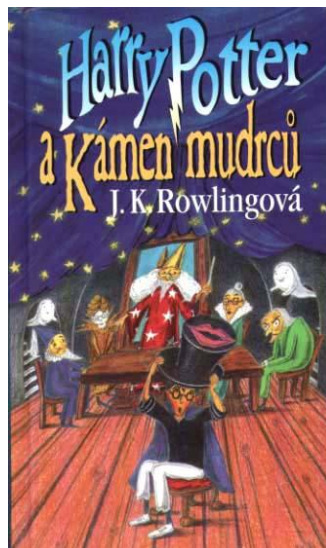
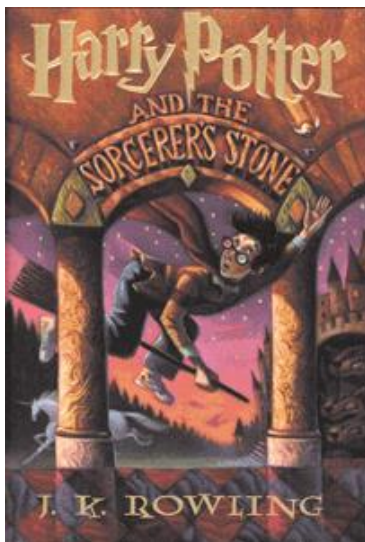




- Znacznie lepsze niż SMT oparte na pojedynczych słowach
- Standardowa technologia stosowana przez Google, Microsoft, Baidu
- Moses – otwarty silnik PB-SMT (open source)
- Najbardziej powszechny system SMT
- Badania finansowane przez KE
- Wykorzystywany przez DCT - MT@EC

MOSES  CORE

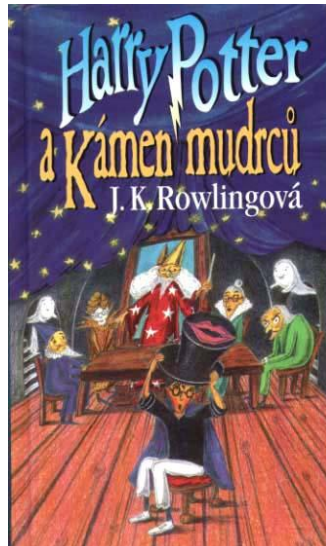
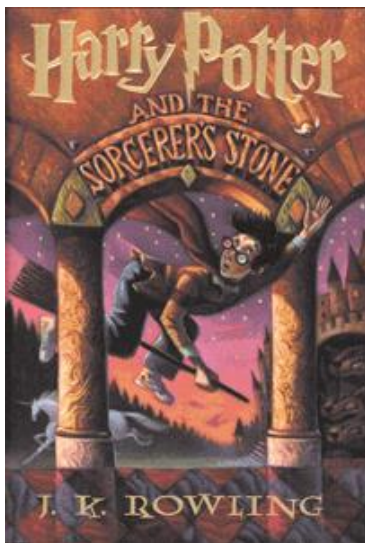
- Systemy tłumaczenia maszynowego są całkowicie zależne od danych
- System SMT uczy się transprzetwarzania (transprocessing) z danych
- Dane
 - istniejące przekłady (dane dwujęzyczne)
 - dane monolingwalne (teksty w jęz. docelowym)
 - słowniki, terminologia, toponimy, nazwy własne
- Silniki SMT są dobre w dziedzinach, z których mają mnóstwo danych odpowiedniej jakości



"Good," said Wizard Wink. "We'll start with Numbers."
"Oh no!" groaned Max and Mick.
The wizard boys didn't like school much. They wanted to play.
Wizard Wink asked them the first question.
"If I had three crystal balls in one hand . . ." he said, "and four
in the other . . . what would I have?"
"BIG HANDS!" shouted Max and Mick.
"Very funny," said Wizard Wink.
"Oooh! I know. I know!" said Zoe. "Seven!"
"Right!" said Wizard Wink.
"Pool!" said Melissa. "That was easy-peasy!"



MOSES  CORE



Protect - Personal Information CIVMEANS7

Legal Aid Agency
Financial Assessment
for Family Mediation
Provider reference/case code: MED121GBHST/1/451
This form must be completed in ink.

Applicant's Details

Surname: Mr/Ms/Miss/Ms _____ First name(s) _____
Surname at birth if different: _____ Date of birth: __/__/____
Address: _____ Postcode: _____
National Insurance number: _____
Job: _____

Financial Eligibility

1. The client has a partner whose means are to be aggregated:

Yes Please provide details of both client's and partner's means.
 No Please provide details of both client's means only.

2. The case is about ownership or possession of assets and / or financial provision:

Yes Go to question 3.
 No Go directly to Part B Capital.

3. The client's assets (held in sole name or jointly held) have been claimed by the opponent:

Yes Please complete Part A Capital - Subject matter of dispute.
 No Go directly to Part B Capital.

The subject matter of dispute disregard only applies to assets that are specially claimed by the opponent. All assets that have not been specifically claimed by the opponent must be included in Part B Capital.

CIVMEANS7 Page 1 Version B April 2013 © Crown Copyright

MOSES  CORE

- System CEF.AT także potrzebuje odpowiedniej jakości danych
- Rządy, samorządy, administracja, instytucje publiczne, organizacje pozarządowe
- Celem CEF jest świadczenie usług językowych pomocnych w kontaktach z obywatelami z danego państwa członkowskiego, z UE i innych podmiotów obsługiwanych przez administrację publiczną

- Pomagając w znalezieniu odpowiednich danych,
- można przyczynić się do pomyślnej realizacji projektu CEF.AT, a tym samym do zapewnienia
 - usług dla Państwa instytucji
 - usług dla obywateli UE
 - wsparcia wielojęzyczności.