Multimedia Department

Statystyczne tłumaczenie mowy polskiej

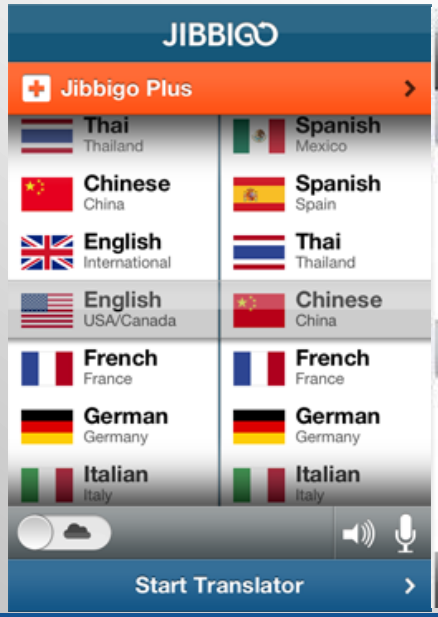
Krzysztof Marasek
Krzysztof Wołk
Łukasz Brocki
Danijel Korzinek
PJATK

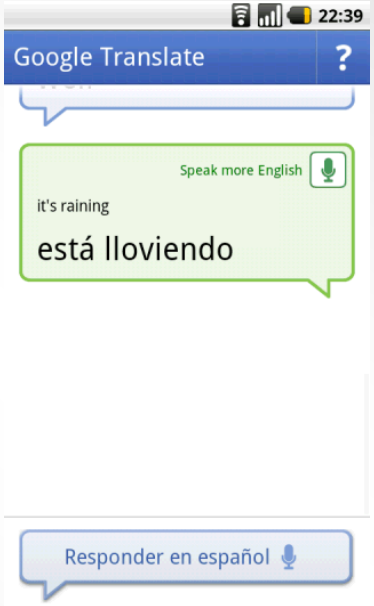
Plan prezentacji

- ✓ Wprowadzenie
- ✓ Komponenty systemu tłumaczącego
- ✓ Aktualne działania
- ✓ Nowe metody

Dlaczego automatyczne tłumaczenie mowy?

- ✓ **Czynniki ludzkie**
  - ✓ Potrzeby komunikacyjne w kurczącym się geograficznie świecie, komercyjne zastosowania SST
  - ✓ Urządzenia mobilne, gadżeciarstwo
- ✓ **Czynniki lingwistyczne**
  - ✓ Struktura mowy jest prostsza niż języka pisanego
  - ✓ Ograniczone słownictwo
  - ✓ (często) prostsza forma gramatyczna





Multimedia Department

Dlaczego automatyczne tłumaczenie mowy?

- ✓ **Czynniki techniczne**
  - ✓ Choć nadal dalekie od perfekcji, metody maszynowego tłumaczenia dokonały w ostatnich latach wyraźnego postępu

    NIST MT workshops (od 2002)

    IWSLT workshops (od 2002)

    TC-STAR workshop (2005-2006)

    ACL/NAACL Shared Tasks (2005-2006)

TASK NAME	CATEGORY	SCOPE AND SCENARIOS
TIMIT	ASR	English phonetic recognition; small vocabulary
WSJ 0&1	ASR	Mid-to-large vocabulary; dictation speech
AURORA	ASR	Small vocabulary, under noisy acoustic environments
DARPA EARS	ASR	Large vocabulary, broadcast, and conversational speech
NIST MT OPEN EVAL	MT	Large scale MT, newswire, and Web text
WMT (EUROPARL/EUROMATRIX)	MT	Large scale MT, newswire, and political text
C-STAR	ST	Limited domain spontaneous speech
DARPA TRANSTAC	ST	Limited domain spontaneous dialog
IWSLT	ST	Limited domain dialog and unlimited free-style talk
TC-STAR	ST	Broadcast speech, political speech
DARPA GALE	ST	Broadcast speech, conversational speech

He, Deng, 11

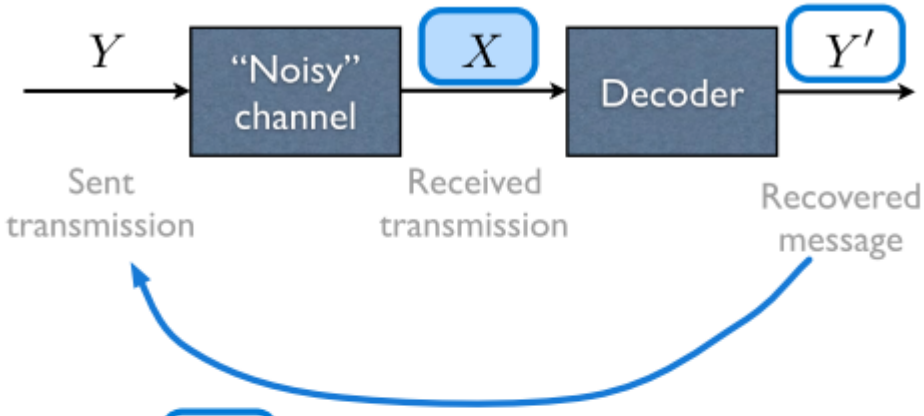
    - ✓ Czynniki sukcesu to:
      - Coraz silniejsze komputery (trening na klastrach PC, dużo RAM)
      - Rozwój metod statystycznych i bazujących na danych (*data driven*)
      - Odpowiednie zasoby lingwistyczne (korpusy bilingualne)
      - Rozwój metod oceny jakości tłumaczenia (miary błędów)
      - Stopniowe komplikowanie zadań
      - Zaangażowanie wielkich firm: Google, Microsoft, IBM, itd..

Lazzari, 06

  - ✓ Rozwój interfejsów głosowych
    - ✓ Rozpoznawanie mowy działa coraz lepiej
    - ✓ Mowę syntetyczną coraz trudniej odróżnić od naturalnej

Podstawy

- ✓ Kanał komunikacyjny z szumem i twierdzenie Bayesa



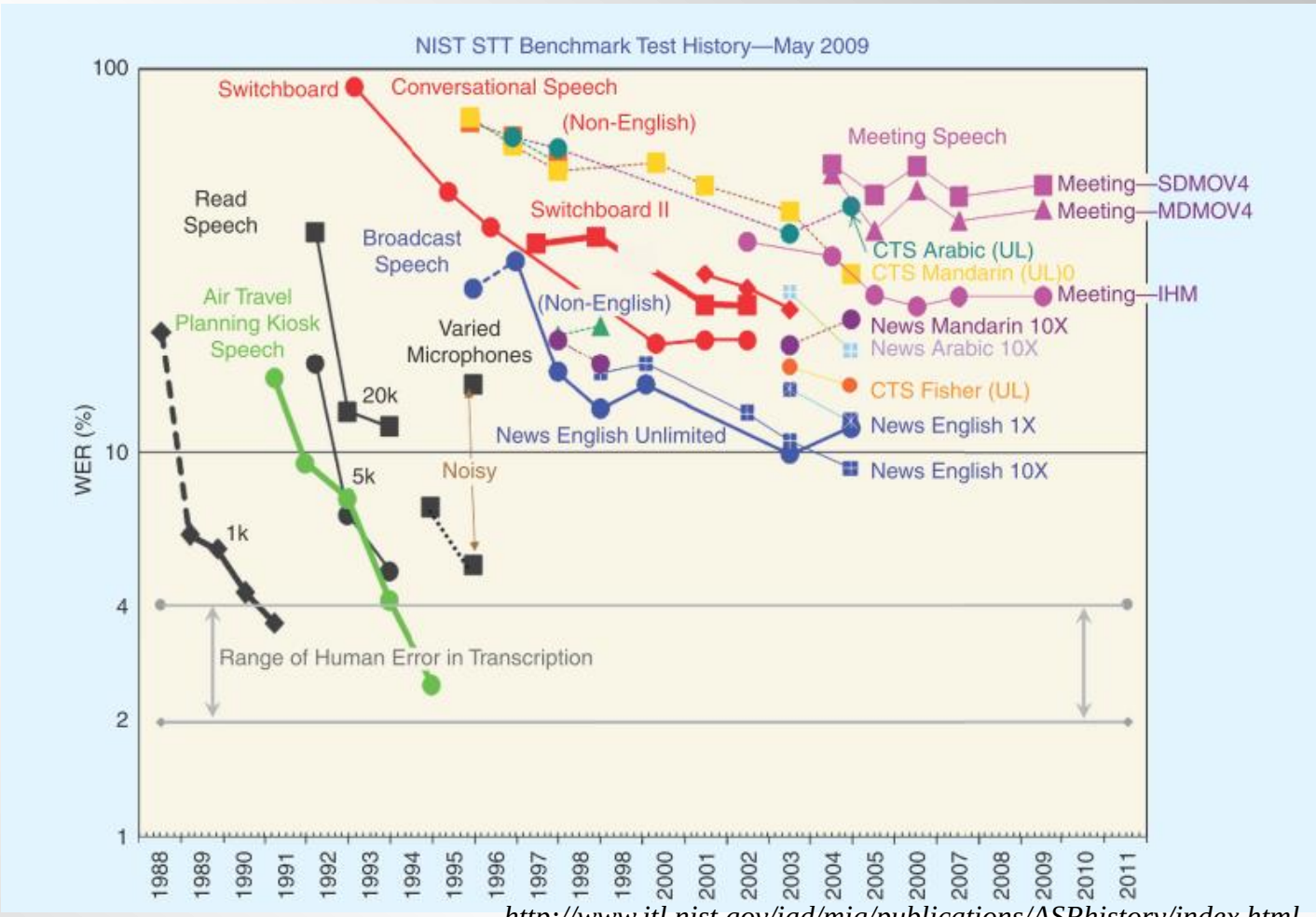
$$y' = \arg\max_y p(y|x)$$

$$= \arg\max_y \frac{p(x|y)p(y)}{p(x)}$$

$$= \arg\max_y p(x|y)p(y)$$

Dyer, 12

Jakość rozpoznawania mowy



http://www.itl.nist.gov/iad/mig/publications/ASRhistory/index.html

- ✓ **ASR da się używać!**
- ✓ ASR działa najlepiej dla ściśle określonego mówcy, im mniejszy słownik tym lepiej, angielski lepiej niż inne języki
- ✓ Stopa błędów rozpoznawania mowy spontanicznej jest co najmniej dwukrotnie większa niż dla czytania, stopa błędów jest wysoka dla konwersacji wielu mówców w trudnym akustycznie środowisku

ASR dla języka polskiego

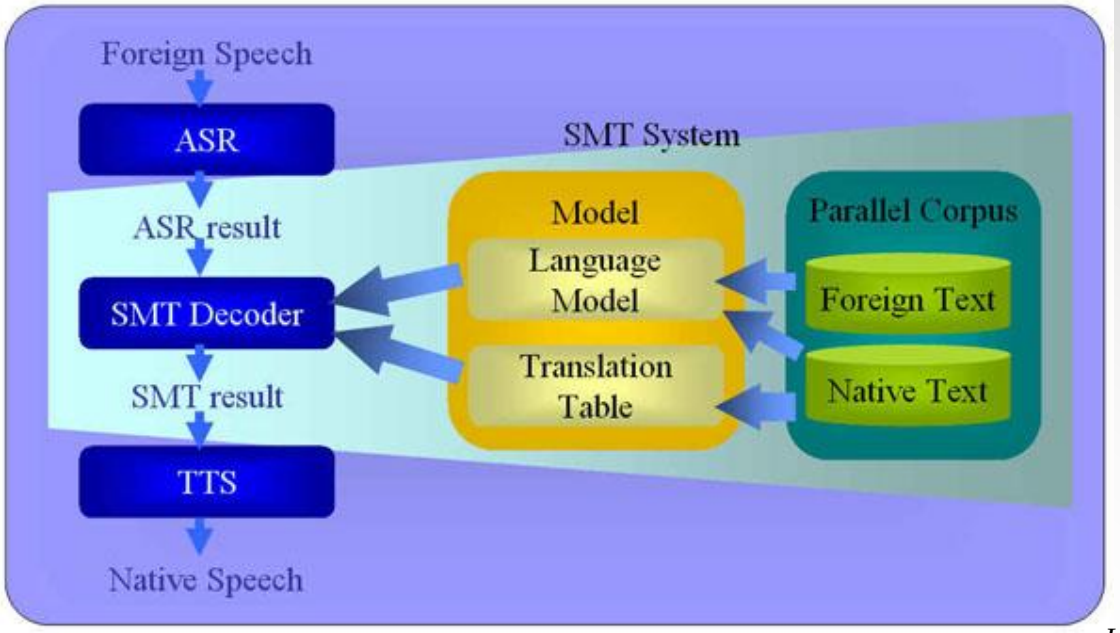
- ✓ Mały wybór produktów komercyjnych
- ✓ Kaldi jako platforma eksperymwntów
  - ✓ Speech server
  - ✓ Speaker dependent models, fMMLR, MMI,…..
- ✓ Własny dekoder LSTM
  - ✓ LSTM w AM i LM
  - ✓ zoptymalizowany - praca w czasie rzeczywistym na słabym laptopie
- ✓ Własny system konwersji zapisu ortograficznego na fonetyczny
  - ✓ Warianty wymowy
- ✓ Spikes – sparse signal representation
  - ✓ Kodowanie sygnału mowy zbliżone do słuchu
  - ✓ Scattering wavelets (Mallat, 11)
- ✓ Zasoby językowe i mowy dla języka polskiego
- ✓ DNN-BLSTM
- ✓ Julius

Domain	WER	Vocabulary size
TV news	15,72	42k
Polish Senate	19,6	87k
Lectures	27,75	210k

Method	WER
Inital trigram	37.37
+LDA/MLLT	34.37
+MMI	32.01
+MPE	32.55
+SAT(fMLLR)	33.51
+MMI	31.81
+fMMI	29.85
+BMMI	29.69
+SGMM	32.39

Komponenty systemu tłumaczącego

- ✓ Rozpoznawanie mowy
- ✓ Tłumaczenie wykorzystujące modele statystyczne

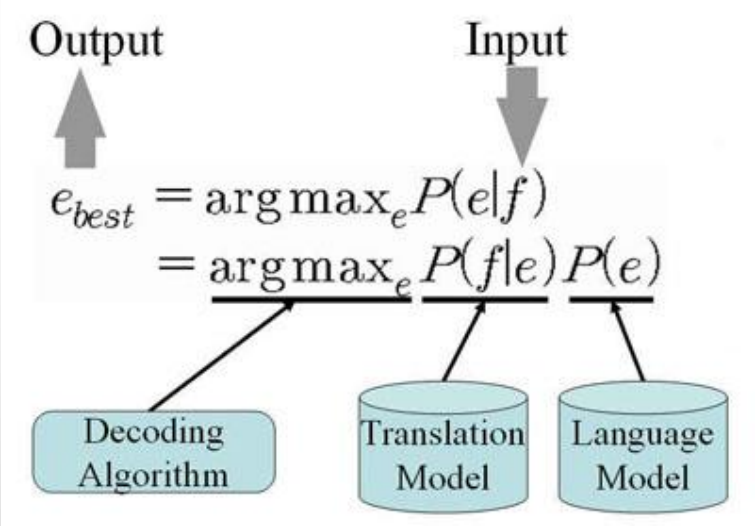


Intelligent Software Lab, 06

- ✓ Synteza mowy

Tłumaczenie statystyczne

- ✓ **Input :** SMT system otrzymuje zdanie do przetłumaczenia
- ✓ **Output :** SMT system generuje zdanie które jest tłumaczeniem zdania wejściowego
- ✓ **Language Model (model języka)** jest modelem określającym prawdopodobieństwo dowolnej sekwencji słów w danym języku
- ✓ **Translation Model (model tłumaczenia)** określa prawdopodobieństwa par tłumaczeń
- ✓ **Decoding Algorithm (dekodowanie**) to algorytm przeszukiwania grafu wyznaczający optymalne przejście przez graf słów

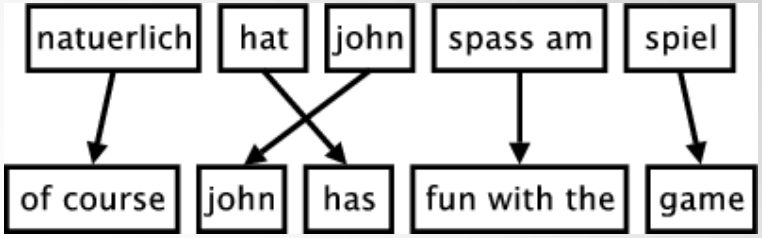


Tłumaczenie wykorzystujące frazy

- ✓ Dopasowany korpus, coraz popularniejsze, coraz bardziej akceptowalne wyniki
  - ✓ Coraz rzadziej systemy regułowe, interlingua, modelowanie
  - ✓ Phrase based translation zamiast analizy składni, parsowania zdań
  - ✓ Każda fraza jest tłumaczona na frazę i wynik może mieć zmienioną kolejność
  - ✓ Model matematyczny

$\text{argmax}_{\mathbf{e}}\ p(\mathbf{e}|\mathbf{f}) = \text{argmax}_{\mathbf{e}}\ p(\mathbf{f}|\mathbf{e})\ p(\mathbf{e})$

model języka **e** i model tłumaczenia $p(\mathbf{f}|\mathbf{e})$



  - ✓ Podczas dekodowania sekwencja słów F jest dzielona na sekwencję *I* fraz f_1^I o jednakowym prawdopodobieństwie, każda z fraz f_i z f_1^I jest tłumaczona na frazę e_i i mogą one zostać poprzestawiane.
  - ✓ Tłumaczenie jest modelowane jako rozkład $\varphi(f_i|e_i)$,
  - ✓ przestawianie fraz ma rozkład $d(\text{start}_i,\text{end}_{i-1}) = \alpha^{|\text{start}_i - \text{end}_{i-1} - 1|}$ z odpowiednią wartością α, wprowadza się też czynnik (koszt słowa) ω>1, aby chętniej generować krótsze zdania

Zatem najlepsze tłumaczenie to

$$\mathbf{e}_{best} = \text{argmax}_{\mathbf{e}}\ p(\mathbf{e}|\mathbf{f}) = \text{argmax}_{\mathbf{e}}\ p(\mathbf{f}|\mathbf{e})\ p\_LM(\mathbf{e})\ \omega^{\text{length}(\mathbf{e})},$$

gdzie $p(\mathbf{f}|\mathbf{e})$ jest przedstawiane jako:

$$p(f_1^I|e_1^I) = \Phi_{i=1}^{I}\ \varphi(f_i|e_i)\ d(\text{start}_i,\text{end}_{i-1})$$

www.statmt.org/moses, 06

Parametry modelu – dopasowanie fraz

- ✓ **IBM model I** $P(f,a|e) = P(m|e)\, P(a|m,e)\, P(f|a,m,e)$

  Dla danego zdania docelowego *e* oblicza się prawdopodobieństwo, że długość jego tłumaczenia *f* wynosi *m*. Kolejno oblicza się prawdopodobieństwo dopasowania *a*. Dopiero na podstawie dopasowania można obliczyć prawdopodobieństwo, że zdanie źródłowe *f* jest tłumaczeniem zdania *e*

- ✓ **IBM model II**

  Prawdopodobieństwo podobne jak w Modelu I, lecz dodatkowo pojawia się prawdopodobieństwo dopasowania, czyli parametr uwzględniający pozycję słowa w zdaniu.

- ✓ **IBM model III**

  W tym modelu występują cztery parametry: „płodność” wyrazu, czyli ilość słów, która musi powstać w celu przetłumaczenia wyrazu; informacje o położeniu słowa w zdaniu; tłumaczenie wyrazu; nieprawdziwe słowa

- ✓ **IBM model IV**

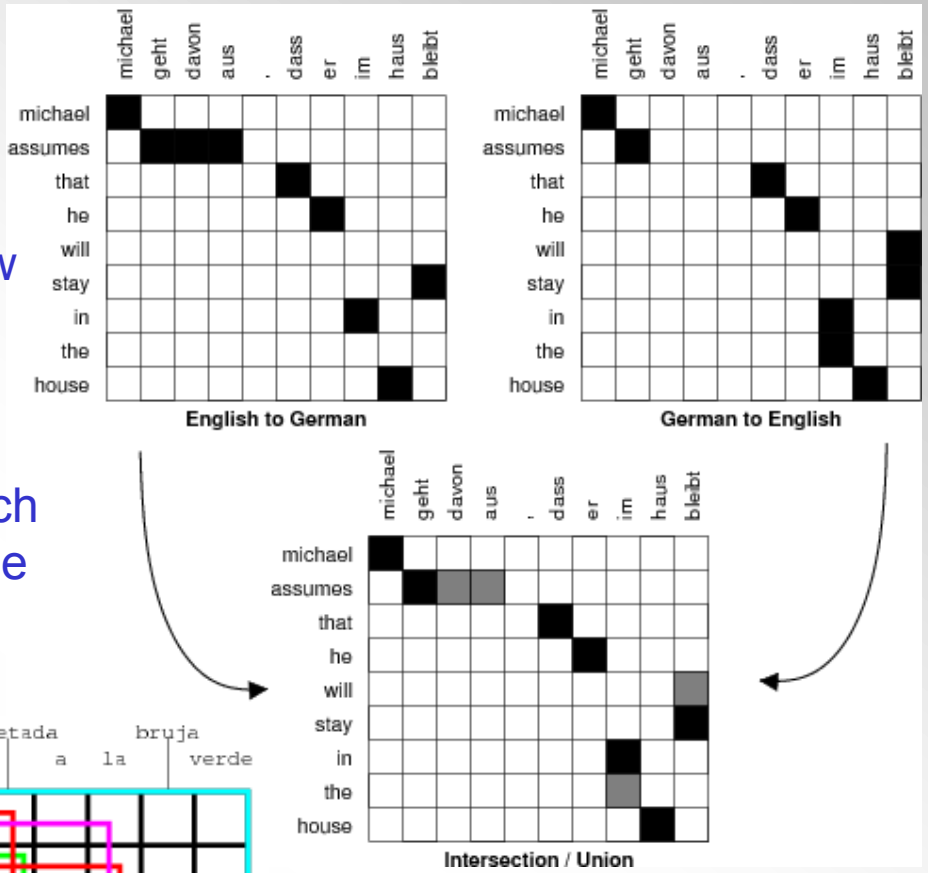
  Umieszczenie kolejnych przetłumaczonych słów zależy od innych wcześniejszych słów pochodzących od tego samego źródła. Wprowadza podział na grupy słów występujące razem i rozłącznie

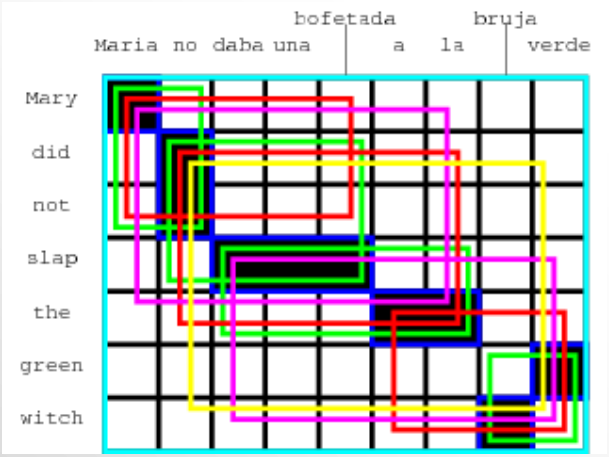
- ✓ **IBM model V**

  Podobny do Modelu IV, lecz uniemożliwia powstawanie zdań, w których brakuje słów, takich gdzie dana pozycja zajmowana jest przez więcej niż jedno słowo oraz umieszczania słów przed pierwszym lub za ostatnim wyrazem przetłumaczonego zdania. Najbardziej skuteczny, jednak zazwyczaj odrzucany na korzyść modeli II, III i IV m.in. ze względu na złożoność obliczeniową

Tłumaczenie fraz, czyli phrase alignment

- ✓ Dopasowanie na podstawie bilingualnego korpusu – modelowanie
  - ✓ Dwukierunkowe dopasowanie słów pozwala wyznaczyć te najbardziej odpowiednie słowa
  - ✓ Na tej podstawie próbuje się tłumaczyć frazy, np. szukamy takich fraz które są dopasowane do siebie z wykorzystaniem tylko słów z ich wnętrza (Och, 2003)





www.statmt.org/moses, 06

Multimedia Department

GIZA++

- ✓ GIZA++ (Och, Ney) IBM model IV

Sentence pair (1) source length 9 target length 6 alignment score : 2.62269e-09

aligned

paczka ciastek i cytrynowy baton .

NULL ({ }) a ({ }) bag ({ }) of ({ }) cookies ({ 1 2 }) and ({ 3 }) a ({ }) lemon ({ 4 }) bar ({ 5 }) . ({ 6 })

not aligned

```
GROW-DIAG-FINAL(e2f,f2e):
  neighboring = ((-1,0),(0,-1),(1,0),(0,1),(-1,-1),(-1,1),(1,-1),(1,1))
  alignment = intersect(e2f,f2e);
  GROW-DIAG(); FINAL(e2f); FINAL(f2e);

 GROW-DIAG():
  iterate until no new points added
    for english word e = 0 ... en
      for foreign word f = 0 ... fn
        if ( e aligned with f )
          for each neighboring point ( e-new, f-new ):
            if ( ( e-new not aligned or f-new not aligned ) and
                 ( e-new, f-new ) in union( e2f, f2e ) )
              add alignment point ( e-new, f-new )
 FINAL(a):
  for english word e-new = 0 ... en
    for foreign word f-new = 0 ... fn
      if ( ( e-new not aligned or f-new not aligned ) and
           ( e-new, f-new ) in alignment a )
        add alignment point ( e-new, f-new )
```

Trening w obu kierunkach, aby uzyskać mapowanie many-to-many

Dopasowanie pl-en heurystyką grow-diag-final

0-0 1-1 4-2 0-3 1-3 2-4 4-5 3-6 4-7 5-8

0-0 1-0 1-1 2-2 2-3 2-4 3-5

0-0 2-1 3-1 1-2 0-3 0-4 4-5

GIZA ++

- ✓ Tabela tłumaczenia *w(e|f) (maximum likelihood)*

```
cytrynowy squash 0.1000000
cytrynowy one 0.0002875
cytrynowy please 0.0000734
cytrynowy flavored 0.1428571
cytrynowy lemon 0.1111111
cytrynowy , 0.0000913

i ||| and ||| 0-0
cytrynowy ||| lemon ||| 0-0
Poproszę pieczonego ||| a backed  ||| 0-0 1-0 1-1
Poproszę pieczonego ziemniaka ||| a backed potato
please ||| 0-0 1-0 1-1 2-2 2-3 2-4
```

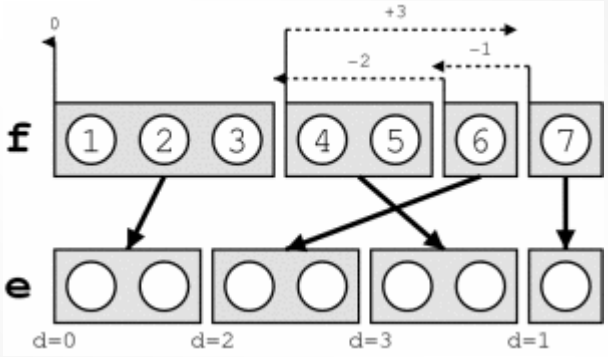
Wyznaczanie wyników tłumaczenia

- ✓ Obliczanie prawdopodobieństwa tłumaczenia frazy *φ(e|f)*
  - ✓ Plik z frazami jest sortowany - tłumaczenia kolejnych fraz są obok siebie
  - ✓ Zliczane są tłumaczenia dla danej frazy i obliczane *φ(e|f)* dla danej frazy obcojęzycznej *f*. Aby wyznaczyć *φ(f|e)* plik z odwrotnościami tłumaczeń jest sortowany i *φ(f|e)* jest obliczane dla danej frazy ENG
  - ✓ Oprócz rozkładów prawdopodobieństw *φ(f|e)* and *φ(e|f)* obliczane są dodatkowe funkcje wpływające na wynik, np. wagi leksykalne, word penalty, phrase penalty, etc. Currently, lexical weighting is added for both directions and a fifth score is the phrase penalty.
- ✓ Zwykle oblicza się :
  - ✓ inverse phrase translation probability *φ(f|e)*
  - ✓ inverse lexical weighting *lex(f|e)*
  - ✓ direct phrase translation probability *φ(e|f)*
  - ✓ direct lexical weighting *lex(e|f)*
  - ✓ phrase penalty (always *exp(1) = 2.718*)

```
pieczonego ziemniaka , zieloną fasolkę i ||| a baked potato , green beans and ||| 1 0.000177825 1
0.00559922 2.718 ||| ||| 1 1
pieczonego ziemniaka , zieloną fasolkę ||| a baked potato , green beans ||| 1 0.000230482 1
0.00601299 2.718 ||| ||| 1 1
pieczonego ziemniaka , zieloną ||| a baked potato , green ||| 1 0.00184386 1 0.012026 2.718 |||
||| 1 1
pieczonego ziemniaka , ||| a baked potato , ||| 1 0.0125997 1 0.0175379 2.718 ||| ||| 1 1
pieczonego ziemniaka . ||| a baked potato . ||| 1 0.0234301 1 0.028652 2.718 ||| ||| 1 1
pieczonego ziemniaka i ||| baked potato and ||| 1 0.0363686 1 0.154584 2.718 ||| ||| 1 1
```

Reordering model

- ✓ Przesuwanie słów dokonywane jest na podstawie miary liniowo-zależnej od odległości słów. Przykładowo, koszt przeskoczenia o dwa słowa jest dwa razy wyższy niż przeskok o jedno słowo…



Tuning - MERT

- In order to combine evidence from different models, it is standard practice to use a discriminative linear model, with the log probabilities as features. If the features of the model are h1, . . . , hr , which depend on e and f, then the best translation is given by

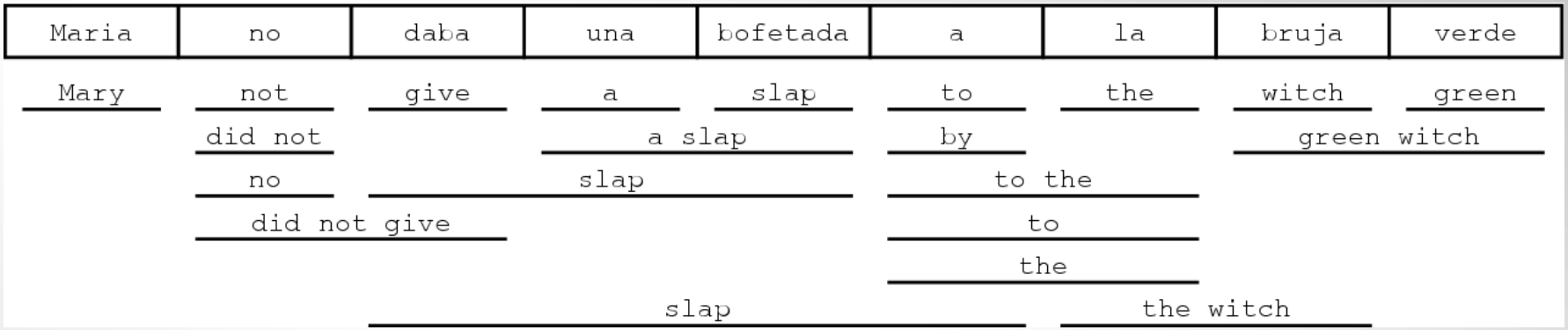
$$e^*(\lambda) = \arg\max_{e} \sum_{i=1}^{r} \lambda_i h_i(e, f)$$

- and depends on the feature weights λ1, . . . , λr .
- The standard solution is to use minimum error rate training (MERT), a procedure introduced by Och (2003), which searches for weights minimizing a given error measure, or, equivalently, maximizing a given translation metric. This algorithm enables the weights to be optimized so that the decoder produces the best translations (according to some automatic metric *Err* and one or more references ref) on a development set of parallel sentences.

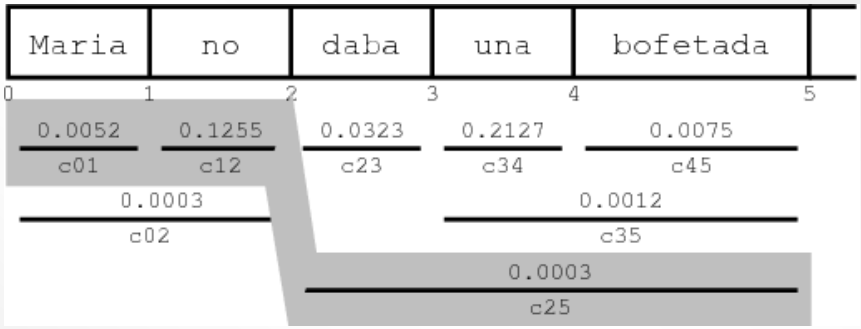
$$\lambda^*(\lambda) = \arg\min_{\lambda} Err(e^*(\lambda); ref)$$

Dekodowanie czyli tłumaczenie

- ✓ System wyszukuje możliwe tłumaczenia fraz



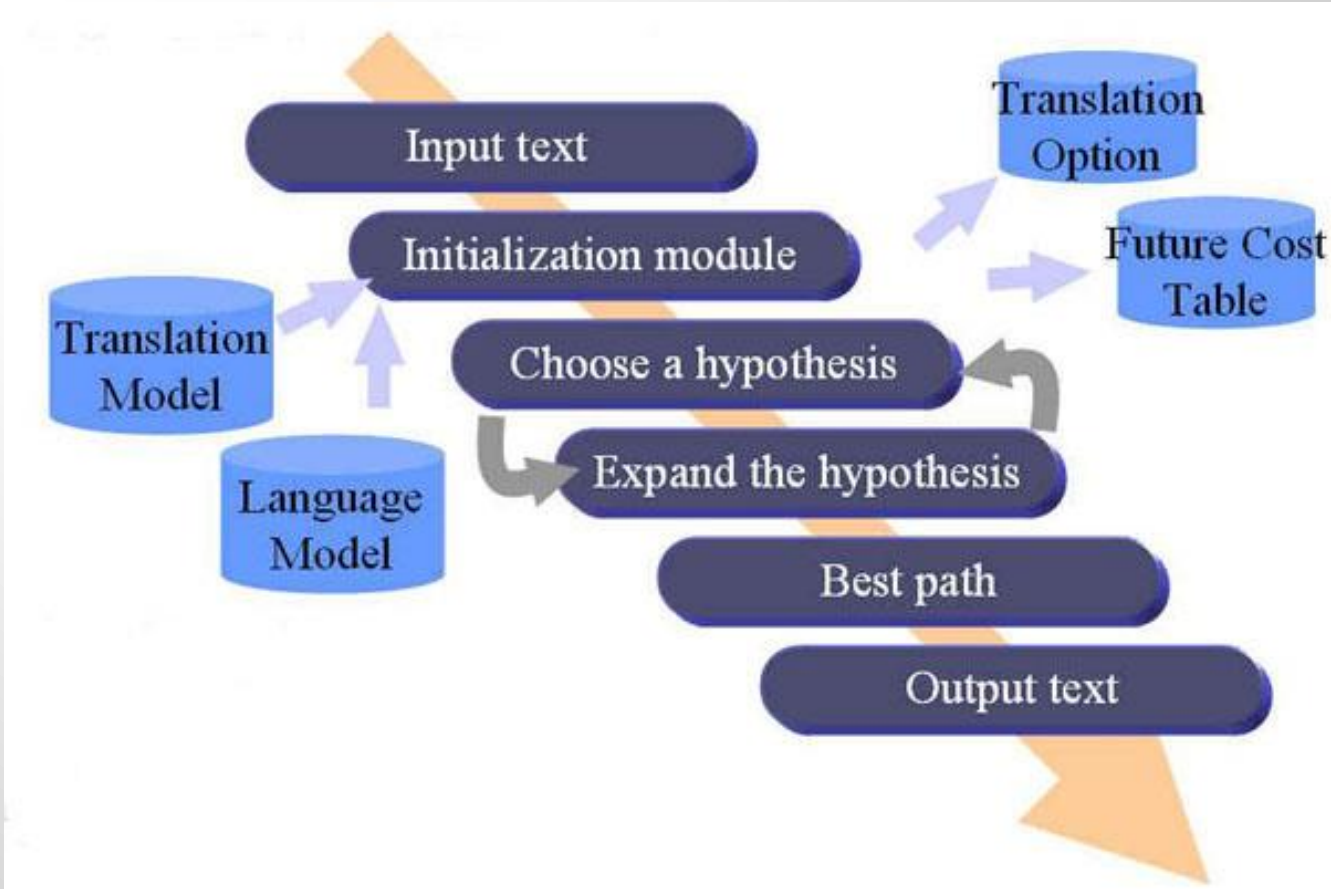
- ✓ Uwzględnia koszt związany z tłumaczeniem, przestawianiem fraz i modelem języka – minimalizuje koszt, czyli maksymalizuje prawdopodobieństwo
- ✓ Metoda wyszukiwania jest zbliżona do używanej w rozpoznawaniu mowy



- ✓ Można generować listę najlepszych tłumaczeń

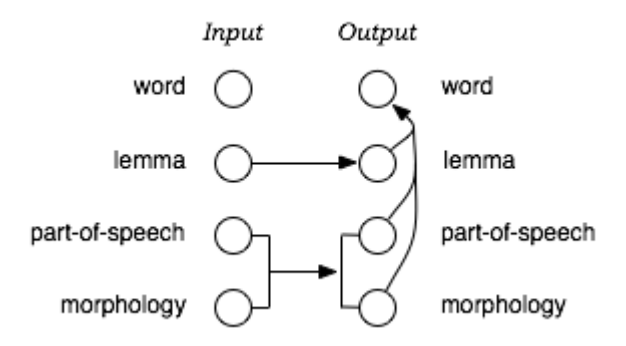
www.statmt.org/moses, 06

Proces dekodowania



Factored models

- ✓ Dodatkowe źródło informacji dla tłumaczeń
- ✓ Wiele aspektów tłumaczenia da się wyjaśnić na poziomie składni, morfologii
- ✓ dla języków o bogatej fleksji lematyzacja może ograniczyć słownik
- ✓ Tłumaczenie i generacja w kolejnych krokach,
  - ✓ Translate input lemmas into output lemmas
  - ✓ Translate morphological and POS factors
  - ✓ Generate surface forms given the lemma and linguistic factors
- ✓ (nasze doświadczenia z modelami języka dla PL, model koneksjonistyczny) – kombinacja słów, lematów i kategorii gramatycznych – ppl o 20% niższe niż 3-gramów



Statystyczne modele języka

- ✓ Zwykle estymacja N-gramów z różnymi heurystykami wygładzania modeli (Good-Turing, Knesser-Ney, inne)
- ✓ Problem z mała ilością tekstów z domeny tłumaczenia (korpusy bilingualne są małe):
  - ✓ Adaptacja modeli języka
- ✓ Wykorzystanie zewnętrznych źródeł danych - jednojęzykowych
- ✓ Mały korpus A i duży korpus B, zapewne spoza domeny
- ✓ Podejście interpolacyjne: model merging (interpolacja liniowa, back-off), dynamic cache models (modele klas i adaptacja), MAP adaptation
- ✓ Inne: Max Entopy, ….
- ✓ Praktycznie – im bliższe domeny A i B tym lepiej

Miary jakości tłumaczenia

- ✓ BLEU – „im bliżej ludzkiego tłumaczenia tym lepiej”, zmodyfikowana precyzja* proponowanego tłumaczenie względem wielu wzorcowych tłumaczeń

  * Zmodyfikowana bo SMT zwykle generuje więcej słów niż jest w oryginale

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{n\text{-}gram \in C} Count_{clip}(n\text{-}gram)}{\sum_{C' \in \{Candidates\}} \sum_{n\text{-}gram' \in C'} Count(n\text{-}gram')}.$$

$$Score = \sum_{n=1}^{N} \left\{ \sum_{\substack{\text{all } w_1 \ldots w_n \\ \text{that co-occur}}} Info(w_1 \ldots w_n) \Big/ \sum_{\substack{\text{all } w_1 \ldots w_n \\ \text{in sys output}}} (1) \right\} \cdot \exp\left\{ \beta \log^2 \left[\min\left(\frac{L_{sys}}{L_{ref}}, 1 \right) \right] \right\}$$

- ✓ NIST – zbliżone do BLEU, ale uwzględnia istotność N-gramów
- ✓ METEOR
- ✓ WER $$WER = \frac{S + D + I}{N}$$
- ✓ MEANT i HMEANT – semantic role labels

who	what	whom	when	where
agent	patient	benefactive	temporal	locative
why	**how**			
purpose	degree, manner, modal, negation, other			

[0] srl ◂ ☑ done ▸

And the problems in the municipality are also gritty and urban .			And the problems in the community are of crucial urban nature .		
head of frame	**role**	**slot filler**	**head of frame**	**role**	**slot filler**
are	agent (who)	the problems	are	agent (who)	the problems
are	locative (where)	in ... municipality	are	locative (where)	in ... community
are	other (how)	also	are	experiencer/patient (what)	of ... nature
are	experiencer/patient (what)	gritty ... urban			

F_i = # correct or partially correct fillers for PRED i in MT

MT_i = total # fillers for PRED i in MT

REF_i = total # fillers for PRED i in REF

$$P = \sum_{matched\ i} \frac{F_i}{MT_i}$$

$$R = \sum_{matched\ i} \frac{F_i}{REF_i}$$

$$P_{total} = \frac{P_{correct} + 0.5 P_{partial}}{\text{total \# predicates in MT}}$$

$$R_{total} = \frac{P_{correct} + 0.5 P_{partial}}{\text{total \# predicates in REF}}$$

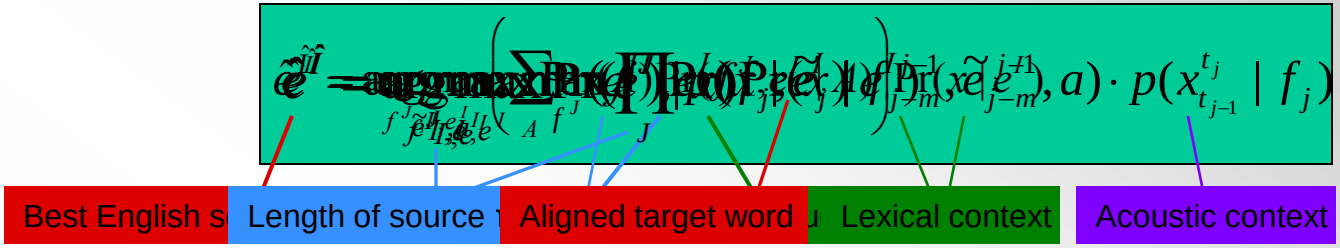
$$\text{HMEANT} = \frac{2 * P_{total} * R_{total}}{P_{total} + R_{total}}$$

(Birch at al.. 13)

Multimedia Department

Sprzężenie rozpoznawania i tłumaczenia

- ✓ Wykorzystanie krat słów jako wejścia do tłumaczenia z użyciem FST
- ✓ Noisy Channel Model – Shannon, Bayes



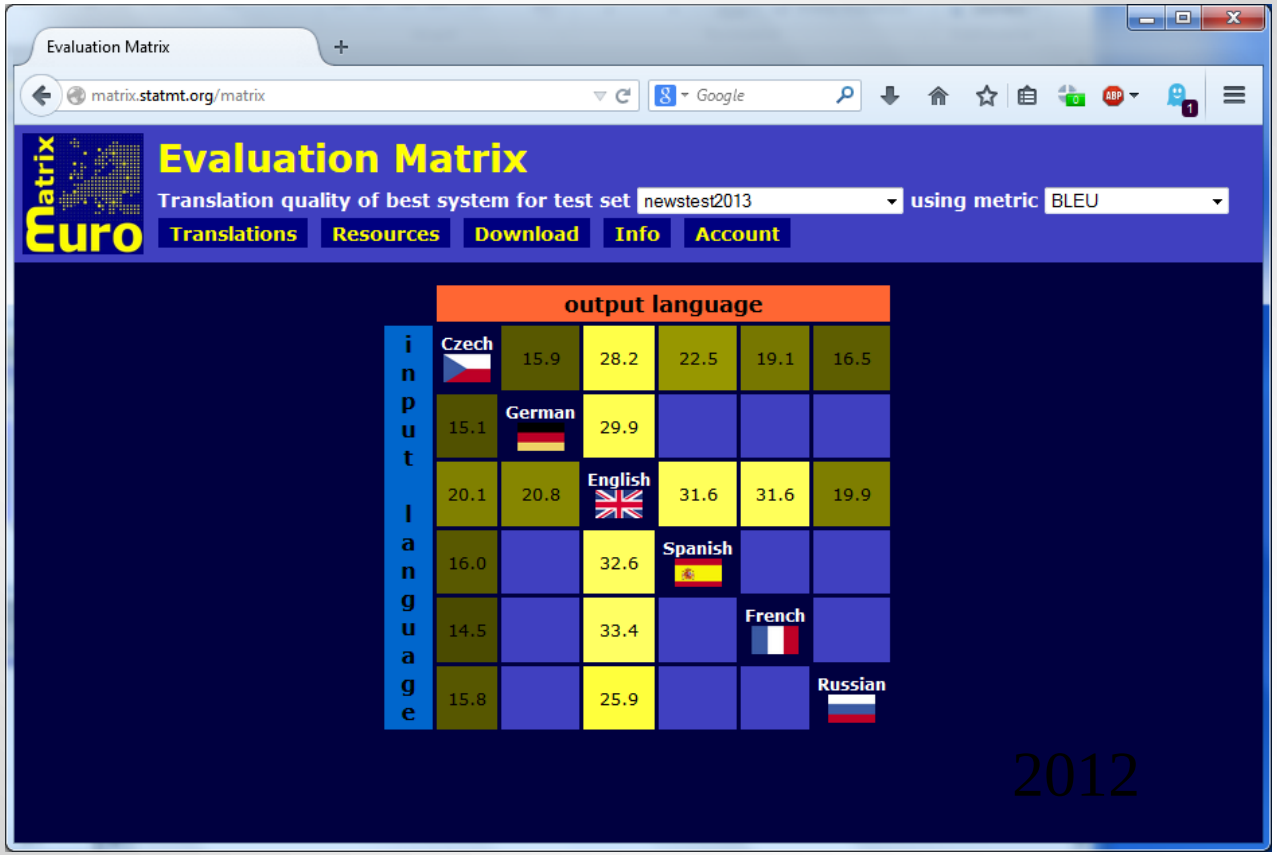
- ✓ Using an alignment model, A
- ✓ Instead of modeling the alignment, search for the best alignment

- ✓ Nikt tego nie wykorzystuje bezpośrednio…..
- ✓ Zwykle tłumaczy się najlepszą hipotezę z ASR

Matusov, 05

Nasze eksperymenty PL-EN

Jakość tłumaczenia (domena TED)



http://matrix.statmt.org/matrix

2012

TED : MT English-Polish test2011 (MT_{EnPl})

System	*case sensitive* BLEU	TER	*case insensitive* BLEU	TER
PJIIT	15.66	68.65	16.61	67.16
UEDIN	13.10	70.96	13.69	69.86

2013

TED : MT Polish-English test2011 (MT_{PlEn})

System	*case sensitive* BLEU	TER	*case insensitive* BLEU	TER
PJIIT	23.29	60.99	24.37	59.36
UEDIN	21.69	62.73	22.57	61.24

2014

System	Year	BLEU
BASE	2010	16,70
BEST	**2010**	**23,74**
BASE	2011	20,40
BEST	**2011**	**28,00**
BASE	2012	17,22
BEST	**2012**	**23,15**
BASE	2013	18,16
BEST	**2013**	**28,62**
BASE	2014	14,71
BEST	**2014**	**18,96**

SLT quality: TED

- ✓ IWSLT benchmark

2012

TED : SLT English-French test 2012(SLT_{EnFr})

System	*case sensitive* BLEU	*case sensitive* TER	case insensitive BLEU	case insensitive TER
KIT	32.21	48.58	32.86	47.65
MSR-FBK	29.92	53.30	31.03	52.10

TED : MT English-French test 2012(MT_{EnFr})

System	*case sensitive* BLEU	*case sensitive* TER	*case insensitive* BLEU	*case insensitive* TER
EU-BRIDGE	42.13	38.72	42.99	37.83
UEDIN	41.21	39.83	42.02	38.94
KIT	41.02	39.22	41.96	38.34
RWTH	40.06	39.95	40.79	39.11
PRKE-IOIT	39.94	41.52	40.64	40.75
MITLL-AFRL	39.76	41.47	40.97	40.31
FBK	39.51	40.56	40.11	39.80

TED : MT English-German test 2012 (MT_{EnDe})

System	*case sensitive* BLEU	*case sensitive* TER	*case insensitive* BLEU	*case insensitive* TER
KIT	23.24	56.17	24.00	55.02
NTT-NAIST	22.86	56.12	24.10	54.57
UEDIN	22.53	57.43	23.26	56.27
RWTH	22.32	57.11	23.04	55.91
POSTECH	20.43	59.14	21.02	58.05

2013

TED : SLT English-French test2011(SLT_{EnFr})

System	*case sensitive* BLEU	*case sensitive* TER	case insensitive BLEU	case insensitive TER
KIT	31.06	50.70	31.93	49.61
MSR-FBK	27.21	56.22	28.32	54.82

TED : MT English-French test 2011(MT_{EnFr})

System	*case sensitive* BLEU	*case sensitive* TER	*case insensitive* BLEU	*case insensitive* TER
EU-BRIDGE	40.71	40.56	41.55	39.72
UEDIN	40.61	40.97	41.48	40.08
MITLL-AFRL	39.35	42.18	40.62	41.08
RWTH	39.25	41.24	40.16	40.29
KIT	39.11	41.74	40.33	40.63
PRKE-IOIT	38.80	42.86	39.54	42.12
FBK	38.41	42.02	39.09	41.25

TED : MT English-German test2011 (MT_{EnDe})

System	*case sensitive* BLEU	*case sensitive* TER	*case insensitive* BLEU	*case insensitive* TER
UEDIN	27.13	50.97	27.75	50.09
KIT	26.29	50.67	26.97	49.76
NTT-NAIST	26.04	50.13	27.27	48.82
RWTH	25.86	51.56	26.58	50.52
POSTECH	23.48	53.71	24.06	52.89

- ✓ Steady progress for all components – closer to real world applications

Polsko-angielskie SLT

- ✓ Initial results for Euronews domain (broadcast news)
- ✓ Problems:
  - ✓ Text denormalization (ASR output – capitalization, punctation)
  - ✓ In-domain parallel text data hardly available, open domain
  - ✓ Vocabulary sizes (huge for PL, smaller for EN)

Data set	BLEU
Original	36,52
Normalized	25,12
ASR output	22,94

Przygotowanie korpusu BTEC dla języka polskiego - 2008

- ✓ Współpraca z ATR Spoken Language Translation Research Laboratories, Kyoto, dr Eiichiro Sumita
- ✓ Ok. 120 000 zdań z rozmów japońskich turystów mówiących po angielsku („angielskie rozmówki”), standardowy benchmark w wielu ewaluacjach, język mówiony, wersje: niemiecki, francuski, arabski, włoski, chiński
- ✓ Cel:
  - ✓ Przygotowanie systemu automatycznego tłumaczenia z mowy na mowę pomiędzy polskim i angielskim dla ograniczonej domeny (rozmówki turystyczne)
  - ✓ Przygotowanie równoległego korpusu polsko-angielskiego (dopasowanie na poziomie zdań i fraz)
- ✓ Nasza praca:
  - ✓ Przetłumaczenie zdań z angielskiego na polski siłami studentów,
  - ✓ Nadzór nad jakością – panie z lektoratu
  - ✓ Oganizacja pracy: KM + studenci
  - ✓ Sformatowanie plików zgodnie z formatem BTEC

ATR Advanced Telecommunications Research Institute International

SLC 音声言語コミュニケーション研究所 Spoken Language Communication Research Laboratories

BTEC - wyniki

- ✓ BLEU=71

```
Jestem Anne Brown.
To będzie około trzy razy więcej.
Musiałem wybrać zły numer.
On jest równym gościem.
Włóż coś wygodnego.
Czy wiesz, czy w tej restauracji wymagane są rezerwacje?
Ile kosztuje podróż w obie strony?
Proszę mówić mi Sue.
Gdzie jest stanowisko British Airlines?
```

```
i 'm anne brown .
it will be about three times more .
i must have dialed the wrong number .
he is równym a guest .
something włóż wygodnego .
do you know if at this restaurant are required reservations ?
how much is the round trip , please ?
please call me sue .
where 's the british airlines ?
```

Wyniki PL-EN BLEU dla wybranych domen

DOMENA	PL-EN	EN-PL
TED 2013	28,62	26,61
EUROPARL	82,48	70,73
SUBTITLES	53,51	52,01
EMEA	76,34	73,32

Co decyduje o polepszeniu wyników?

- ✓ *Is no data like more data*
- ✓ Cierpliwość i dokładność w czyszczeniu danych:
  - ✓ Tokenizacja
  - ✓ Jasne oznaczanie końca zdania: na końcu kropka, po której musi znaleźć się ENTER (\n)
  - ✓ Jedno zdanie w jednej linii
  - ✓ uboga interpunkcja i brak znaków specjalnych @#$%^&*_=+-><;:'"|
  - ✓ Precyzyjne dopasowanie korpusów billingualnych
  - ✓ Usunięcie tłumaczeń na inne języki (np. francuski)

```
1 Działania podjęte w wyniku rezolucji Parlamentu: Patrz protokół.
2 Składanie dokumentów: patrz protokół.
3 Oświadczenia pisemne (art. 116 Regulaminu): patrz protokół.
4 Teksty porozumień przekazane przez Radę: patrz protokół.
5 Skład Parlamentu: patrz protokół.
6 Skład komisji i delegacji: patrz protokół.
7 Przyszłe działania w dziedzinie patentów (złożone projekty rezolucji): Patrz protokół.
8 Porządek dzienny następnego posiedzenia: patrz protokół.
9 Zamknięcie posiedzenia.
10 (La seduta è tolta alle 23.55).
11 Otwarcie posiedzenia.
```

- ✓ Lepszy model języka (wytrenowany na dodatkowych danych i zaadaptowany do danej domeny (konieczne zbliżone dane)
- ✓ Konieczne zbalansowanie słowników (PL 200 tys, EN – 80 tys słów)
- ✓ Szyk zdania SVO? Lematyzacja?

U-STAR PL<->EN

- ✓ U-STAR consortium (leader NICT): 23 languages,
- ✓ Distributed server structure, standards adopted by ITU
- ✓ VoiceTra4U application for iPhone, iPad
- ✓ Simultaneous talk of 5 people at the same time
- ✓ domain: travel phrasebook
- ✓ PJIIT:
  - ✓ PL <-> EN, based on the BTEC
  - ✓ ASR Julius, CentOS, Tomcat
  - ✓ **~~own Android client~~**

Eu-Bridge

- ✓ EU-Bridge (www.eu-bridge.eu), use cases:
  - ✓ Subtilting on-line (Red Bee Media)
  - ✓ Europarliament – automatic translation of voting sessions, list of proper names and common words
  - ✓ webinars (Andrexen)
  - ✓ lecture translation (KIT,KIT2)
  - ✓ Euronews, TED talks

- ✓ PJIIT in EU-Bridge
  - ✓ Development of speech translation technology for the Polish language(ASR, MT)
  - ✓ After 3 years: big step forward – see demos
  - ✓ ASR, PL-EN for lecture translation, EN-PL for Webinars

Multimedia Department

Mobile technologies

- Applications for Android and iOS devices
- Main purpose is S2S translation
- Purely as a “cloud”-based solution
- A combination of 3 technologies:
  - **ASR** - Julius, Kaldi or our own engine
  - **MT** – Moses
  - **TTS** - our own unit selection (Festival), diphone (MBrola) or commercial (Loquendo)

MCML protocol (U-Star)

- Java libraries and a Tomcat based server
- Hierarchical structure
- Pivoting for MT
- Detailed XML-based protocol (lots of metac
- ITU-T standardization: F.745 and H.625
- Links: *demo* *documentation*

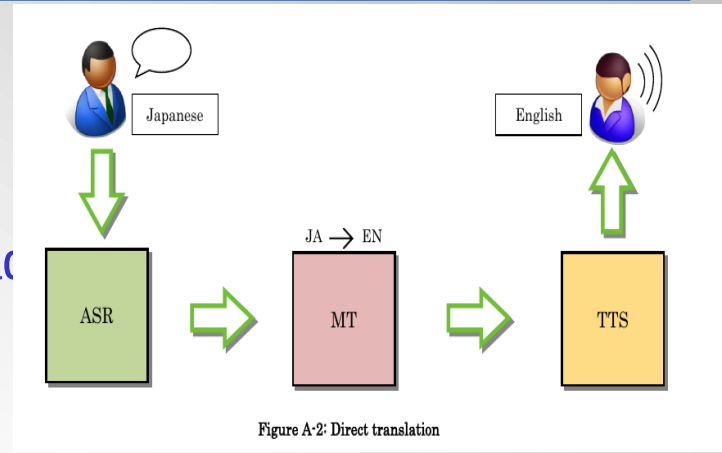


Figure A-2: Direct translation

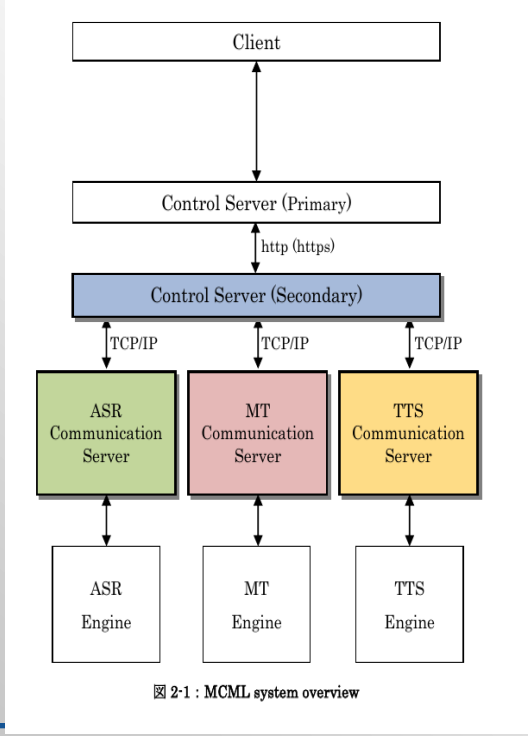


図 2-1 : MCML system overview

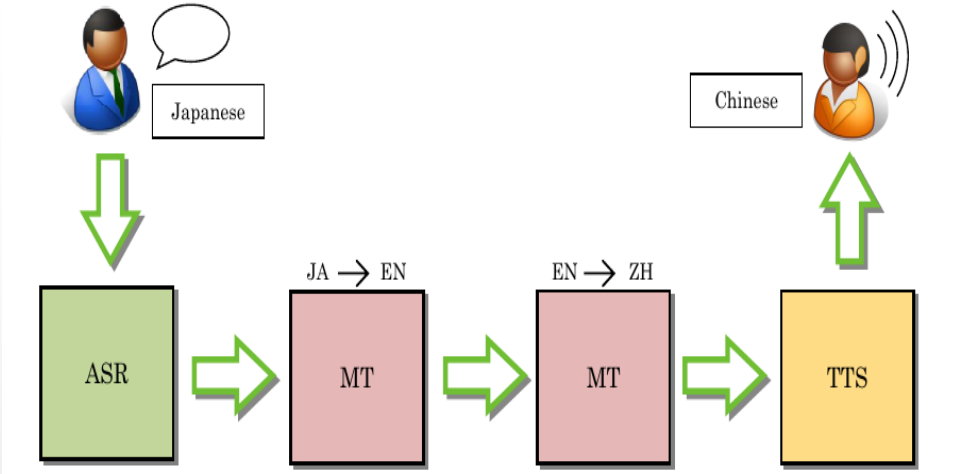


Figure A-3: Pivot translation

MCloud protocol (EU-bridge / Mobile Tech.)

- Adaptation of the protocol used by Jibbigo
- Very straightforward XML-based (serialized) protocol
- Single point of control (the mediator)
- User applications and engines are clients of the mediator
- Links: *demo* *documentation*

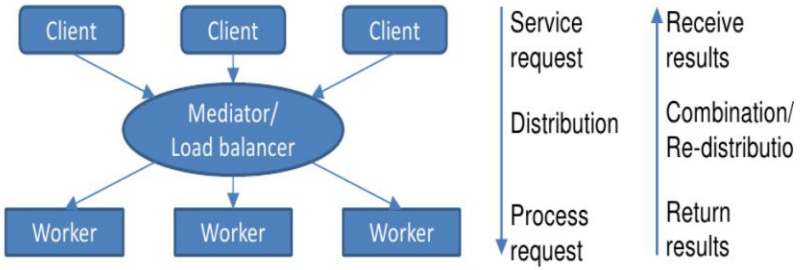


Figure 1.2: Overview of the service architecture

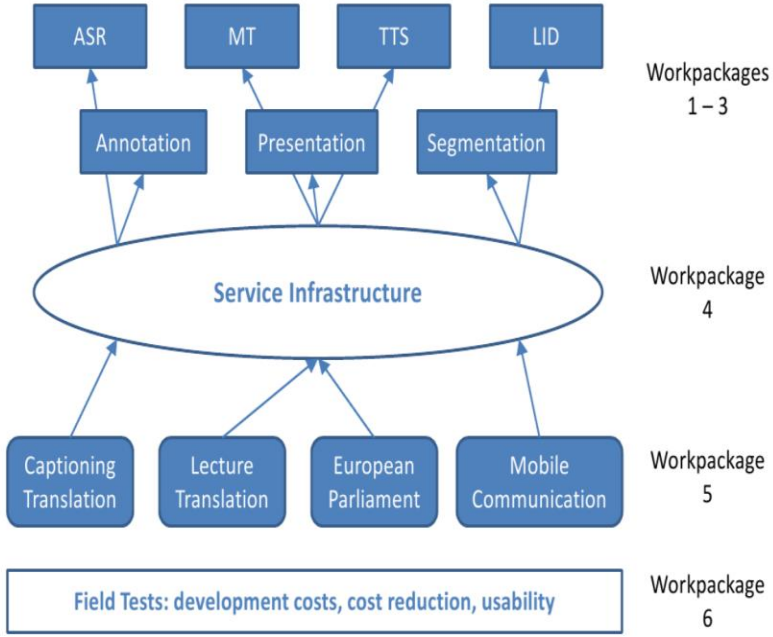
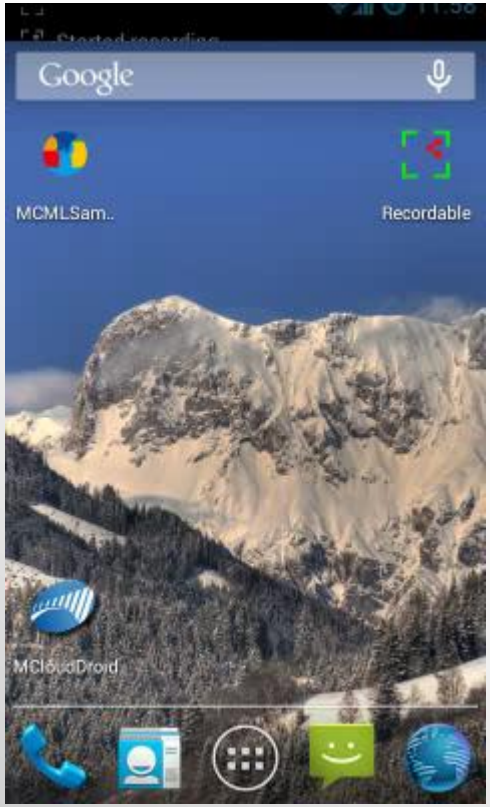
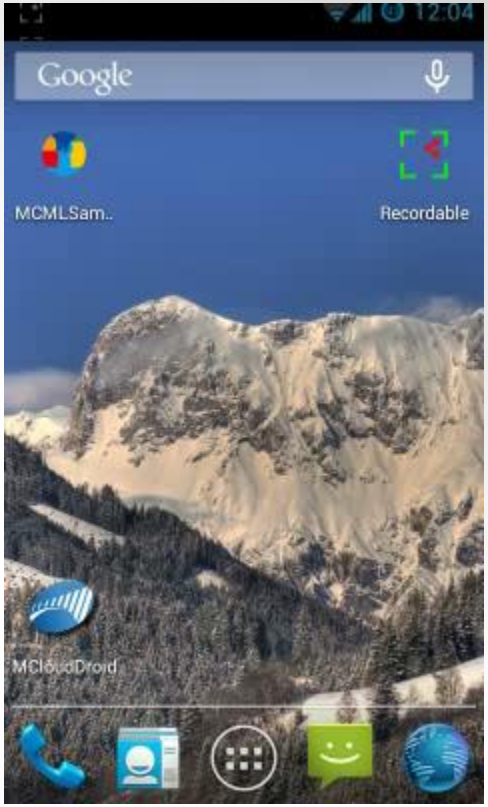


Figure 1.1: Placement of the service architecture within the project

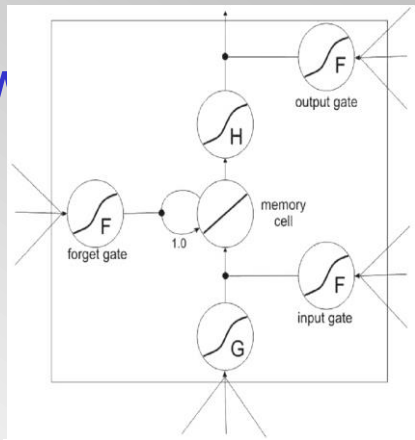
Mobile demos

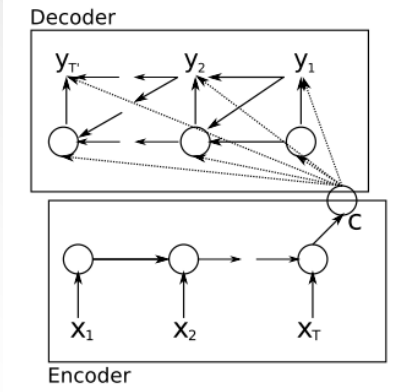




Sieci neuronowe w tłumaczeniu maszynowym

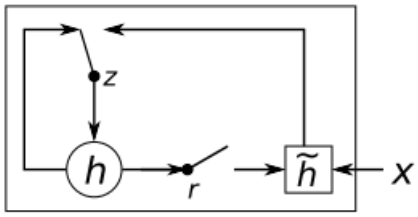
- ✓ Deep learning: przełom w rozpoznawaniu wzorców
- ✓ Rekurencyjne sieci neuronowe, sieci LSTM
- ✓ Dekoder-enkoder w jednej sieci rekurencyjnej
  - Zapis słów i fraz w postaci wektora liczb
  - Zdekodowanie wektorów do symboli z języka wyjściowego
- ✓ RNNEncDec maksymalizuje





$$\max_{\boldsymbol{\theta}} \frac{1}{N} \sum_{n=1}^{N} \log p_{\boldsymbol{\theta}}(\mathbf{y}_n \mid \mathbf{x}_n),$$

Dla parametrów sieci i par (x_n, y_n) (wejscie, wyjście), przy czym neurony warstwy ukrytej wybiórczo zapamiętują bądź zapominają poprzednie stany



Cho at al.., 14

Wektoryzacja słów : słowa zbliżone

```
Enter word or sentence (EXIT to break): dworzec

Word: dworzec  Position in vocabulary: 25130
```

Word	Cosine distance
parking	0.693664
lotnisko	0.650875
dworca	0.643771
autobus	0.601179
peron	0.593517
parkingu	0.572672
mikrobus	0.564531
hotelu	0.562353
pociąg	0.561882
tramwaj	0.558675

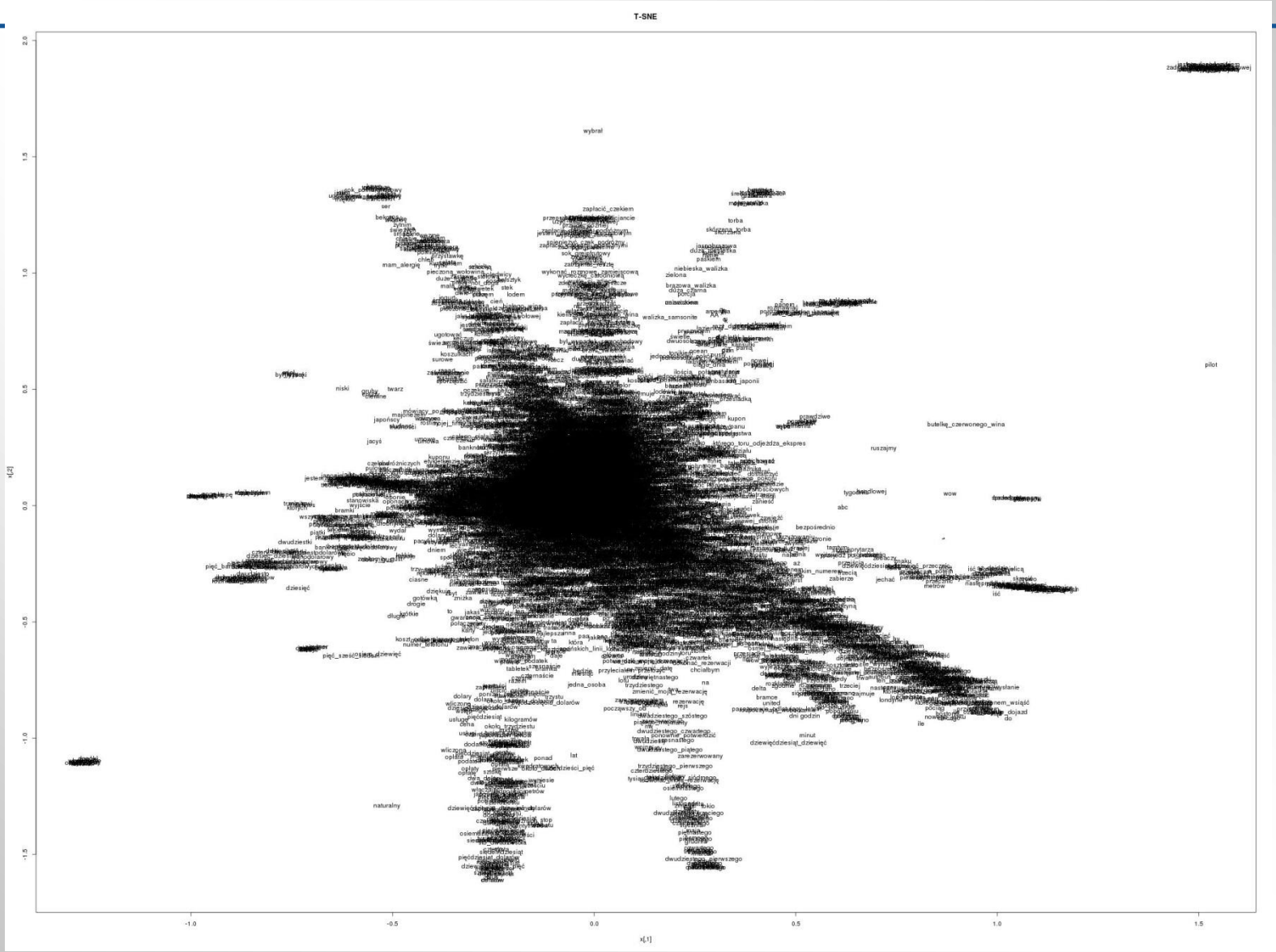
Wektoryzacja słów : słowa zbliżone

```
Enter word or sentence (EXIT to break): ludzkie_zdrowie

Word: ludzkie_zdrowie  Position in vocabulary: 529810
```

Word	Cosine distance
zdrowie_ludzkie	0.596982
środowisku_naturalnym	0.587597
środowisko_naturalne	0.576780
zdrowie	0.561626
środowisko_przyrodnicze	0.558820
zanieczyszczenia_środowiska	0.552765
dobrostan_zwierząt	0.549674
szkodliwego_wpływu	0.531352
stan_środowiska_naturalnego	0.531025
jakość_powietrza	0.530011
środowisko	0.524710
szkodliwe_oddziaływanie	0.521243
zdrowotność	0.521103
zdrowie_polaków	0.520689
dla_środowiska_naturalnego	0.520205
zagrożenie_klimatyczne	0.519684
ochronę_środowiska_naturalnego	0.519296
bezpieczeństwo_żywności	0.517315

Wektoryzacja słów: mapy



Word2vec (Mikolov,13) , t-SNE (van der Maaten and G.E. Hinton, 08), korpus 750 mln słów)

Przykłady tłumaczeń (BTEC, 50000 iteracji)

```
Input: do you know of any interesting restaurants ? <eol>
Target: znasz jakieś ciekawe restauracje ? <eol>
Input:   do you know of any interesting restaurants ? <eol>
Output:  czy znasz jakieś ciekawe restauracje w okolicy ? <eol>

Input: is the swimming pool free of charge ? <eol>
Target: czy basen jest wolny od opłat ? <eol>
Input:   is the swimming pool free of charge ? <eol>
Output:  czy basen jest wolny od opłat ? <eol>

Input: how much will it cost to go to the airport by taxi ? <eol>
Target: ile kosztuje przejazd na lotnisko taksówką ? <eol>
Input:   how much will it cost to go to the airport by taxi ? <eol>
Output:  ile kosztuje przejazd do miasta ? <eol>

Input: my watch is three minutes fast . <eol>
Target: mój zegarek śpieszy się o trzy minuty . <eol>
Input:   my watch is three minutes fast . <eol>
Output:  mój zegarek spóźnia się o trzy minuty . <eol>

Input: you change trains at the next station . <eol>
Target: zmieniasz pociągi na następnej stacji . <eol>
Input:   you change trains at the next station . <eol>
Output:  zmieniasz pociągi w następny stacji . <eol>

Input: i cannot hear you very well . <eol>
Target: nie za dobrze pana słyszę . <eol>
Input:   i cannot hear you very well . <eol>
Output:  słabo pana słyszę . <eol>
```

Podsumowanie

- ✓ Dzięki uczestnictwu w Eu-Bridge rozwinęliśmy technologię SLT dla polskiego
- ✓ Dla wybranych domen (medyczne, rozmówki) jakość tłumaczenia jest zaskakująco dobra i bliska zastosowaniom komercyjnym
- ✓ Do stworzenia systemów potrzebne są duże zasoby i moce obliczeniowe
- ✓ Dominują modele statystyczne, szczególnie obiecujące jest zastosowanie RNN i DL

Multimedia Department

Dziękuje za uwagę!

☺

kmarasek@pjwstk.edu.pl