

Universitat Politècnica de València

Master en Inteligencia Artificial, Reconocimiento de Formas e Imagen Digital

Master in Artificial Intelligence, Pattern Recognition and Digital Imaging

2022-2023

MACHINE TRANSLATION

1. Introduction to Machine Translation

Francisco Casacuberta

`fcn@prhlt.upv.es`

November 8, 2022

Index

- 1 Introduction ▷ 2
- 2 Approaches to MT ▷ 10
- 3 Brief history of MT and MT systems ▷ 14
- 4 Text-input machine translation ▷ 19
- 5 Speech-input machine translation ▷ 28
- 6 Computer-assisted translation ▷ 32
- 7 Notation ▷ 35
- 8 Bibliography ▷ 37

Index

- 1 *Introduction* ▷ 2
- 2 Approaches to MT ▷ 10
- 3 Brief history of MT and MT systems ▷ 14
- 4 Text-input machine translation ▷ 19
- 5 Speech-input machine translation ▷ 28
- 6 Computer-assisted translation ▷ 32
- 7 Notation ▷ 35
- 8 Bibliography ▷ 37

Translation and Machine Translation

Translation:

The process of translating words or text from one language into another ¹.

Machine translation:

Translation carried out by a computer ¹.

¹ English Oxford Dictionaries

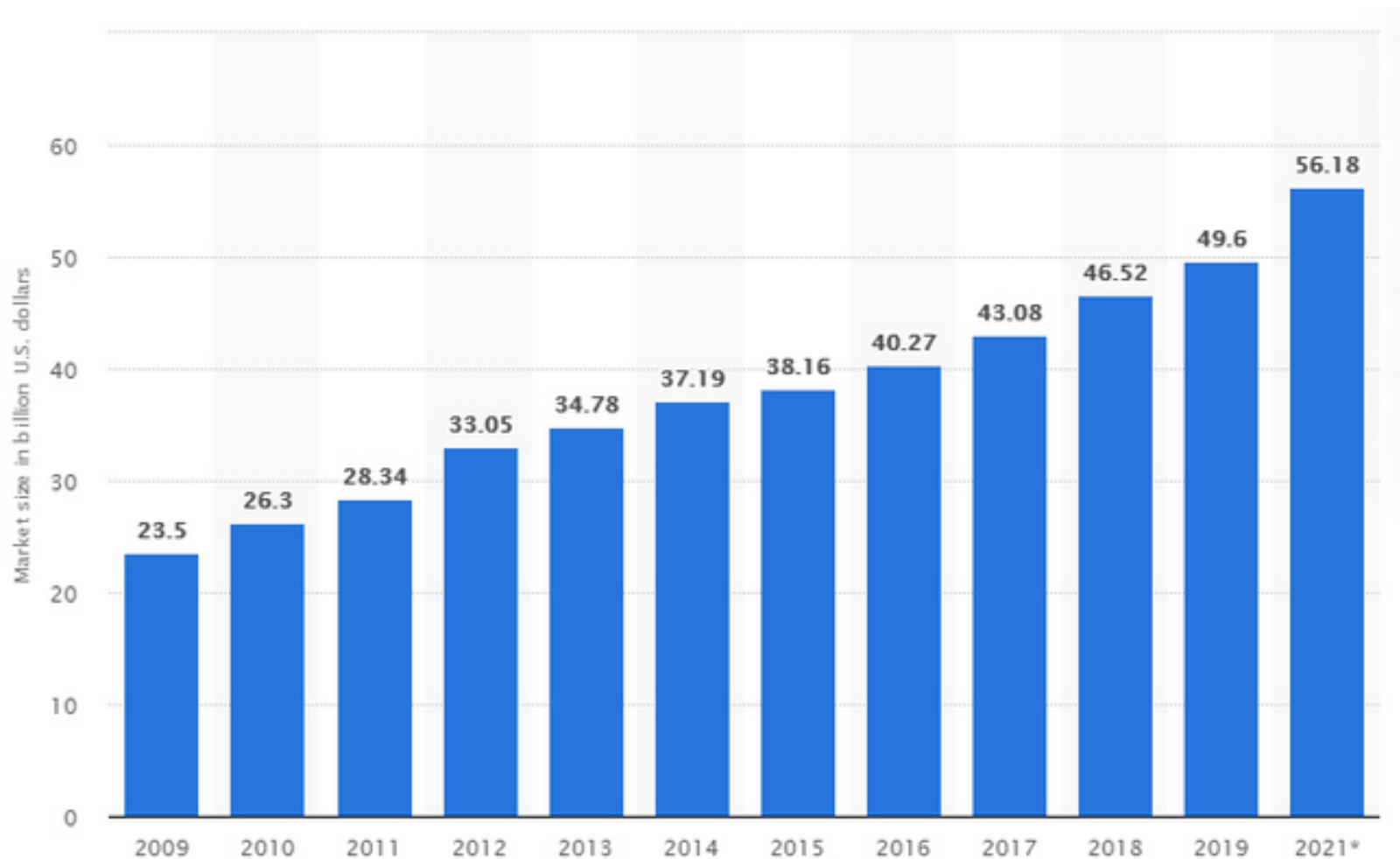
Translation Industry Statistics¹

- 7,139 languages being spoken today.
- Approximately 42% of languages are endangered.
- English is the most spoken language, while Mandarin Chinese has the most native speakers.
- The most widely used language on the internet is English, the second is Chinese, and third is Spanish at 8%.

¹Lim 2022 Translation Industry Trends and Stats

<https://redokun.com/blog/translation-statistics>

Translation Industry Statistics¹



© Statista 2021

¹Lim 2022 Translation Industry Trends and Stats

<https://redokun.com/blog/translation-statistics>

Translation Industry Statistics¹

- Europe at 49% of language services market, North America at 39.4%.
- The translation service market size grew by 40% during COVID-19.
- The machine translation market:
 - USD 650 million in 2020.
 - Expected to reach USD 3 billion by 2027.
- Increasing of demand in healthcare, medical industries, social networking, ...

¹Lim 2022 Translation Industry Trends and Stats

<https://redokun.com/blog/translation-statistics>

Translation Industry Statistics¹

- 640,000 human translators worldwide,
- 88% of full-time professional translators use at least 1 CAT tool.
- Translators estimate that using translation software can boost their productivity by at least 30%.
- The percentage of projects for end-clients using machine translations climbed from 13% in 2019 to 24% in 2020.

¹Lim 2022 Translation Industry Trends and Stats

<https://redokun.com/blog/translation-statistics>

MT objectives: Facts

- MT is useful.
- There are many situations that MT systems produce reliable, if less than perfect, translations at high speed. In some circumstances, MT systems can produce good quality outputs.

The first book translated by a MT system to French (12/9/2018): “Deep Learning” by Goodfellow, Bengio & Courville.

<https://www.usinenouvelle.com/editorial/le-premier-livre-traduit-par-une-ia-est-un-manuel-de-deep-learning.N740664>

- Human parity and Super-Human Performance in MT [Toral 2020]
<https://www.youtube.com/watch?v=wMhyEiZUCY4&t=1404s> (From 15:45 min)
- MT does not threaten translators' jobs: High demand of translations and too repetitive translation jobs.
- Speech-to-speech MT is still a research topic, but ...
- There are many open research problems in MT.

Need of pre/post-editing

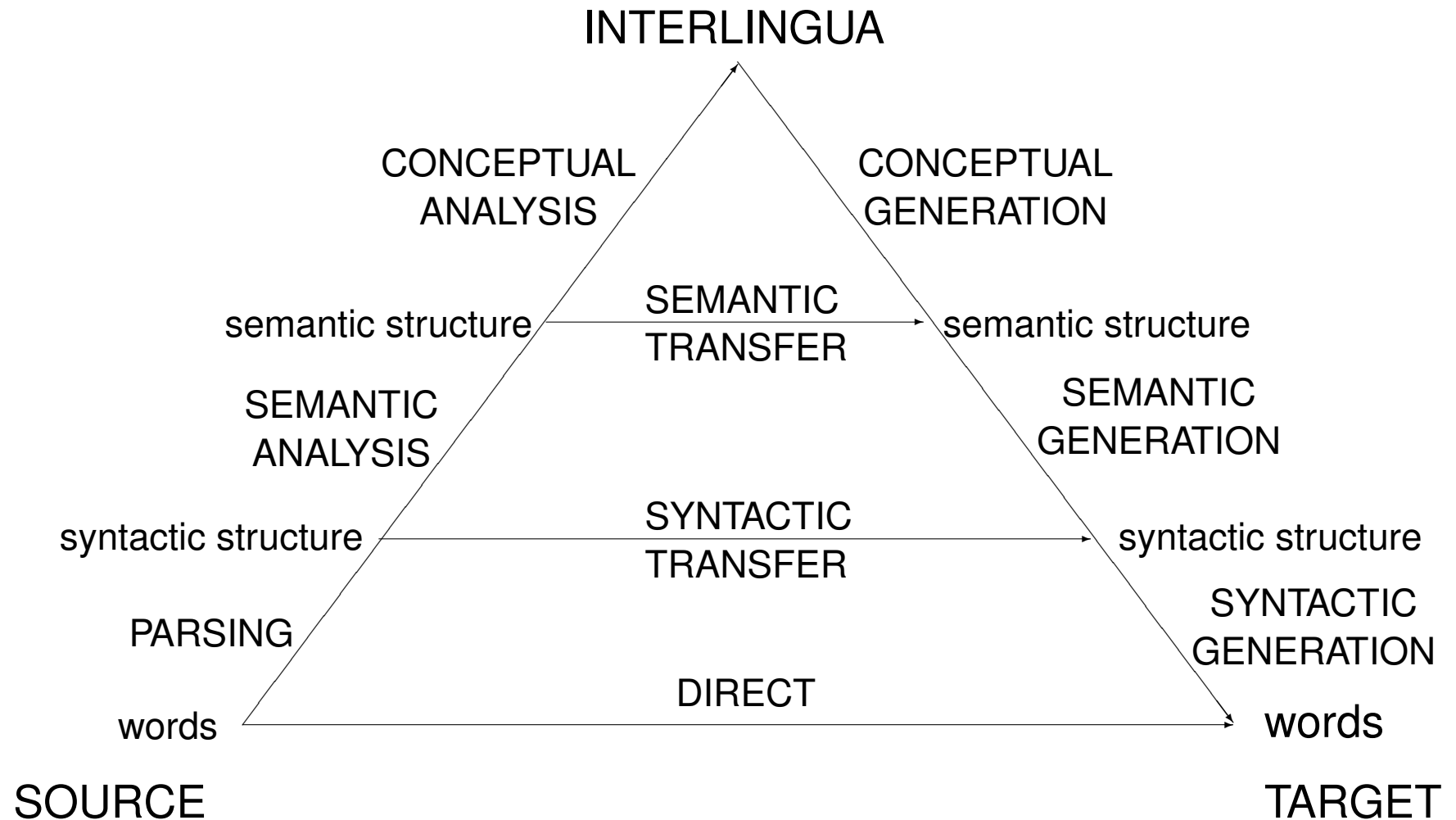
- While the number of errors and bad constructions is high, “post-editing” can make the result useful.
- Many problems could have been avoided by making the source text “simpler”.



Index

- 1 Introduction ▷ 2
- 2 *Approaches to MT* ▷ 10
- 3 Brief history of MT and MT systems ▷ 14
- 4 Text-input machine translation ▷ 19
- 5 Speech-input machine translation ▷ 28
- 6 Computer-assisted translation ▷ 32
- 7 Notation ▷ 35
- 8 Bibliography ▷ 37

Approaches to MT: Vauquois triangle



Approaches to MT: Technologies

- (Linguistic) knowledge-based methods
- (Memorized) example-based methods
 - Translation memories
- Statistical models
 - Alignment models
 - Finite-state models
 - **(Deep) Neural networks**
- Hybrid models

Assessment

- Manual from test sentences:
 - Human evaluation (expensive)
- Automatic from test sentences with reference translation:
 - Editing Distances: **Translation Error Rate (TER)**
 - Multi-reference TWER.
 - N-gram based: **BLEU** and NIST scores.
 - Meteor
 - **Beer**
 - ...
- Automatic from test sentences with reference translation for CAT:
 - Word-Stroke Ratio (WSR).
 - Key-stroke ratio (KSR).
 - Mouse Action Ratio (MAR)
 - ...

Index

- 1 Introduction ▷ 2
- 2 Approaches to MT ▷ 10
- 3 *Brief history of MT and MT systems* ▷ 14
- 4 Text-input machine translation ▷ 19
- 5 Speech-input machine translation ▷ 28
- 6 Computer-assisted translation ▷ 32
- 7 Notation ▷ 35
- 8 Bibliography ▷ 37

Brief history of MT

- **1949** Weaver: Information-theory based approach.
- **1957** Chomsky: Natural language is not governed by statistics.
- **1960** ALPAC (Automatic Language Processing Advisory Committee) report: No useful MT results are foreseen.
- **1960-2000**
 - SYSTRAN system: based on dictionaries.
 - Several (linguistic) knowledge-based approaches.
- **1989-95** “Empiricists” methods are introduced: corpus-based and statistical approaches (IBM, 1989)
- **1995-nowadays** “Empiricists” methods are thriving. Speech-to-speech MT in limited domains.

Recent history of MT: “Empiricists” methods

- **1989-1995** Statistical approach to MT by IBM Yorktown Heights researchers:
 - Hansards Parallel English/French transcriptions of parliamentary discussions.
 - DARPA competitive assessment (1994): Results comparable to those achieved by traditional approaches.
- **1995-2010** Development of statistical techniques and other empiricists methods:
 - Great progress on the statistical approach.
 - Other empiricist techniques: Memory-based, finite-state, etc.
 - Computer assisted translation.
- **2010-2015** Practical use of statistical machine translation and a new approach:
 - Industrial use of Moses.
 - **Continuous space representation: Neural machine translation.**
 - Adaptation, online learning, multimodal, multilingual, data filtering, ...
 - Advanced computer assisted translation: interactive, multimodal, adaptative, ...
- **2016-** Practical use of neural machine translation:
 - Industrial use of NMT.
 - New neural models: **Transformer.**
 - Topics: learning from monolingual corpora, low-resource languages, **pre-trained models for NLP, prompting, ...**

Machine translation systems

- Text translation systems

- **GoogleTrans:** http://www.google.com/language_tools
- **DeepL:** <https://www.deepl.com/en/translator>
- **SYSTRAN:** <https://www.systran.net/en/translate/>
- **Microsoft Translator:** <https://www.bing.com/translator>
- **Amazon Translate:** <https://aws.amazon.com/es/getting-started/hands-translate-text-between-languages-cloud/>
- **SisHiTra:** <https://demosmt.prhlt.upv.es/sishitra/>
- **Apertium:** <http://www.apertium.org/>

Machine translation systems

- Computer assisted (aided) translation systems

- **TRANSTYPE:**

- <http://rali.iro.umontreal.ca/rali/?q=en/node/1282>

- **Déjàvu:**

- <http://www.atril.com/>

- **TRADOS:**

- <https://www.sdltrados.com/products/>

- **CasMaCat:**

- <https://demosmt.prhlt.upv.es/matecat-test/translate/demo-eutt2.xliff/es/1-fmttqpvc/demo@4002#1>

- **Neural Machine Translation**

- <https://demosmt.prhlt.upv.es/inmt/>

Index

- 1 Introduction ▷ 2
- 2 Approaches to MT ▷ 10
- 3 Brief history of MT and MT systems ▷ 14
- 4 *Text-input machine translation* ▷ 19
- 5 Speech-input machine translation ▷ 28
- 6 Computer-assisted translation ▷ 32
- 7 Notation ▷ 35
- 8 Bibliography ▷ 37

General Framework

- Every sentence y from a *target* language is considered as a possible translation of a given sentence x from a *source* language.
- For each possible pair of sentences y, x , there is a probability $\Pr(y | x)$.
- $\Pr(y | x)$ should be *low* for pairs (y, x) such as:
(*una habitación con vistas al mar , are all expenses included in the bill ?*)
- $\Pr(y | x)$ should be *high* for pairs such as:
(*¿ hay alguna habitación tranquila libre ? , is there a quiet room available ?*)
- Goal of statistical machine translation: given a source sentence x search for a target sentence \hat{y} such that (**Risk minimization**):

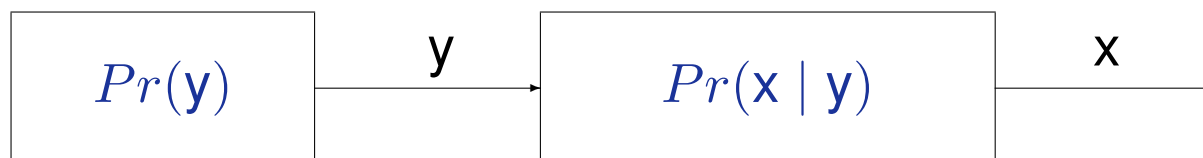
$$\hat{y} = \operatorname{argmax}_y \Pr(y | x)$$

An inverse (generative) approach

Decompose $\Pr(y \mid x)$ using Bayes' rule:

$$\hat{y} = \operatorname{argmax}_y \Pr(y \mid x) = \operatorname{argmax}_y \frac{\Pr(x \mid y) \cdot \Pr(y)}{\Pr(x)} = \operatorname{argmax}_y \Pr(x \mid y) \cdot \Pr(y)$$

A “distorted (noisy) channel model”



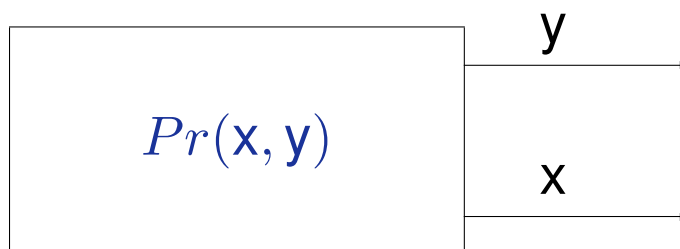
Need: a target-language model + alignment and lexicon models

A generative (finite-state) approach

The direct probability can be decomposed in a different way:

$$\hat{y} = \operatorname{argmax}_y \Pr(y \mid x) = \operatorname{argmax}_y \frac{\Pr(x, y)}{\Pr(x)} = \operatorname{argmax}_y \Pr(x, y)$$

A “joint” model



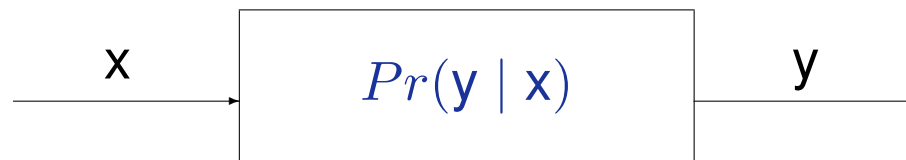
A stochastic finite-state transducer can model the joint distribution

A direct (discriminative) approach

Search for a target sentence with maximum *posterior* probability:

$$\hat{y} = \operatorname{argmax}_y \Pr(y \mid x)$$

A “direct model”



Log-linear combination of models

Encoding-Decoding models

Log-linear combination of models

Search for a target sentence with maximum *posterior* probability:

$$\hat{y} = \operatorname{argmax}_y \Pr(y \mid x)$$

$$\hat{y} = \operatorname{argmax}_y \frac{\exp \left(\sum_{k=1}^K \lambda_k h_k(x, y) \right)}{\sum_{y'} \exp \left(\sum_{k=1}^K \lambda_k h_k(x, y') \right)} = \operatorname{argmax}_y \sum_{k=1}^K \lambda_k h_k(x, y)$$

- $h_1(x, y) = \log \Pr(y)$, a language model
- $h_2(x, y) = \log \Pr(y \mid x)$, a translation model model
- $h_3(x, y) = \log \Pr(x \mid y)$, an inverse translation model model
- ...

Encoder-decoder models

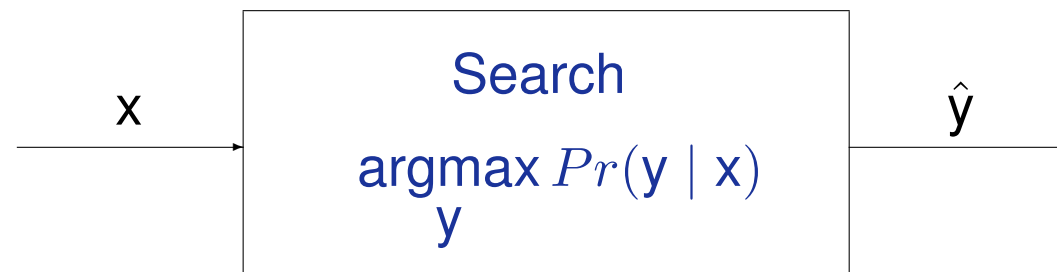
Search for a target sentence with maximum *posterior* probability:

$$\hat{y} = \operatorname{argmax}_y \Pr(y \mid x)$$

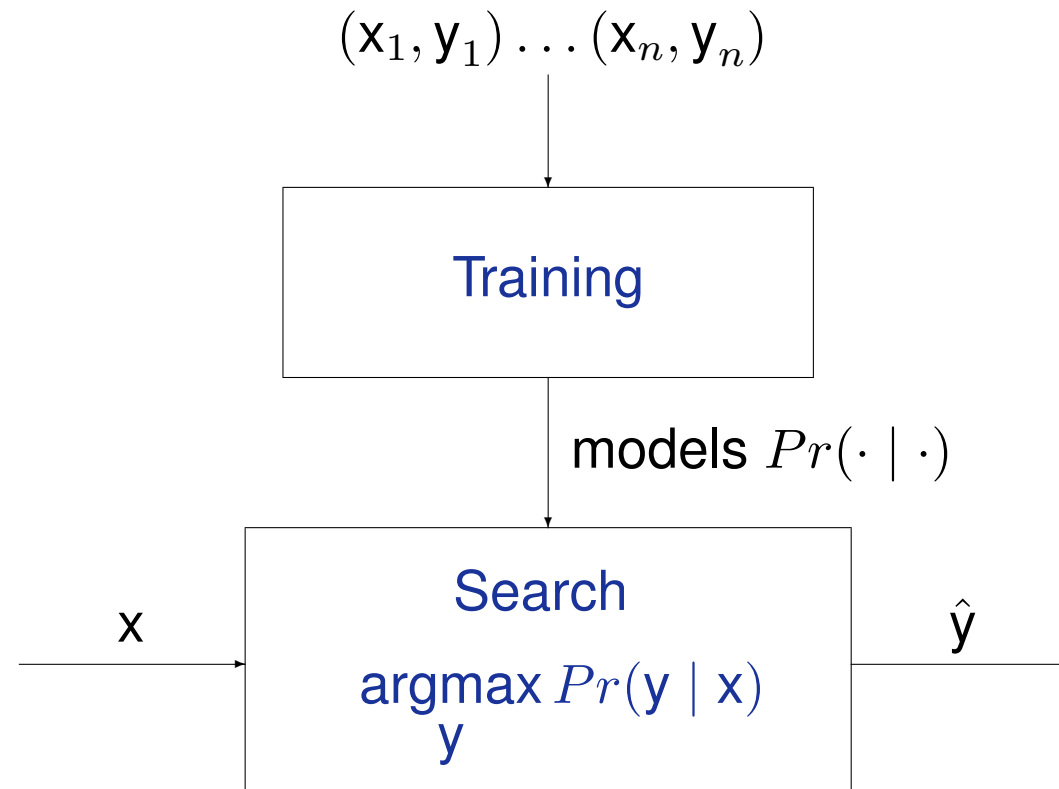
$$\hat{y} = \operatorname{argmax}_y \prod_{i=1}^I P(y_i \mid y_1^{i-1}, a(x))$$

- Recurrent neuronal networks with an attention model.
- Full attention non-recurrent model: Transformer.
- A convolutional neuronal network.

Training and search



Training and search



Index

- 1 Introduction ▷ 2
- 2 Approaches to MT ▷ 10
- 3 Brief history of MT and MT systems ▷ 14
- 4 Text-input machine translation ▷ 19
- 5 *Speech-input machine translation* ▷ 28
- 6 Computer-assisted translation ▷ 32
- 7 Notation ▷ 35
- 8 Bibliography ▷ 37

Speech-input translation

Given an input acoustic sequence \mathbf{v} , search for a target sentence with maximum posterior probability:

$$\hat{y} = \operatorname{argmax}_y \Pr(y \mid \mathbf{v})$$

But this can be seen as a “*two-step process*”:

$$\mathbf{v} \longrightarrow \mathbf{x} \longrightarrow y$$

where the “*hidden variable*” \mathbf{x} accounts for all possible input decodings of \mathbf{v} :

$$\hat{y} = \operatorname{argmax}_y \sum_{\mathbf{x}} \Pr(y, \mathbf{x} \mid \mathbf{v}) = \operatorname{argmax}_y \sum_{\mathbf{x}} \Pr(\mathbf{x}, y) \cdot \Pr(\mathbf{v} \mid \mathbf{x})$$

(with the assumption: $\Pr(\mathbf{v} \mid \mathbf{x}, y)$ does not depend on the target sentence y)

Speech-input translation: Integrated approaches

$$\operatorname{argmax}_y \Pr(y \mid \mathbf{v}) \approx \operatorname{argmax}_y \max_x (\Pr(x, y) \cdot \Pr(\mathbf{v} \mid x))$$

- $\Pr(\mathbf{v} \mid x) \approx$ ACOUSTIC MODELS
- $\Pr(x, y) \approx$ STOCHASTIC FINITE-STATE TRANSDUCERS

$$\operatorname{argmax}_y \Pr(y \mid \mathbf{v}) \approx \operatorname{argmax}_y \prod_{i=1}^I P(y_i \mid y_1^{i-1}, a(\mathbf{v}))$$

- ENCODER-DECODER MODELS

Speech-input translation: A serial approach

$$\operatorname{argmax}_y \max_x (\Pr(x, y) \cdot \Pr(v \mid x)) = \operatorname{argmax}_y \max_x (\Pr(y \mid x) \cdot \Pr(x) \cdot \Pr(v \mid x))$$

1. $\hat{x} \approx \operatorname{argmax}_x \Pr(x) \cdot \Pr(v \mid x) = \operatorname{argmax}_x \Pr(x \mid v)$
2. $\hat{y} \approx \operatorname{argmax}_y \Pr(y \mid \hat{x})$

- $\Pr(v \mid x) \approx$ ACOUSTIC MODELS
- $\Pr(y \mid x) \approx$ TRANSLATION MODELS
- $\Pr(x) \approx$ SOURCE LANGUAGE MODELS

Index

- 1 Introduction ▷ 2
- 2 Approaches to MT ▷ 10
- 3 Brief history of MT and MT systems ▷ 14
- 4 Text-input machine translation ▷ 19
- 5 Speech-input machine translation ▷ 28
- 6 *Computer-assisted translation* ▷ 32
- 7 Notation ▷ 35
- 8 Bibliography ▷ 37

Text prediction for Interactive Machine Translation in CAT

Given a source text x and a “correct” *prefix* y_p of the target text, search for a *suffix* \hat{y}_s , that maximizes the posterior probability over all possible suffixes:

$$\hat{y}_s = \operatorname{argmax}_{y_s} \Pr(y_s \mid x, y_p)$$

Taking into account that $\Pr(x, y_p)$ does not depend on y_s , we can write:

$$\begin{aligned} \hat{y}_s &= \operatorname{argmax}_{y_s} \frac{\Pr(y_p, y_s \mid x)}{\Pr(y_p \mid x)} \\ &= \operatorname{argmax}_{y_s} \Pr(y_p y_s \mid x) \end{aligned}$$

Main difference with text-input machine translation: **search over the set of suffixes.**

Extension: Fix all correct segments,

Target language dictation in CAT

A *human* translator *dictates* the translation of a source text, x , producing a *target language* acoustic sequence v .

Given v and x , the system should search for a most likely decoding of v :

$$\hat{y} = \operatorname{argmax}_y \Pr(y \mid x, v)$$

By the assumption that $\Pr(v \mid x, y)$ does not depend on x ,

$$\hat{y} = \operatorname{argmax}_y \Pr(v \mid y) \cdot \Pr(x \mid y) \cdot \Pr(y)$$

- $\Pr(v \mid y) \approx$ (TARGET LANGUAGE) ACOUSTIC MODELS
- $\Pr(x \mid y) \approx$ TRANSLATION MODEL
- $\Pr(y) \approx$ TARGET LANGUAGE MODEL

Similar to plain speech decoding, where: $\hat{y} = \operatorname{argmax}_y \Pr(v \mid y) \cdot \Pr(y)$

Index

- 1 Introduction ▷ 2
- 2 Approaches to MT ▷ 10
- 3 Brief history of MT and MT systems ▷ 14
- 4 Text-input machine translation ▷ 19
- 5 Speech-input machine translation ▷ 28
- 6 Computer-assisted translation ▷ 32
- 7 *Notation* ▷ 35
- 8 Bibliography ▷ 37

Notation and basic concepts

- x and y will generally denote **source** and **target** texts, respectively
- **Unconditional probabilities:** $\Pr(Y = y) \equiv \Pr(y)$
- **Conditional probabilities:** $\Pr(Y = y \mid X = x) \equiv \Pr(y \mid x)$
- **Joint probability:** $\Pr(x, y) = \Pr(x) \cdot \Pr(y \mid x)$
- **Bayes' rule:** $\Pr(x \mid y) \cdot \Pr(y) = \Pr(y \mid x) \cdot \Pr(x)$
- **Chain rule:** $\Pr(x_1^I) = \Pr(x_1) \cdot \Pr(x_2 \mid x_1) \cdots \Pr(x_I \mid x_1^{I-1})$
- **Marginal distribution:** $\Pr(x) = \sum_{\forall y} \Pr(x, y)$
- **Maximum approximation:** $\sum_{x \in A} \Pr(x) \approx \max_{x \in A} \Pr(x)$
- **Expected values** $E_P(f(x)) \equiv E_x(f(x)) = \sum_x f(x) P(x)$

Index

- 1 Introduction ▷ 2
- 2 Approaches to MT ▷ 10
- 3 Brief history of MT and MT systems ▷ 14
- 4 Text-input machine translation ▷ 19
- 5 Speech-input machine translation ▷ 28
- 6 Computer-assisted translation ▷ 32
- 7 Notation ▷ 35
- 8 *Bibliography* ▷ 37

Bibliography

1. P. Brown, J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, J. Lafferty, R. Mercer, P. Roosin: *A statistical approach to machine translation*. Computational Linguistics, 16:79–85, 1990.
2. P. Brown, S. Chen, S. Della Pietra, V. Della Pietra, A. Kehler, R. Mercer: *Automatic speech recognition in machine aided translation*. Computer Speech and Language, 8(3): 177–87. 1994.
3. F. Casacuberta, H. Ney, F. J. Och, E. Vidal, J. M. Vilar, S. Barrachina, I. García-Varea, D. Llorens, C. Martínez, S. Molau, F. Nevado, M. Pastor, D. Picó, A. Sanchis, Ch. Tillmann: *Statistical and finite-state approaches to speech-to-speech translation*. Computer Speech and Language, 18:25-47, 2004.
4. P. Koehn. *Statistical Machine Translation*. Cambridge University Press. 2010.
5. P. Langlais, G. Foster, G. Lapalme; *Unit Completion for a Computer-aided Translation Typing System*. Machine Translation, 15(4): 267–294, 2000.
6. A. Lopez. *Statistical Machine Translation*. ACM Computing Surveys, 40(3):1–49, 2008
7. E. Vidal, F. Casacuberta, L. Rodríguez, J. Civera and C. Martínez *Computer-assisted translation using speech recognition*. IEEE Transactions on Speech and Audio Processing, 14(3):941-951, 2006.
8. S. Barrachina, O. Bender, F. Casacuberta, J. Civera, E. Cubel, S. Khadivi, A. Lagarda H. Ney, J. Tomás, and E. Vidal. *Statistical approaches to computer-assisted translation*. Computational Linguistics, 35(1):3-28, 2009.
9. Francisco Casacuberta, Marcello Federico, Hermann Ney, Enrique Vidal. Recent Efforts in Spoken Language. Translation. IEEE Signal Processing Magazine, 2008. Vol. 25 (3), pp. 70-79.