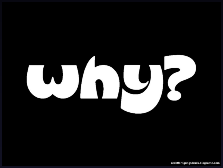
Introduction to Machine Translation

Gideon Maillette de Buy Wenniger
Supervisor: Khalil Sima'an
website : http://staff.science.uva.nl/~gemaille/

Statistical Language Processing and Learning Lab
Institute for Logic Language and Computation
University of Amsterdam, the Netherlands

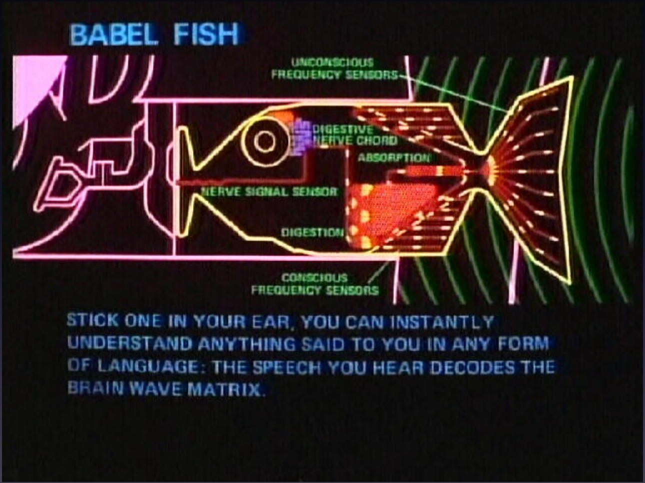
February 22th, 2012

Why Machine Translation?



- Reliable translators may not be available for all languages
- Translation demand > translator capacity
- Large volume translation (i.e. internet) can only be tackled with automatic translation
- Cheapness makes new things affordable
- Imperfect, often sufficient, saves over fully manual translation
- It just rocks, need I say more? ☺

The ultimate Translation System?

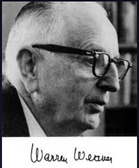


Part 0 : A very short historical background

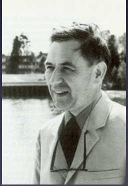
MT by Selected events (1/3)

The early days of MT

- *1933 Kickoff MT*
  Patents Georges Artsrouni (France) and Petr Trojanskij (Russia)
  - Artsrouni : general-purpose machine / mechanical multilingual dictionary
  - Trojanskij : mechanical dictionary / encoding and interpretation gramatical funtions



- *1946 and 1947 Proposal : Computer based translation*
  Andrew Booth (a British crystallographer) and Warren Weaver (US scientist, mathematician, science administrator).



- *1951 and 1952 First MT Conference*
  Yehoshua Bar-Hillel (MIT). Interviews + Report : State-of-the-art basic approaches MT. June 1952 first MT conference.

MT by Selected events (2/3)

Extreme highes and extreme lows

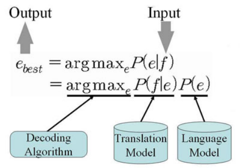
- *1954 MT goldrush*
  Léon Dostert,Paul Garvin (Georgetown University) + Peter Sheridan (IBM) public demonstration of the feasibility of MT
  Toy example selected sample of 49 Russian sentences 250 words and just 6 grammar rules. Lots of media attention, major attraction research funding.

- *1966 ALPAC Report: MT deathsentence ☹*
  Automatic Language Processing Advisory Committee (ALPAC) advised on US funding MT research. “there is no immediate or predictable prospect of useful machine translation” (ALPAC 1966). (but some continued anyway ☺)

MT by Selected events (3/3)

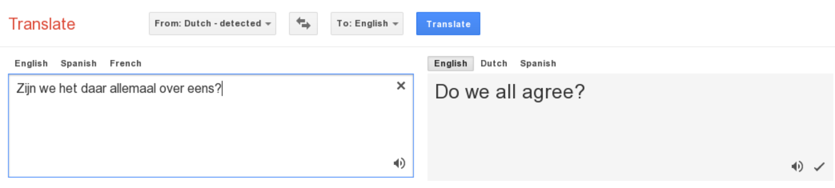
Age of Statistical MT

- *1988 Start age statistical MT - the IBM gang*
  Brown et. al, "A statistical approach to language translation."
- *2002 Automated evaluation metrics - start BLEU period*
  "BLEU: a Method for Automatic Evaluation of Machine Translation", Papineni et. al
- *2004 Phrase based translation*
  "The Alignment Template Approach to Statistical Machine Translation", F. Och H. Ney
- *2005 Hierarchical phrase-based SMT* "A hierarchical phrase-based model for statistical machine translation." D. Chiang

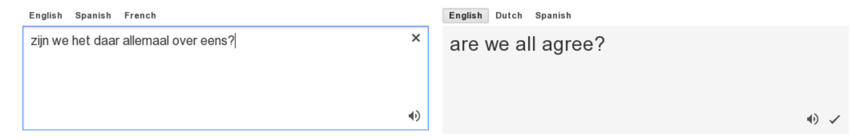


Part 1 : The complexities of translation

Is translation difficult?



(c) Example sentence in standard format ☺



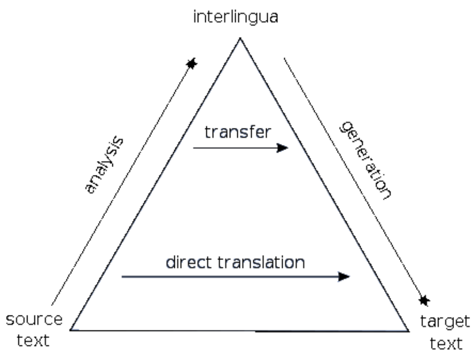
(d) Example sentence slightly changed ☹

Is translation difficult (continued)?

(e) Example sentence in standard format ☺

(f) Example sentence slightly changed ☹

Intermezzo 1 : The impertinent (Translation) pyramid



(g) Translation pyramid

(h) Other contestant

Intermezzo 2 : Links with core areas of AI

- **Modelling** :
  constructing the translation model. Links to statistics, linguistics, logic, discrete mathematics etc,
- **Parameter optimization / Learning** :
  Link to huge optimization and Machine Learning (also datamining) fields
- **Decoding/Search** :
  Links to important AI field of search methods, information retrieval, databases, distributed computing

Why is translation difficult?

- **Data challenges**
- **Inherent ambiguity of language**
- **Problems with compositionality**
- **Structure/Modelling challenges**
- **Computational complexity challenges**

Data challenges

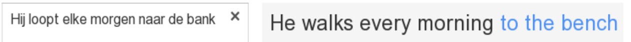
- Language is combinatoric ⇒ infinite number of sentences
- Finite training material available ⇒ Generalization required

Parallel Corpus (L1-L2)	Sentences	L1 Words	English Words
Danish-English	1,785,775	46,102,455	48,833,481
German-English	1,739,154	45,607,269	47,978,832
Dutch-English	1,822,036	50,315,412	49,938,127
Spanish-English	1,786,594	51,551,485	49,411,045
Estonian-English	469,622	9,318,986	12,452,336
Finnish-English	1,742,553	34,123,013	47,601,416
French-English	1,825,077	54,568,499	50,551,047

Figure: Corpus sizes for some of the Europarl language pairs

Inherent ambiguity of language

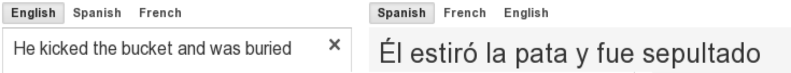
(a) Interpretation 1



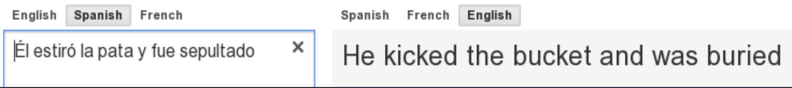
(b) Interpretation 2

Problems with compositionality (1/2)

- Translation not completely compositional



(c) Idiomatic transltation English-Spanish ☺



(d) Idiomatic transltation English-Spanish ☺

(e) Idiomatic transltation breaks for English-Dutch ☹

Problems with compositionality (2/2)

- Common assumption: limited length translation pieces
- Keeps translation efficient and number of rules manageable
- But ... words that translate together not always close

Hij legde het wapen heel langzaam weg | He put the weapon away very slowly

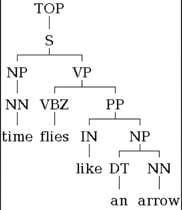
(f) Limited length fragment compositionality still works

Hij legde het wapen heel langzaam maar toch zeker en zonder al teveel twijfel weg | He put the gun slowly but surely and without much doubt

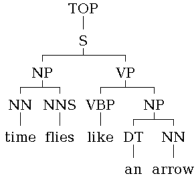
(g) Limited length fragment compositionality breaks

Structure/Modelling challenges (1/2)

- Structure and meaninging are hidden within sentences



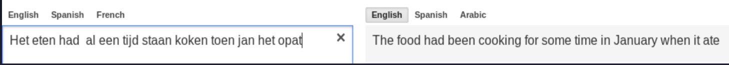
(h) Time goes very fast



(i) Preferences of "time flies"

Structure/Modelling challenges (2/2)

- Argument structure is essential to meaning but implicit in sentence



(j) Argument structure gone wrong

Computational complexity challenges

- A too complex model makes searching the space of possible translations infeasible
- More complex models make it impossible to efficiently combine evidence contributing to same translations
- Translation models/grammars may become huge. Filtering becomes nescessary but may be costly to do on a per-sentence level.

Part 2 : Translation models

Do we need to use models?

- Without model assumptions we can only interpolate the data we have and not really generalize
- Without generalization no prediction of future data
- There is an inherent trade-off between accurarcy of estimating the model we assume (variance) and the aggresiveness of the simplifying assumptions we make (bias).

Illustration: Bias-Variance trade-off

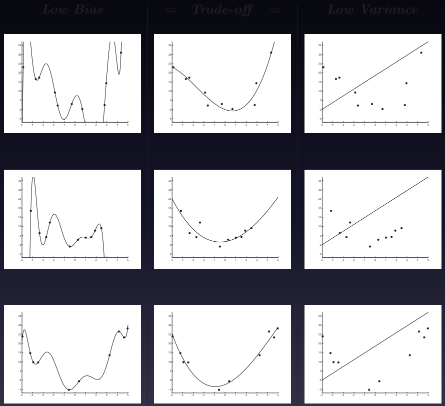


Figure: Trade-off bias/variance. To generalize robustly from data a limitation of the model complexity is required. [1]

[1] Source: Learning the Latent Structure of Transltation. PhD thesis. Markos Mylonakis, 2012

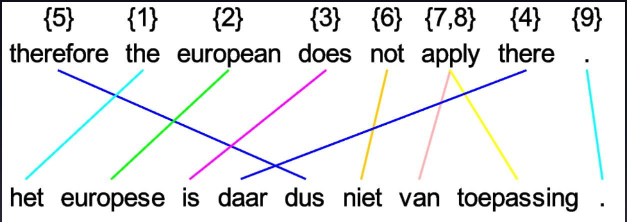
On the pervasiveness of hierarchy

“Scientific knowledge is organized in levels, not because reduction is impossible but because nature is organized in levels, and the pattern at each level is most clearly discerned by abstracting from the detail to the levels far below.
(The pattern of a halftone does not become clearer when we magnify it so the individual spots of ink become visible.)
And nature is organized in levels because hierarchic structure - systems of Chinese boxes - provide the most viable form for any system of even moderate complexity.”

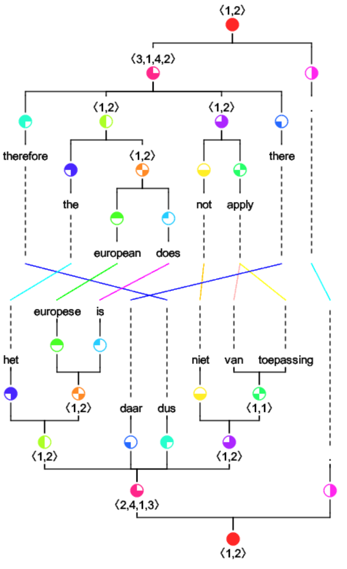
Herbert A.Simon (1973). The organization of Complex Systems.

Part 3 : Some own research in 2 slides

Word Alignments



Hierarchical Alignment Trees



Acknowledgements/Sources

- *“Machine Translation: A Concise History”* W. John Hutchins
- *Supervision + some material* Khalil Sima’an
- *“Statistical Machine Translation“ (tutorial)* Adam Lopez.
- *Bias-Variance trade-off* Markos Mylonakis.
- *Beamer Presentation theme* David Carlisle, Shawn Lankton. http://www.shawnlankton.com/2008/02/beamer-and-latex-with-keynote-theme/
- *Stage time* Raquel Fernández Rovira

Summary

- History MT
- MT Pyramid
- Links with core areas AI
- What makes MT difficult
- Translation Models
- Hiearchy
- Alignments and HATs

That's it...

Questions?