CIT CORK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

Computer Science

http://www.cit.ie

Natural Language Processing

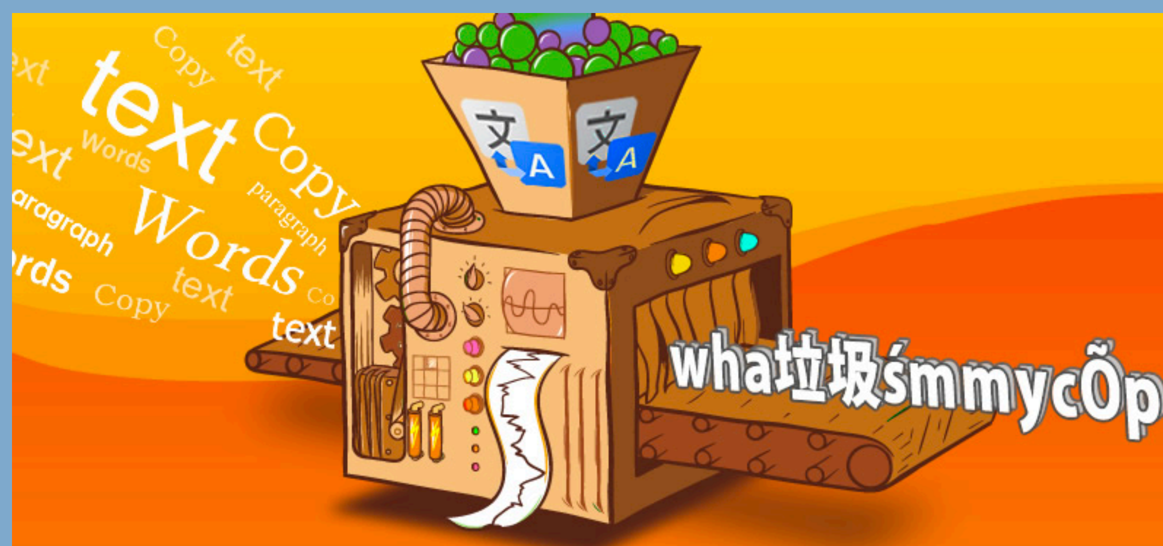
Week7: Machine Translation

Dr. Haithem Afli

Haithem. afli@cit.ie

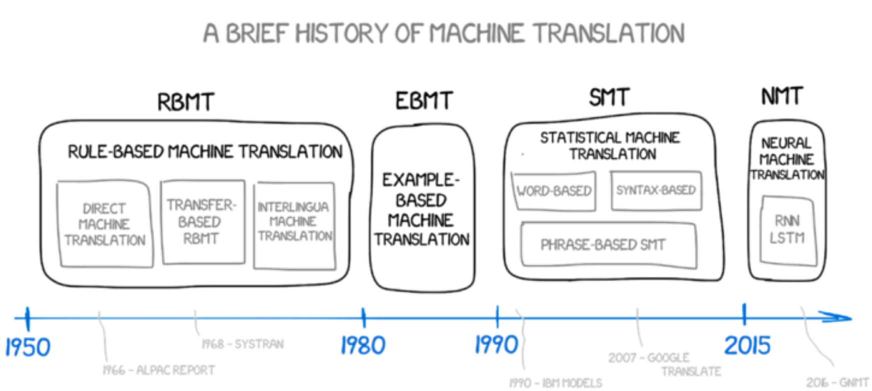
@AfliHaithem

2020/2021



Overview

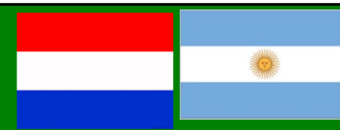
- Why MT?
- How are people using MT?
- What's the Role of the Human Translator?
- Why Corpus-Based MT?
- How might you go about translating languages you have no knowledge of?
- The Importance of Data



Thanks to Prof Andy Way

Volume

The 2014 FIFA World Cup was the biggest event yet for Twitter with **672 million tweets**

Requested translation from Twitter (words)				Grand Total from all World Cup matches
	6,459,830	5,141,360	4,847,590	**85,047,110**

Top 3 languages

English

Portuguese

Spanish

- Source→Target traffic:
- EN→ES 13,614,450 (EN to all languages: 50,545,460)
- ES→EN 5,569,200 (ES to all languages: 10,609,420)
- PT→EN 1,831,750 (PT to all languages: 4,230,880)

Volume

The 2014 FIFA World Cup was the biggest event yet for Twitter with **672 million tweets**

Requested translation from Twitter (words)			Grand Total from all World Cup matches
	6,459		**85,047,110**

85,047,110 total words = 2,835,000 words per day (30 days)

equivalent to 1,134 human translators working full-time for 30 days

Top 3 languages ... ish

- Source→Targe
- EN→ES 13, ... ages: 50,545,460)
- ES→EN 5, ... all languages: 10,609,420)
- PT→EN 1, ... 50 (PT to all languages: 4,230,880)

13 languages and 24 language pairs

Irish	Spanish	Korean
German	Italian	Farsi
French	Portuguese	Greek
Croatian	Japanese	Chinese

↔ English

24 MT systems trained on Microsoft slator Hub and used by

SIZE MACHINE TRANSLATION MARKET IS $250 MILLION – TAUS PUBLISHES NEW MARKET REPORT

Mission
History
Services
People
Team
Advisors
Representatives
Partners
Newsroom
Press releases
Newsletters
Working at TAUS
Contact

August 26, 2014, Amsterdam - TAUS estimates the size of the machine translation market at $250 million in its newly published machine translation market report. The 60 page report offers a detailed overview of all facets of the market machine translation with sections on the different types of offerings, the players, open-source systems, challenges, opportunities and trends.

"*The size of the MT market is relatively small compared to its innovation power and impact*", says Jaap van der Meer, one of the co-authors of the report. "*MT technology is a key enabler and a force multiplier for new services and growth. MT technology finds a high adoption rate among language service providers. Innovative companies in information technology and other sectors are converging MT technology in new applications and products or they use MT to enhance their existing products.*"

For this market report TAUS has identified 65 different MT operators. More than 80 companies responded to the surveys and the TAUS team interviewed 37 users and developers of MT. The largest MT providers in alphabetical order are: CSLi, Google, IBM, LionBridge, Microsoft, PROMT, Raytheon BBN, SDL, Smart Communications, SYSTRAN. The MT market is a vibrant sector with new companies entering the market place and long-term players being acquired. Around 20 of the 65 identified MT players started business in the last five years. At the start of this year SYSTRAN was acquired by CSLi from South Korea and AppTek was acquired by eBay.

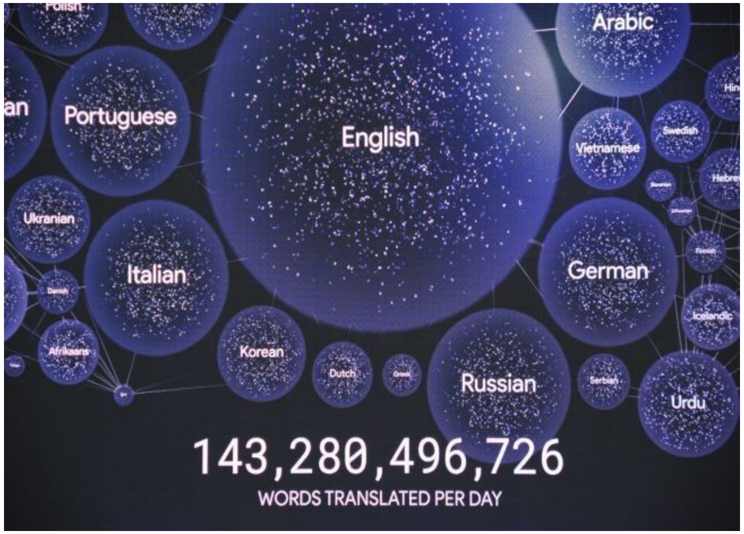
"*The dynamics in the MT market have changed dramatically in the last five years*", says Achim Ruopp, product development manager at TAUS and co-author of the report. "*The increased availability of easy to use and integrate MT with sufficient quality has ignited the emergence of new business models. This has been promoted by many new MT suppliers that base their offering on the open source statistical MT system Moses. A bigger impact still has come from some of the internet giants like Google, Microsoft and Yandex that offer free or very cheap MT services.*"

The TAUS Machine Translation Market Report is available on the TAUS website.

Highlights from TAUS Claims

- Size of the market: €250,000,000
- "MT technology is a key enabler and a force multiplier for new services"
- "Innovative companies in IT and other sectors are converging MT technology in new applications and products or they use MT to enhance their existing products"
- "The increased availability of easy to use and integrate MT with sufficient quality has ignited the emergence of new business models. This has been promoted by many new MT suppliers that base their offering on the open source statistical MT system Moses"

MT is being used every day...



- Google Translate provides a billion translations a day for 200 million users
- Amount of text translated daily is more than what's in a million books
- Surpasses what professional translators handle in a year

Client-customised engines

- Improve productivity,
- Translate content previously not feasible due to time or cost constraints,
- Reduce time to market, and
- Reduce translation costs.

Lots of successful case studies

- Adobe & ProMT
- Church of Jesus Christ of Latter-day Saints & Microsoft Translator Hub
- Dell & Safaba/welocalize
- DuDu & CapitaTI
- Ford & Systran/SAIC
- Sajan & Asia Online
- text&form & LucySoft

The time for MT is now!

- At *MT Summit XIV* in France, for the first time the number of commercial attendees exceeded those from academia.
- Ruopp (2013): for the first time in a TAUS survey, largest group of respondents was LSPs & translation agencies, not research institutes.
- Trends likely to continue, with more large multinational companies, LSPs and MT developers attending such events ...

Not everyone agrees ...

Why Corpus-Based MT?

- the (relative) failure of rule-based approaches
- the increasing availability of machine-readable text
- the increase in capability of hardware (CPU, memory, disk space) with decrease in cost

Why is MT Hard?

- Human languages are:
  - Elegant
  - Efficient
  - Flexible
  - Complex
- One word/sentence may mean many things
- Many ways of saying the same thing
- Meaning depends on context
- Literal and figurative language (metaphor)
- Language and culture (different ways of conceptualising the same thing)
- Word order
- Morphology
- …

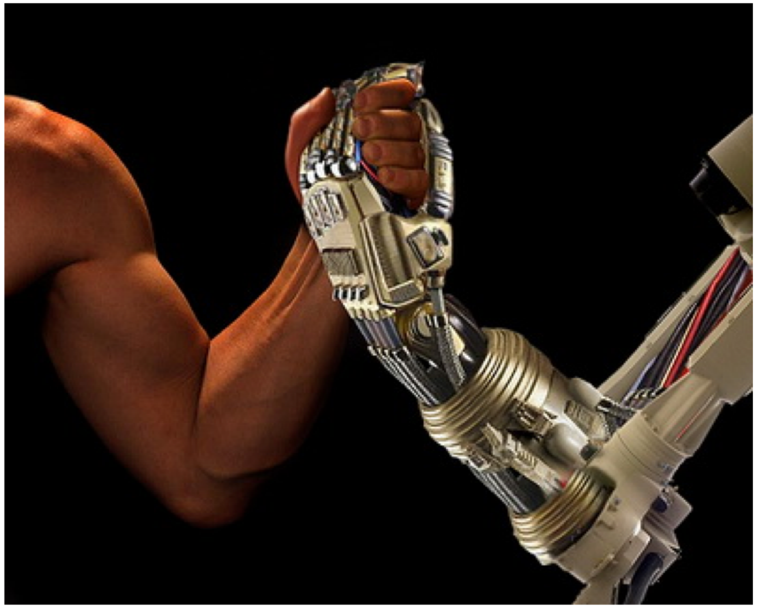


Image: http://workingtropes.lmc.gatech.edu/wiki/index.php/File:Man-vs-machine.jpg
License: CC BY-NC-SA 3.0

Why is MT Hard?

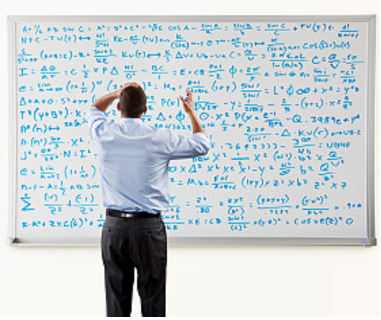
Newspaper Headlines:

1. Minister Accused Of Having 8 Wives In Jail
2. Juvenile Court to Try Shooting Defendant
3. Teacher Strikes Idle Kids
4. Miners refuse to work after death
5. Local High School Dropouts Cut in Half
6. Red Tape Holds Up New Bridges
7. Clinton Wins on Budget, but More Lies Ahead
8. Hospitals Are Sued by 7 Foot Doctors
9. Police: Crack Found in Man's Buttocks

Thanks to Chris Manning

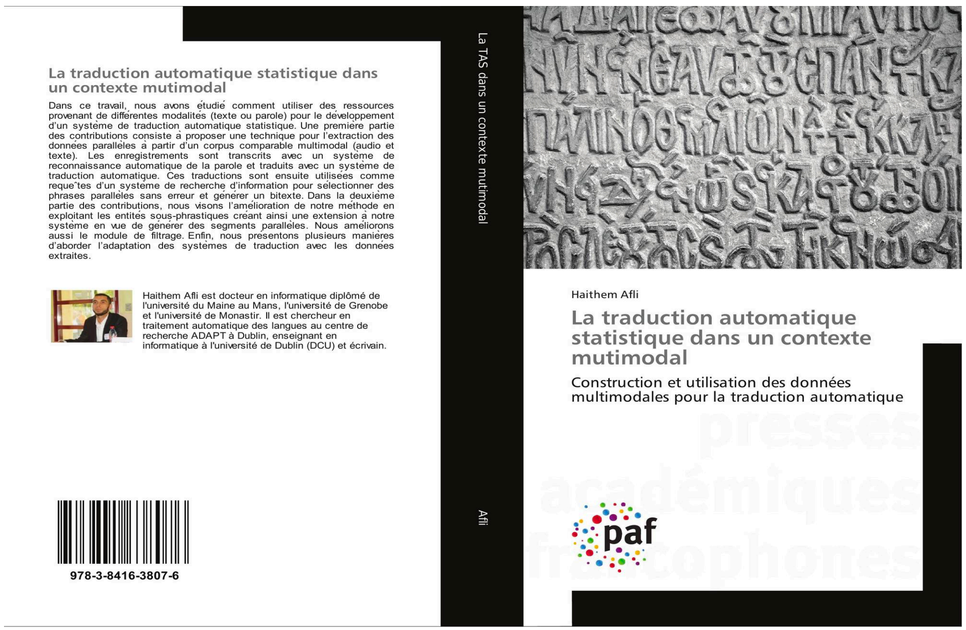
Language & Translation is Complex

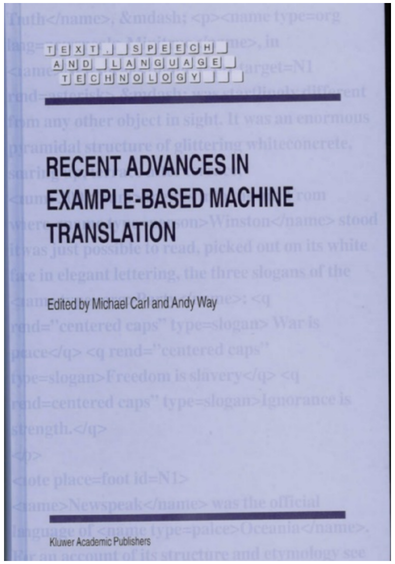
- Language/translation is complex
- We cannot compute it exactly
- We tried: rule-based MT and LT …
- What do we do *now*?
- Machine Learning
  - Learns from data ⇒ data is (mostly) all important
  - Approximate solution ⇒ not perfect, needs help
    - Human Professional Translators
    - Post-editing
    - Automated Translation ≠ Automatic

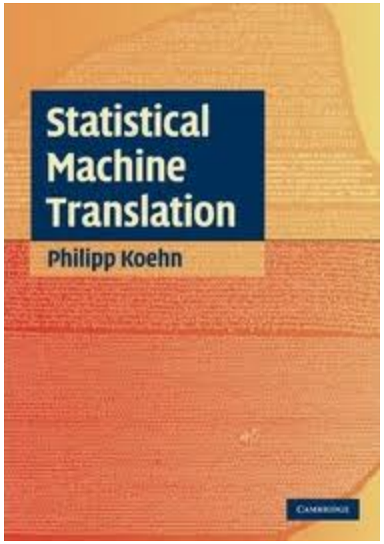


Types of Corpus-based MT

- Example-Based MT (Nagao, 1984)
- Statistical MT
  - 1988: word-based (IBM)
  - 2002—now: phrase-based (Moses)
  - 2005—now: tree-based (Hiero)







- Neural MT (1997 … and 2013 onwards)

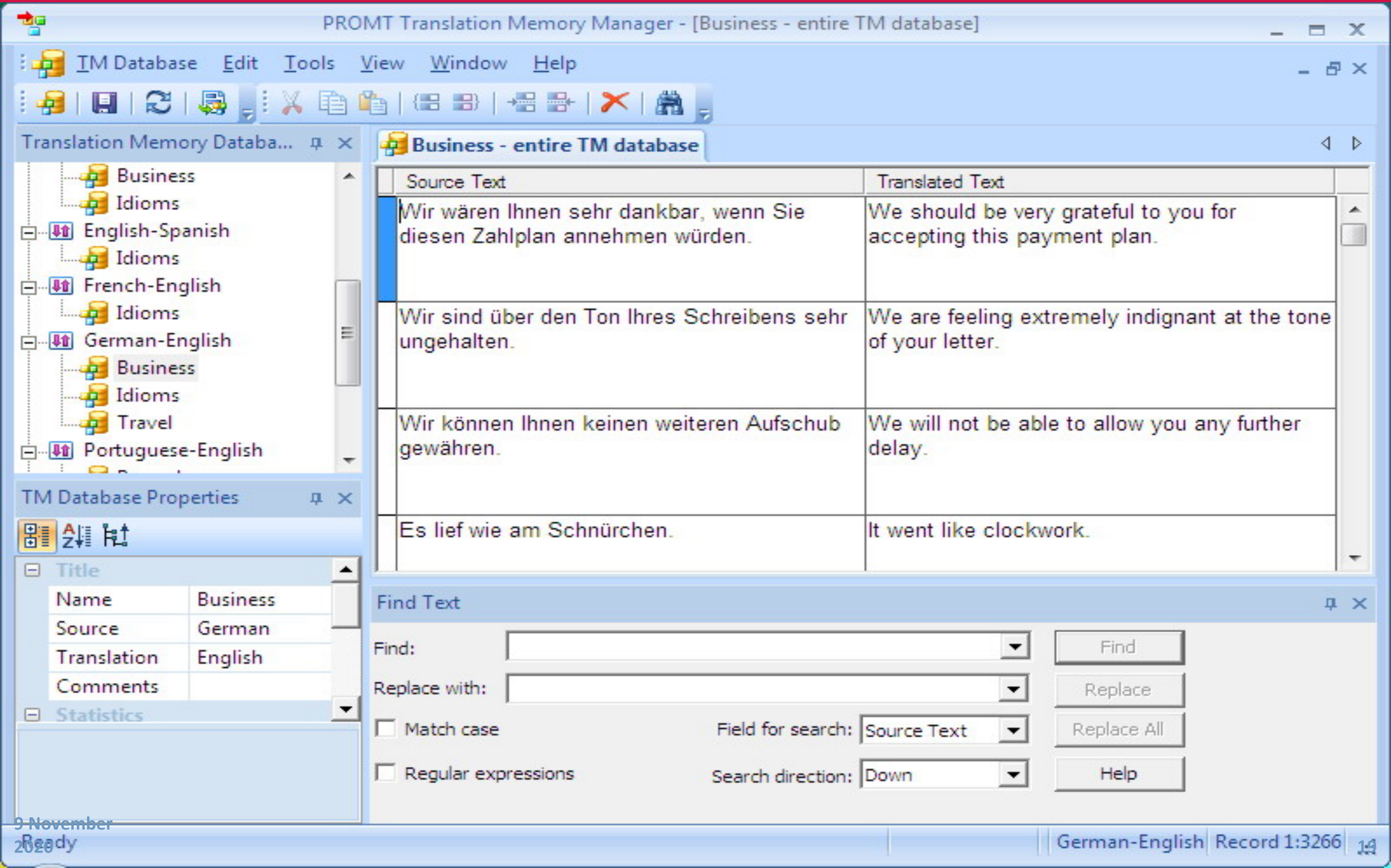
Prerequisite

A prerequisite for Data-Driven MT (and also TM, which is *not* MT, but rather CAT):

- Example-Based MT (EBMT)
- Statistical MT (SMT) & Neural MT (NMT)
- Hybrid Models which use some probabilistic processing

is a *parallel corpus* (or *bitext*) of aligned sentences.

Parallel data prerequisite for corpus-based MT



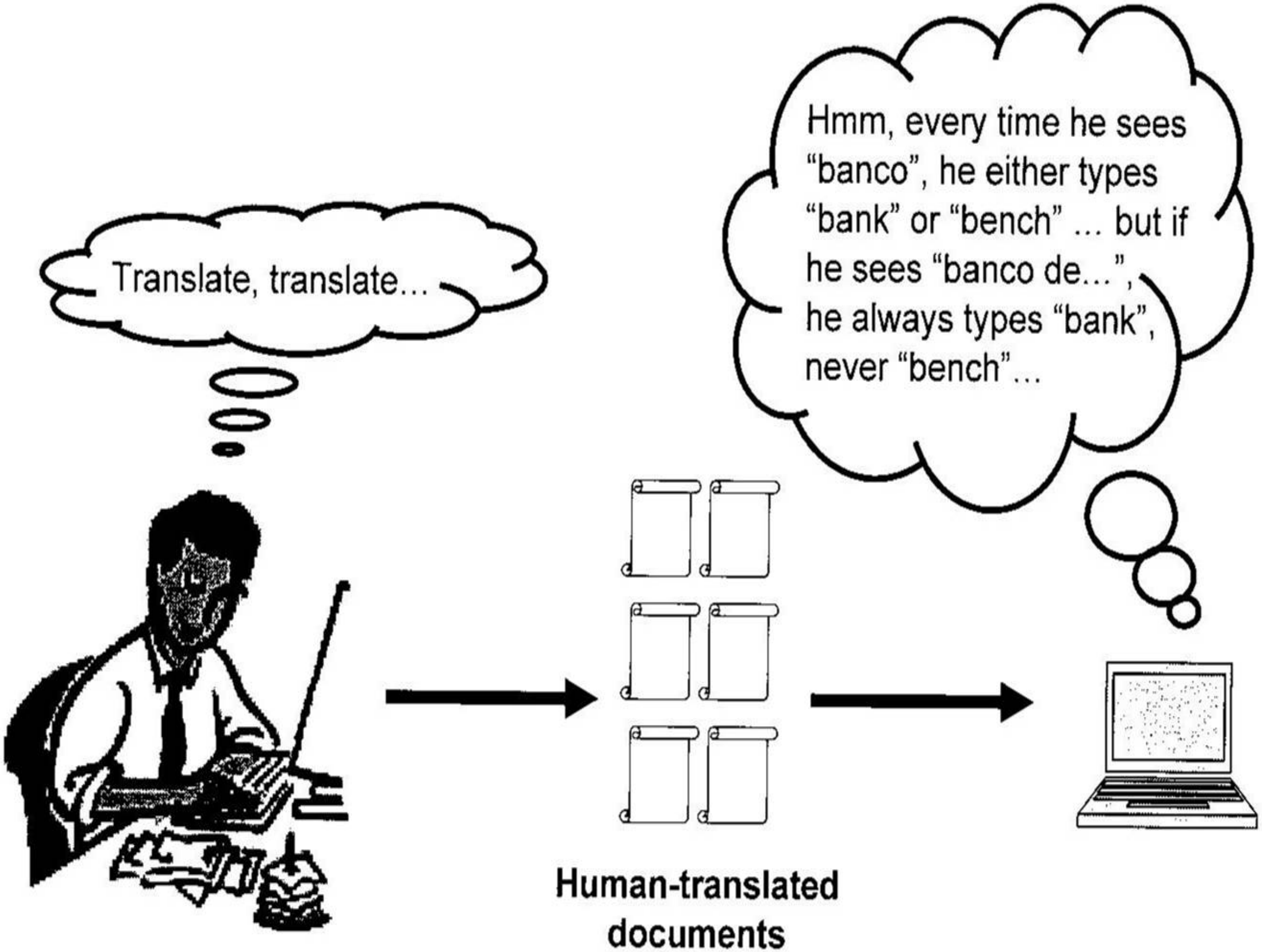
Parallel data prerequisite for corpus-based MT

So how does SMT work?

How might *you* go about translating between two languages you know

nothing about?!

Statistical Machine Translation



Thanks to Kevin Knight …

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok **farok** ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok **farok** izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok **farok** ororok lalok sprok izok enemok . 7b. wat **jjat** bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok **farok** izok stok . 5b. totat **jjat** quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok **farok** ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok **farok** izok stok . 5b. totat jjat quat cat .	11a. lalok nok **crrrok** hihok yorok zanzanok . ??? 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok **hihok** ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok **hihok** yorok zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok **hihok** mok . 12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok **hihok** ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok **yorok** ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok **hihok** **yorok** zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok **hihok** mok . 12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok **clok** . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok **clok** . ??? 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . process of elimination 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok kantok ok-yurp

1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat . cognate?
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

Centauri/Arcturan [Knight, 1997]

Your assignment, put these words in order: **{ jjat, arrat, mat, bat, oloat, at-yurp }**

1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat . zero fertility
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

Some more to try …

- iat lat pippat eneat hilat oloat at-yurp.
- totat nnat forat arrat mat bat.
- wat dat quat cat uskrat at-drubel.

Some more to try …

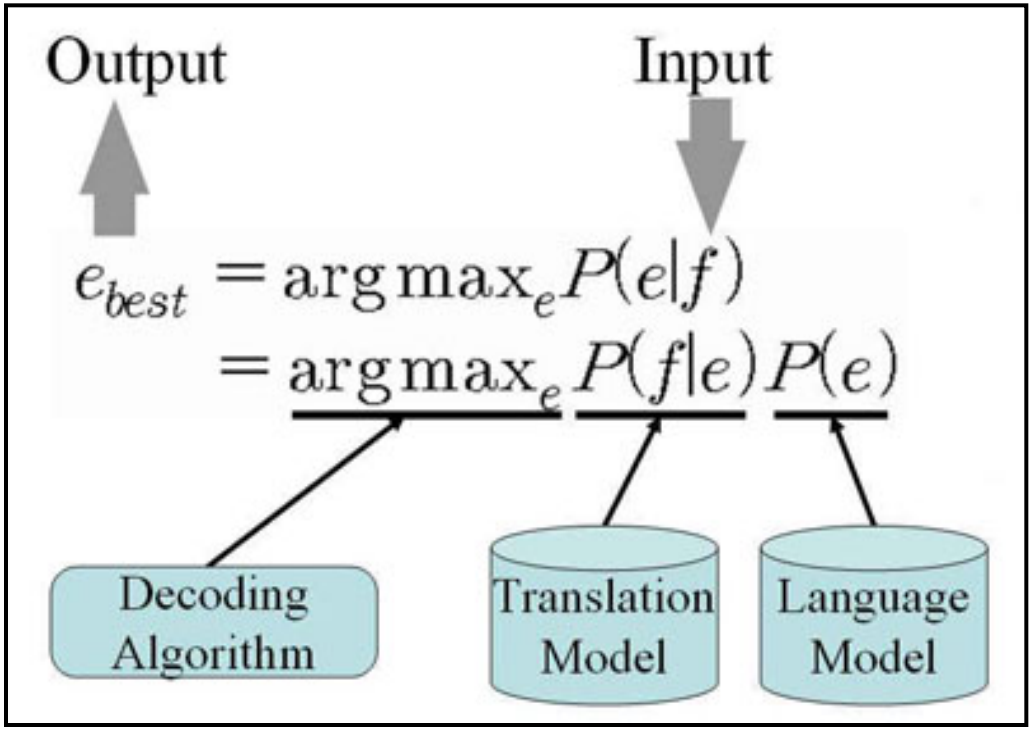
- iat lat pippat eneat hilat oloat at-yurp.
- totat nnat forat arrat mat bat.
- wat dat quat cat uskrat at-drubel.

… if you have trouble sleeping at night!

Discussion

37

How does SMT Work?



- No(t much) maths today …
- Instead:
  - The story of SMT in pictures …
  - It's (mostly) all about the **Data** …

How does SMT Work?

Statistical MT learns from data

Two kinds of data:

- Source documents and their human translations
- Target language collections

- The more data the better!
- Also: the right kind of data!

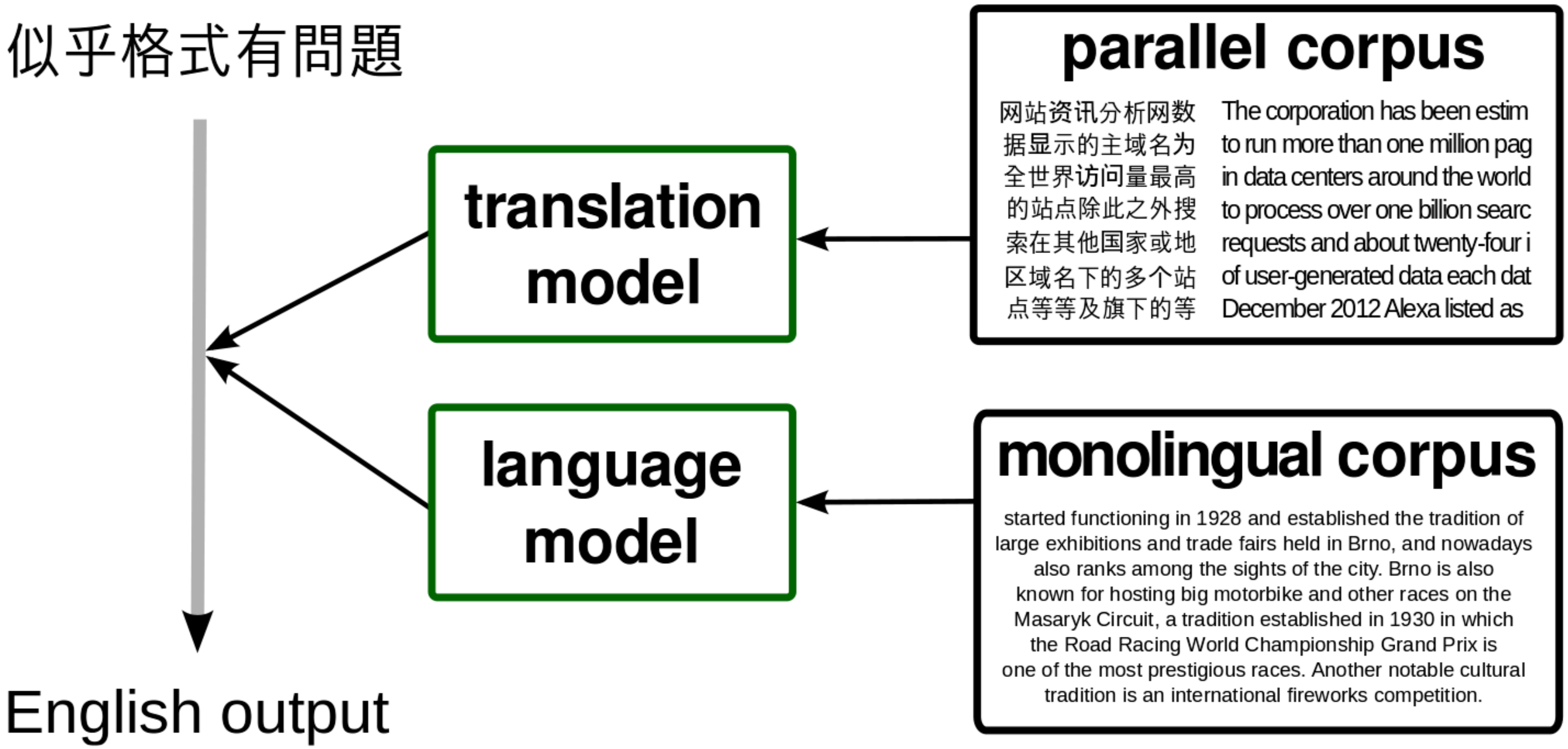
GERMAN	ENGLISH	FRENCH
Einleitung	Introduction	Introduction
I. Von dem Unterschiede der reinen und empirischen Erkenntnis	*I. Of the difference between Pure and Empirical Knowledge*	*I. De la différence de la connaissance pure et de la connaissance empirique.*
Daß alle unsere Erkenntnis mit der Erfahrung anfange, 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 Verstandestätigkeit 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.	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.	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.

What can/do we learn from Data?

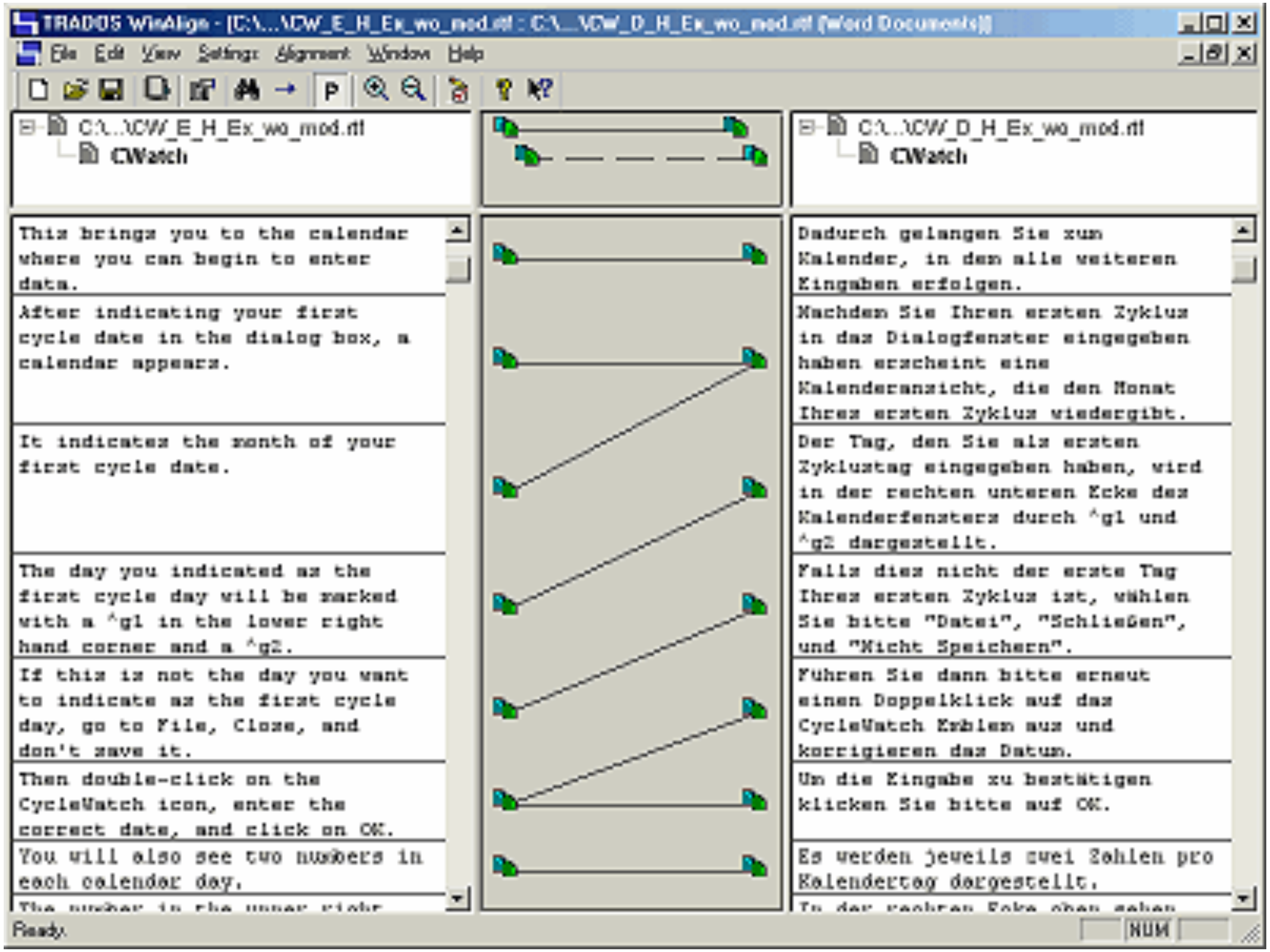
- Which sentences translate as which: sentence alignment
- Which words translate as which: word alignment + translation probabilities => translation model
- What do good target sentences look like: language model

GERMAN	ENGLISH	FRENCH
Einleitung	Introduction	Introduction
I. Von dem Unterschiede der reinen und empirischen Erkenntnis	*I. Of the difference between Pure and Empirical Knowledge*	*I. De la différence de la connaissance pure et de la connaissance empirique.*
Daß alle unsere Erkenntnis mit der Erfahrung anfange, 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 Verstandestätigkeit 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.	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.	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.

How does SMT Work?



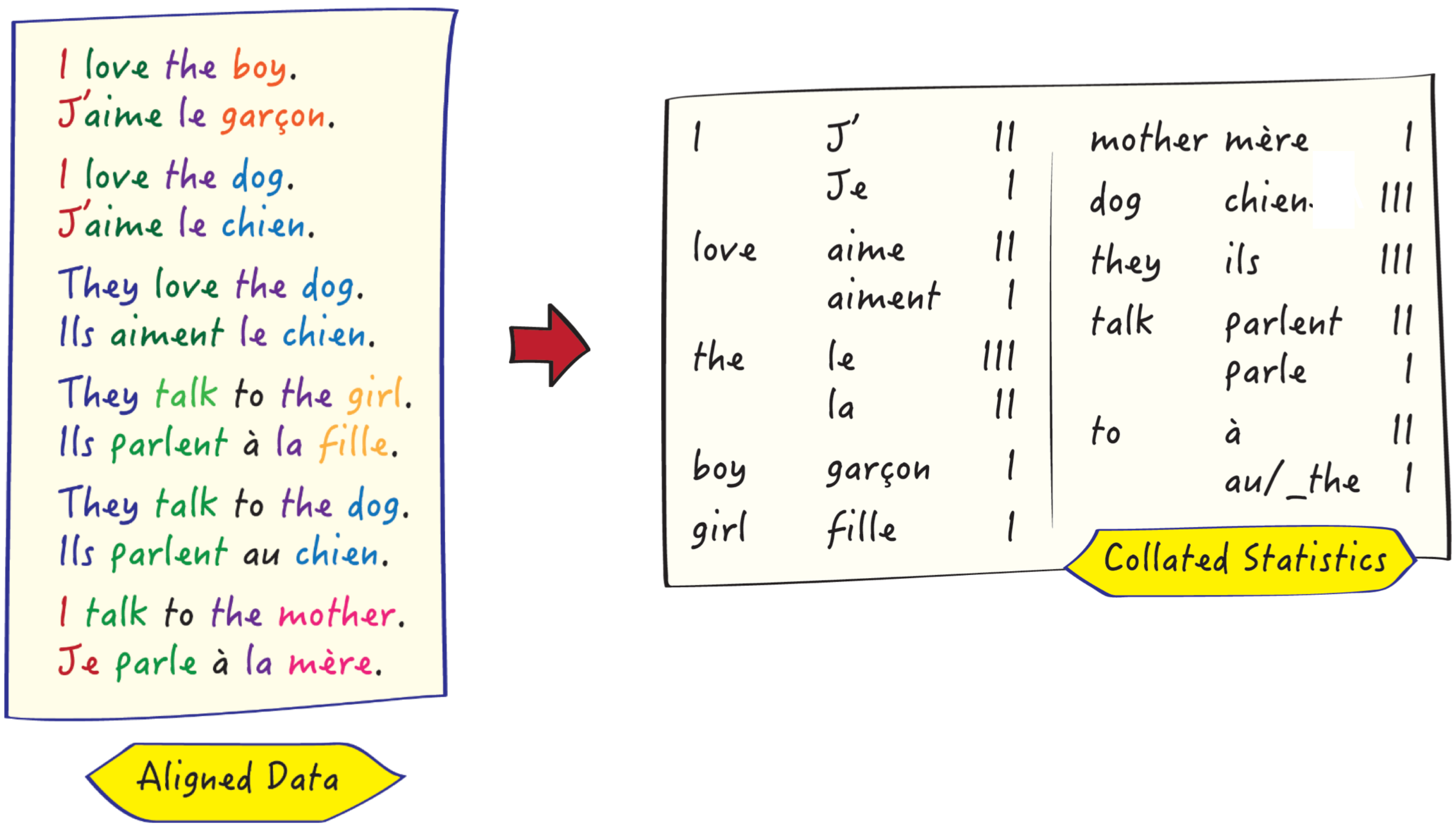
Sentence Alignment



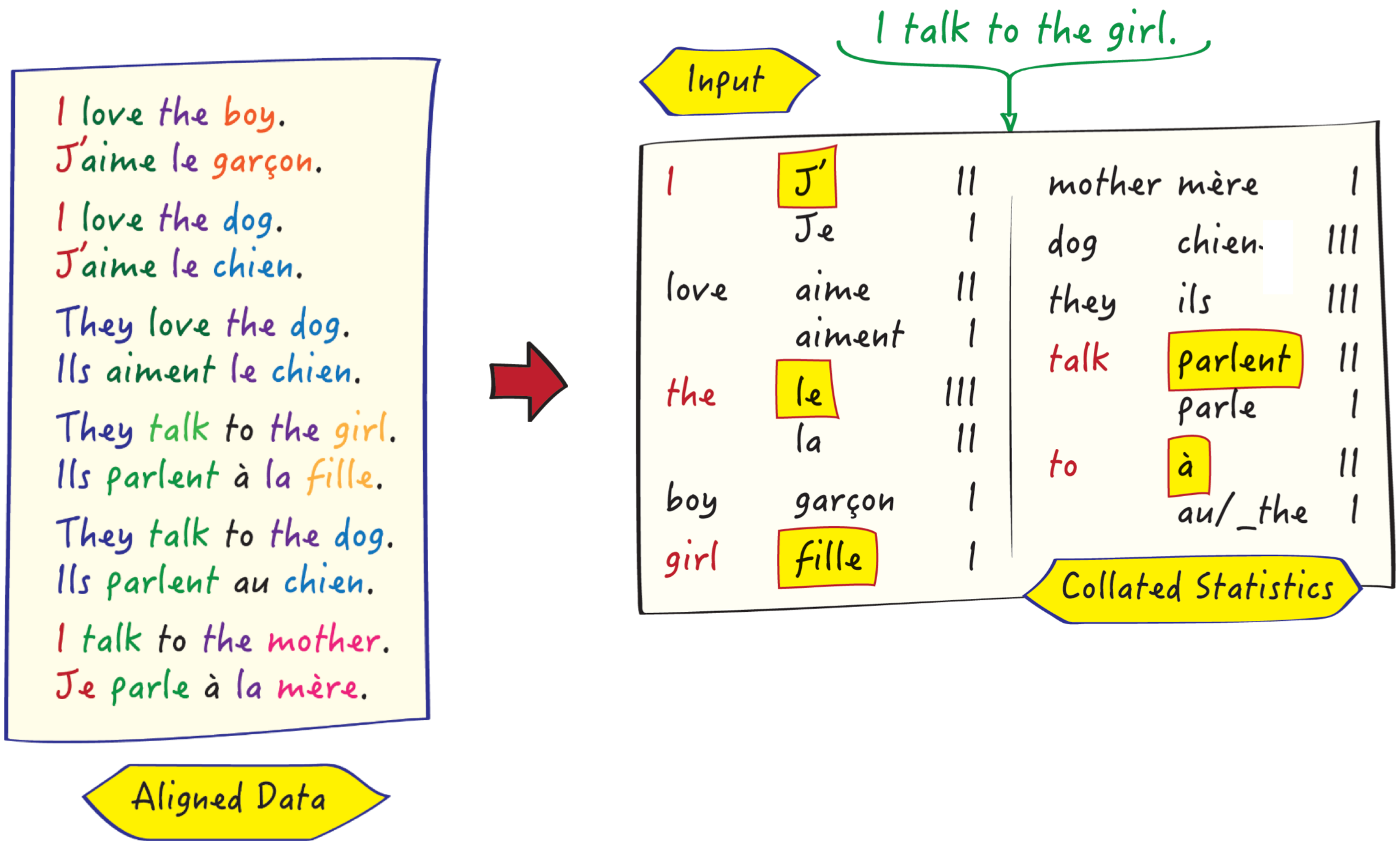
Let’s try it for a another language pair …

Statistical MT

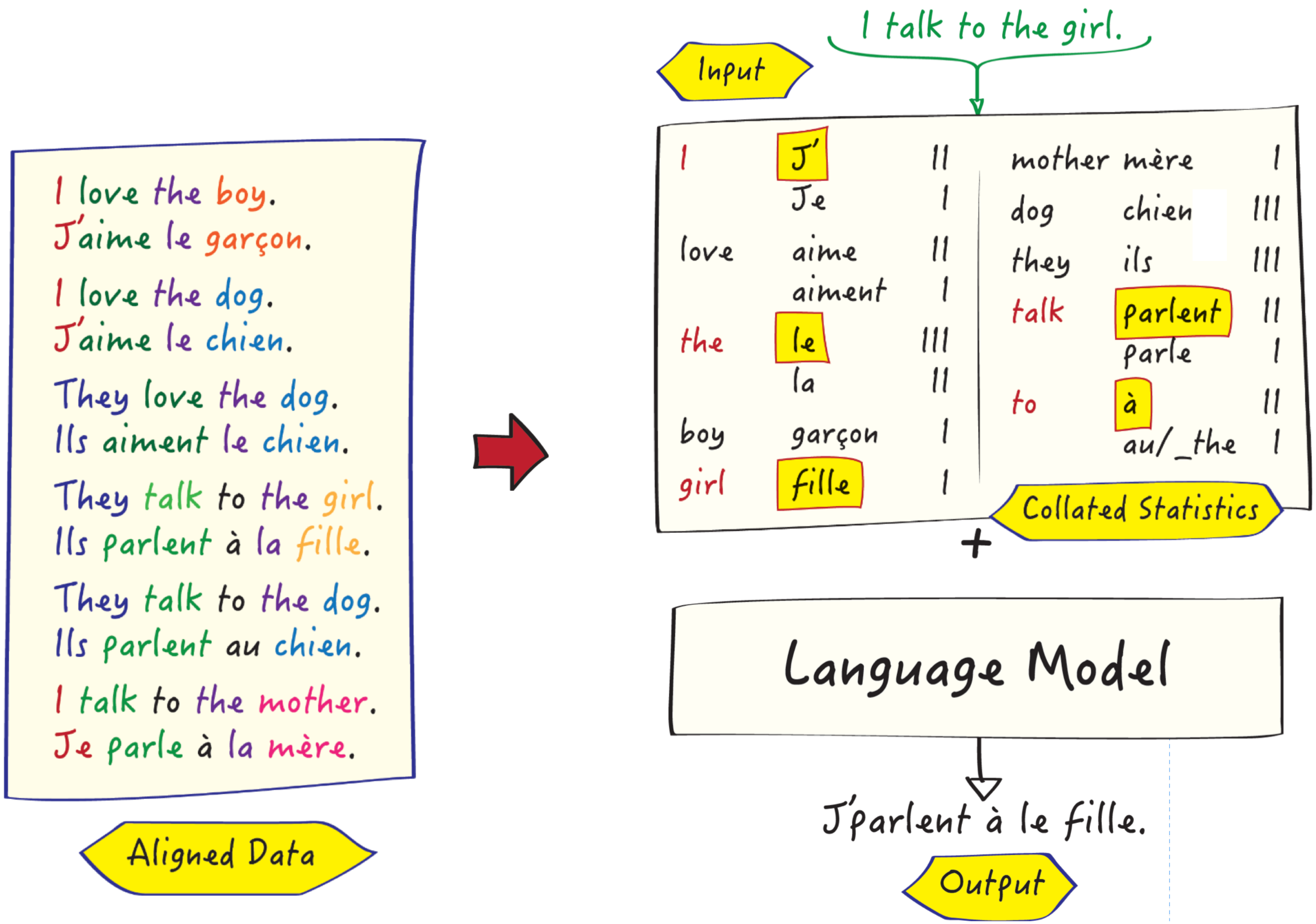
Statistical MT



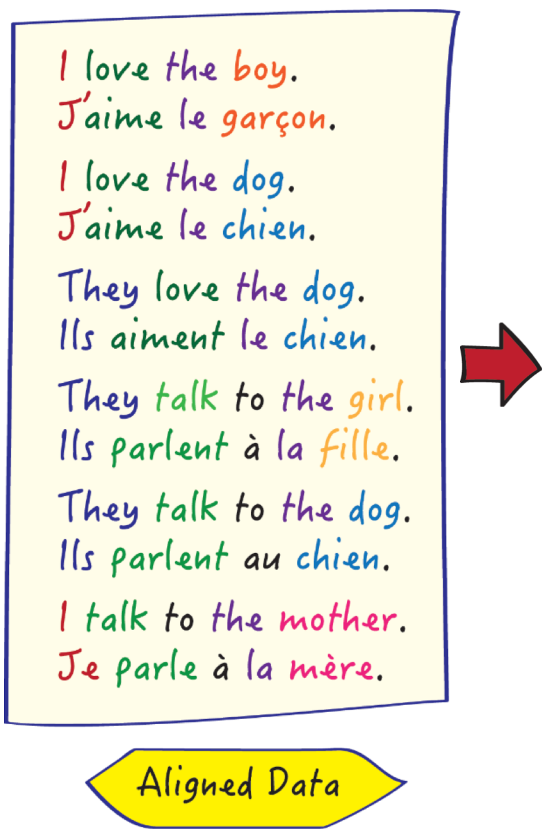
Statistical MT



Statistical MT



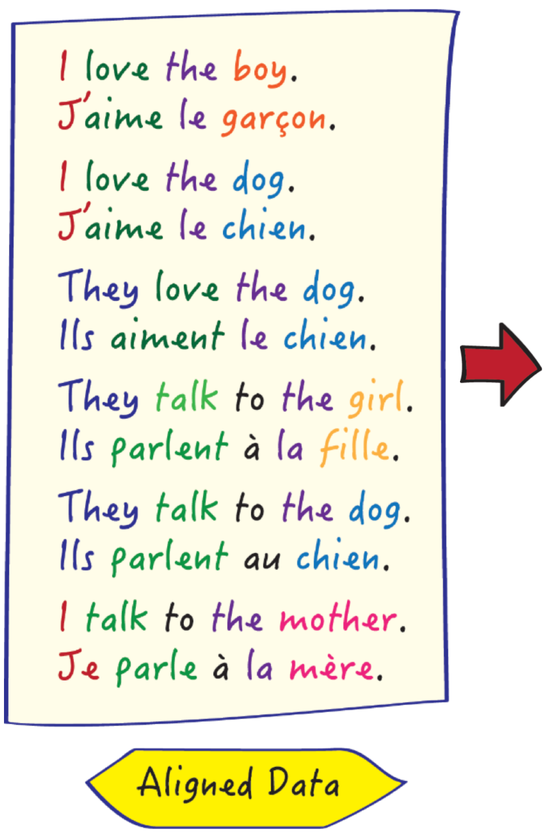
Statistical MT



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	

How to choose?

Statistical Machine Translation



The Language Model:

- What is good target language?
- Which words can follow which words and which can't? The "grammar"!
- Learnt from the data …
  - *Je parle* is good …
  - *J' parlent* is bad …
  - *la fille* is good …
  - *le fille* is bad …
- *Je parle à la fille* >> *J' parlent à le fille*

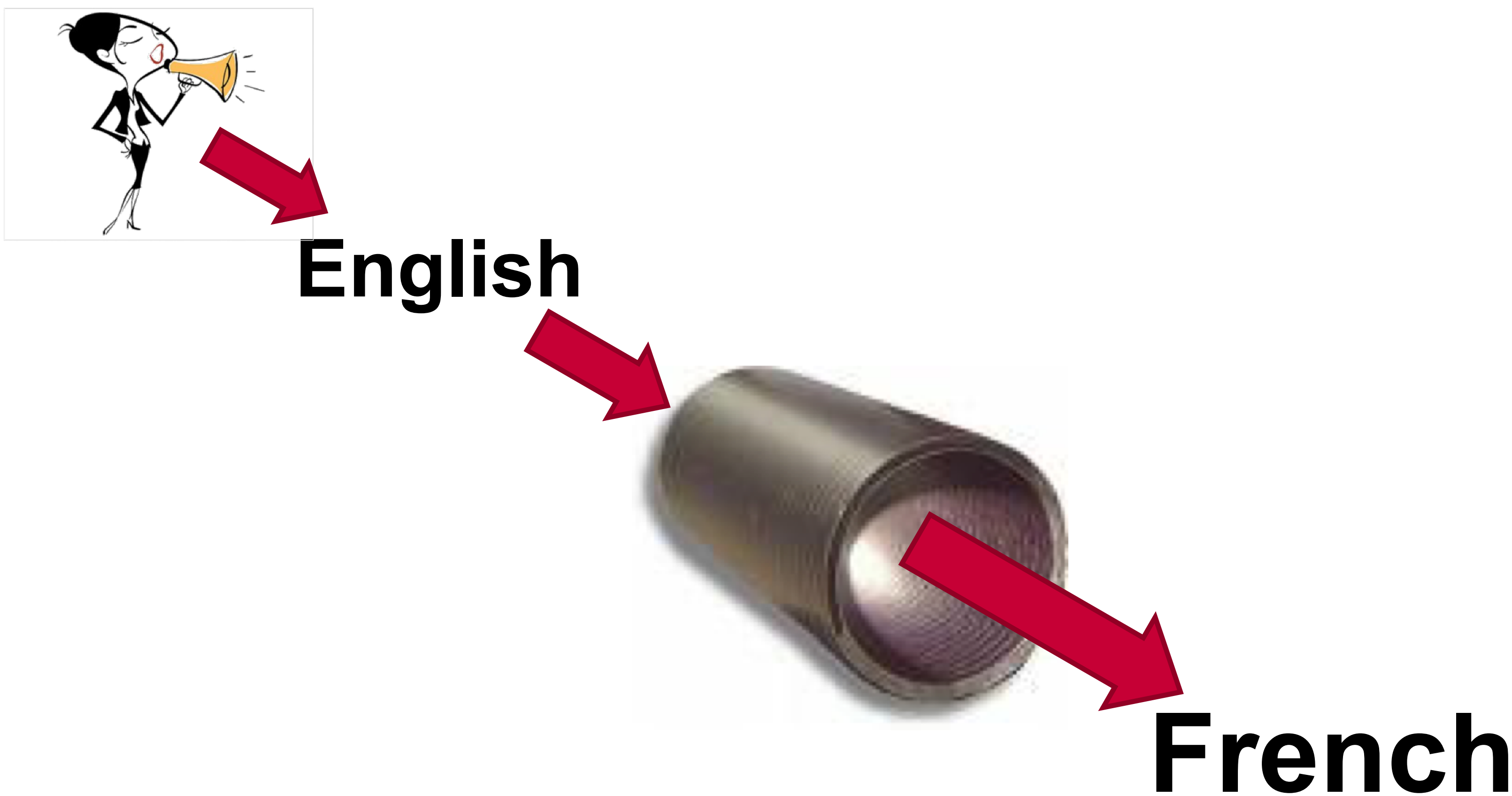
Noisy Channel Framework

English

French

Noisy Channel Framework

$$p(e) \rightarrow e \rightarrow p(f|e) \rightarrow f$$



Noisy Channel Framework

$$p(e) \rightarrow e \rightarrow p(f|e) \rightarrow f$$

$$p(f) = p(e)p(f|e)$$

English

French

Noisy Channel Framework

Applying Bayes' Rule, we have:

$$p(e|f) = \frac{p(e)p(f|e)}{p(f)}$$

Thus:

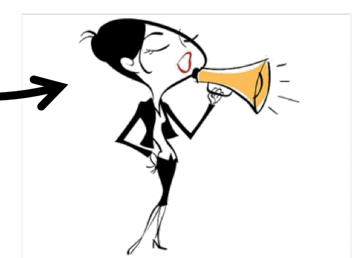
$$\hat{e} = \underset{e}{\operatorname{argmax}}\, p(e|f) = \underset{e}{\operatorname{argmax}}\, p(e)p(f|e)$$

SMT Components

$$\hat{e} = \underbrace{\operatorname*{argmax}_{e}}\ \underbrace{p(e)}\ \underbrace{p(f|e)}$$

Translation Model

Decoder



Language Model

1990s-2010s: Statistical Machine Translation

- Core idea: Learn a probabilistic model from data
- Suppose we're translating French → English.
- We want to find best English sentence *y*, given French sentence *x*

$$\text{argmax}_y P(y|x)$$

- Use Bayes Rule to break this down into two components to be learnt separately:

$$= \text{argmax}_y \underbrace{P(x|y)}_{\text{Translation Model}} \underbrace{P(y)}_{\text{Language Model}}$$

Translation Model

Models how words and phrases should be translated.
Learnt from parallel data.

Language Model

Models how to write good English.
Learnt from monolingual data.

Noisy Channel Framework

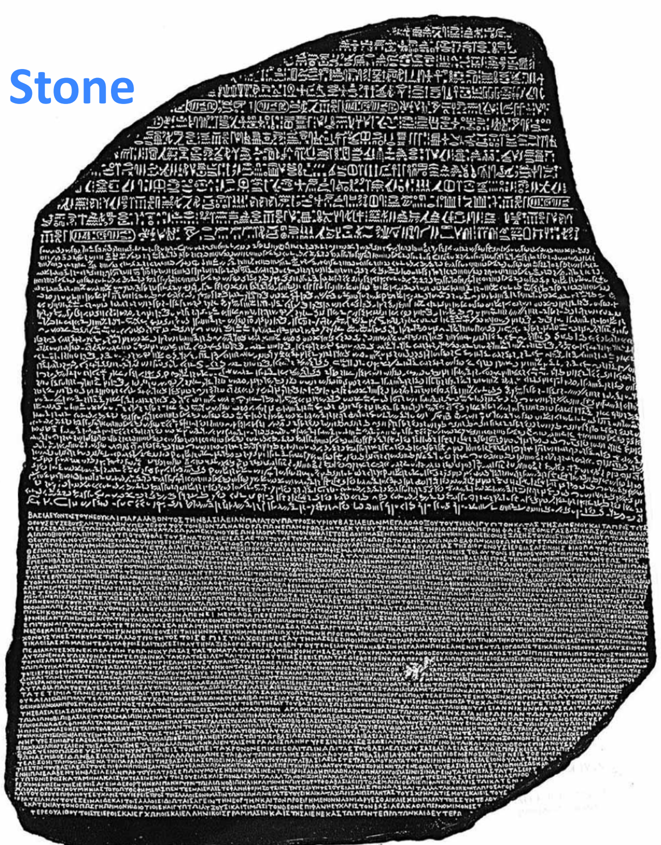
- The *translation model* models how likely it is that f is a translation of e – adequacy.
- The *language model* models how likely it is that e is an acceptable sentence – fluency.
- The *decoder* searches for the most likely e.

Discussion

57

1990s-2010s: Statistical Machine Translation

- Question: How to learn translation model $P(x|y)$?
- First, need large amount of parallel data
  (e.g. pairs of human-translated French/English sentences)



1990s-2010s: Statistical Machine Translation

- Question: How to learn translation model $P(x|y)$?
- First, need large amount of parallel data
  (e.g. pairs of human-translated French/English sentences)

- Break it down further: we actually want to consider

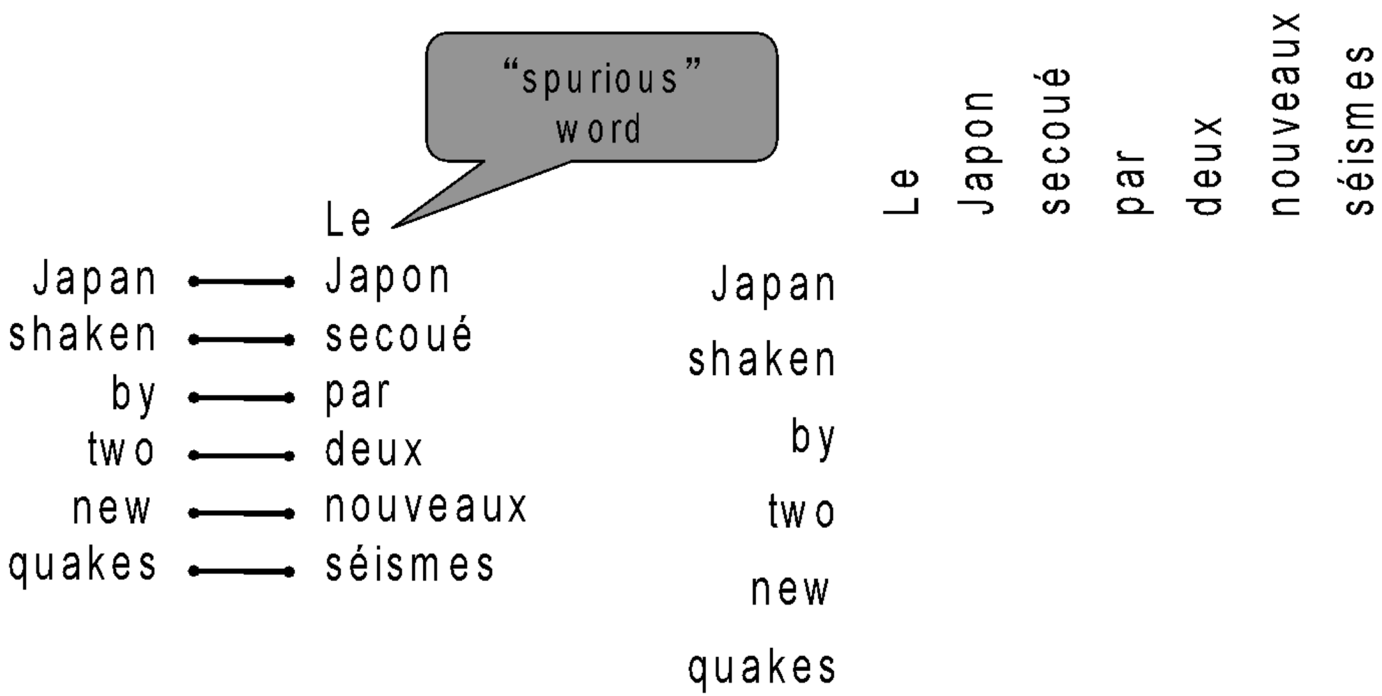
$$P(x, a|y)$$

  where *a* is the alignment, i.e. word-level correspondence between French sentence *x* and English sentence *y*

What is alignment?

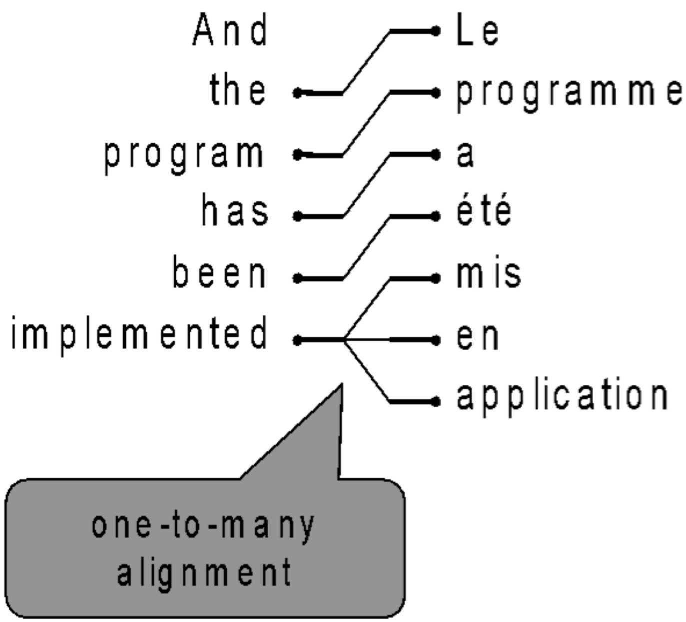
Alignment is the correspondence between particular words in the translated sentence pair.

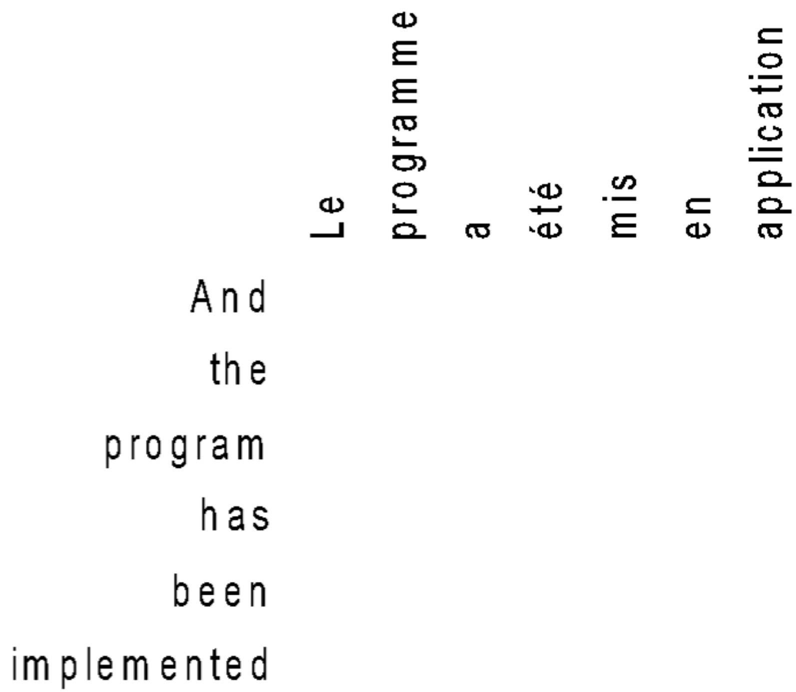
- Note: Some words have no counterpart



Alignment is complex

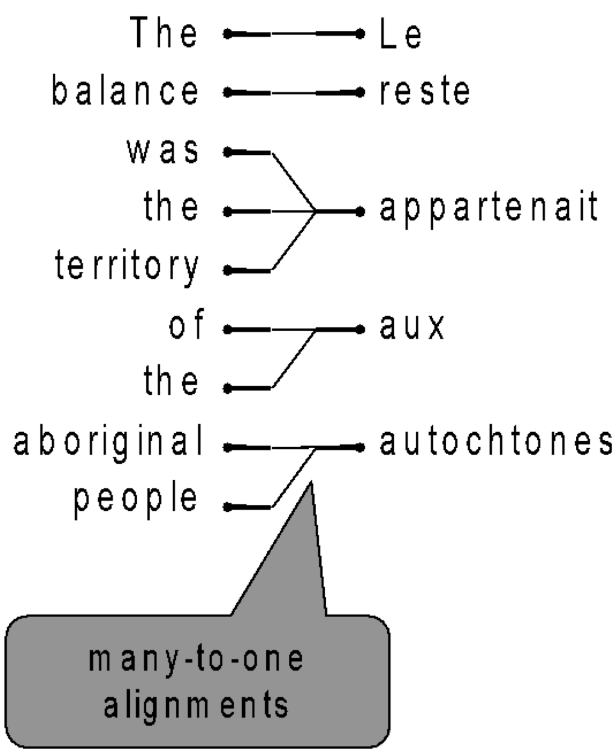
Alignment can be one-to-many (these are “fertile” words)

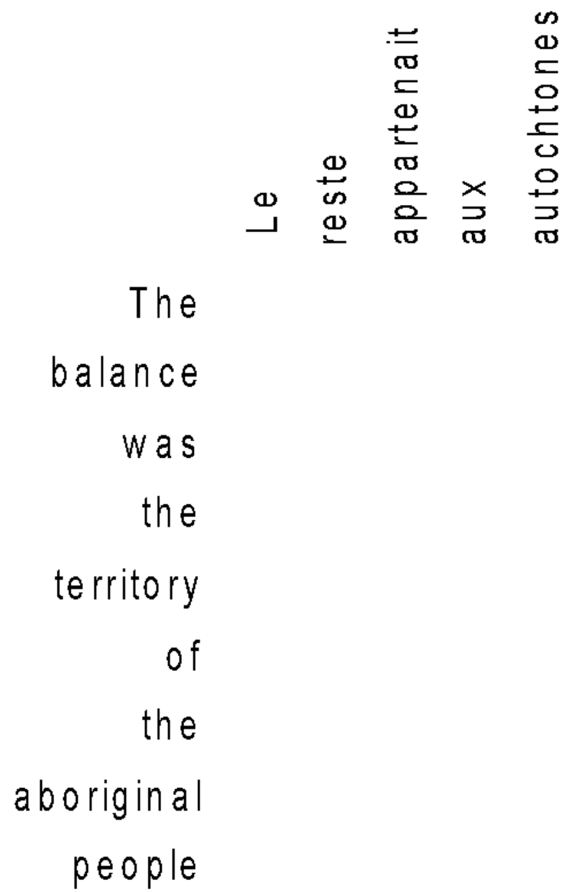




Alignment is complex

Alignment can be many-to-one





Phrase-Based SMT

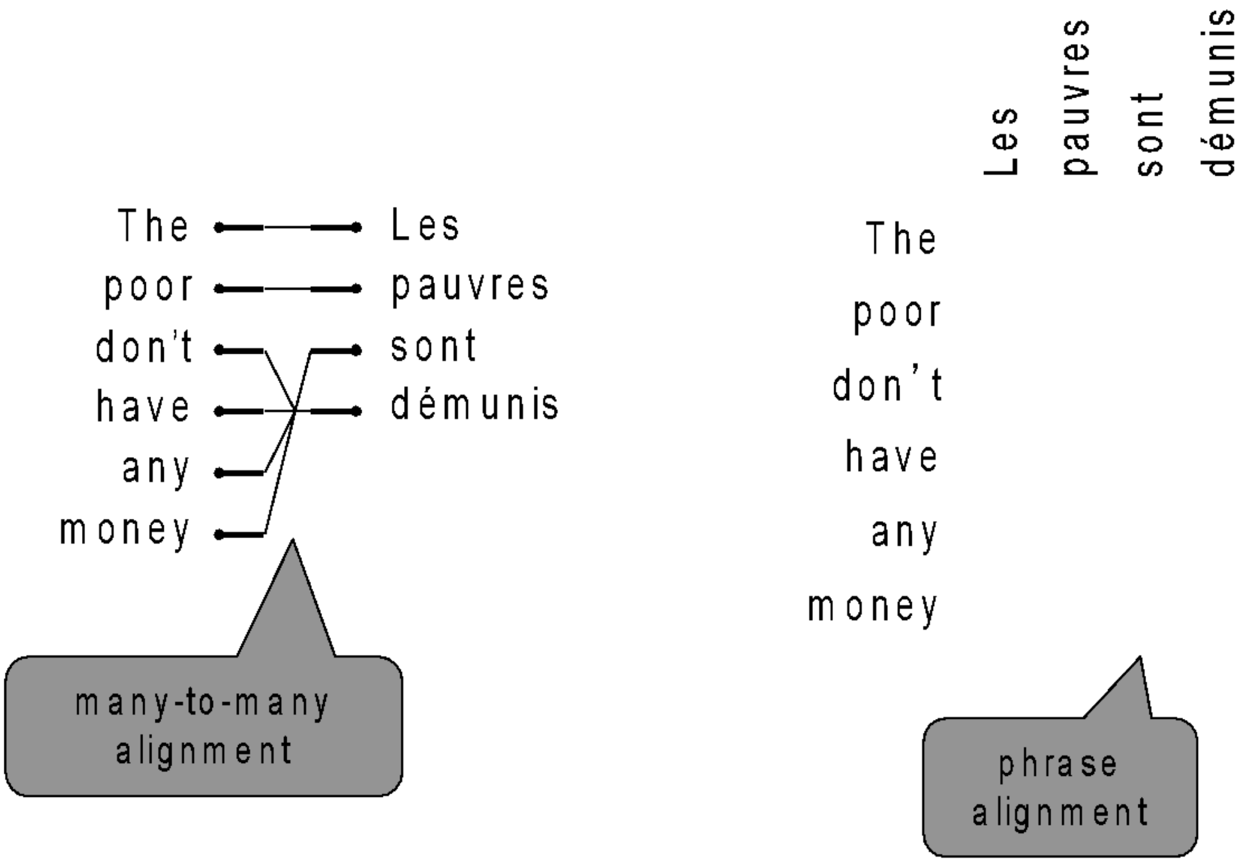
- So far: translating single words
- Loses context: such as agreement (******le fille* ...) etc.
- To some extent "repaired" by language model

- A better model:
  - Not just translations of single words
  - But also phrase translations:

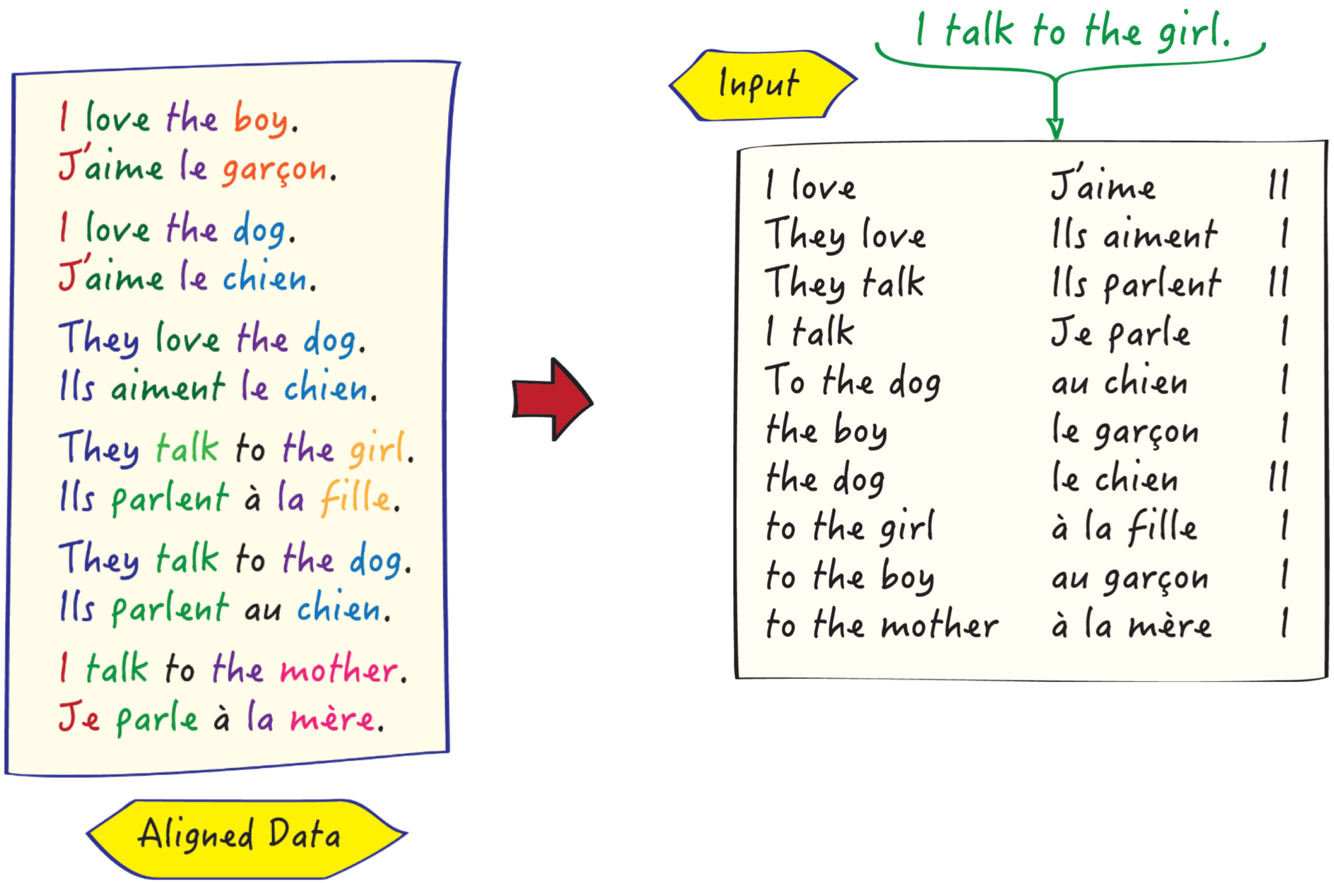
– *the girl : la fille*
– *to the girl : à la fille*
– *I talk : Je parle*

Alignment is complex

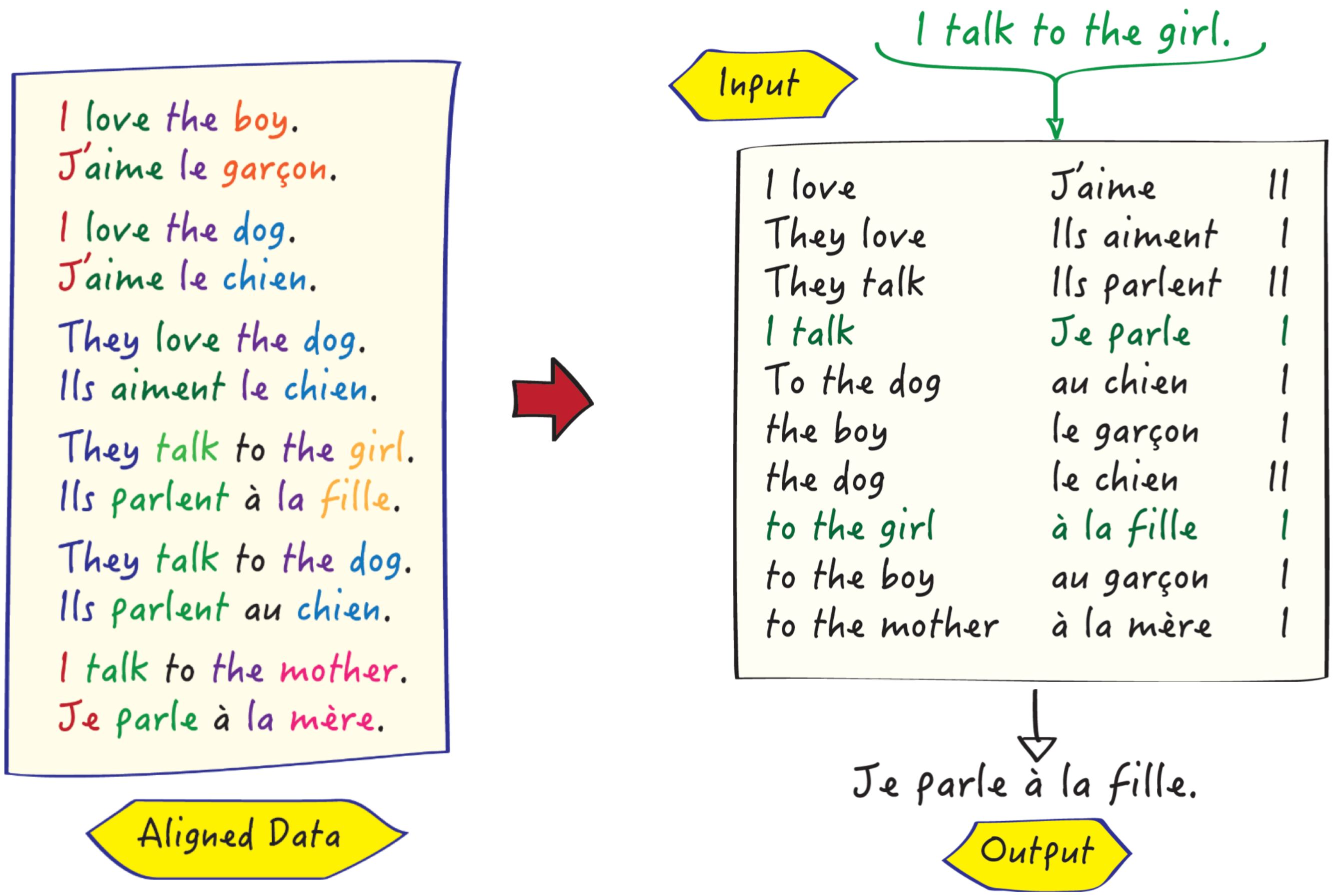
Alignment can be many-to-many (phrase-level)



Phrase-Based SMT



Phrase-Based SMT



Phrase-Based SMT

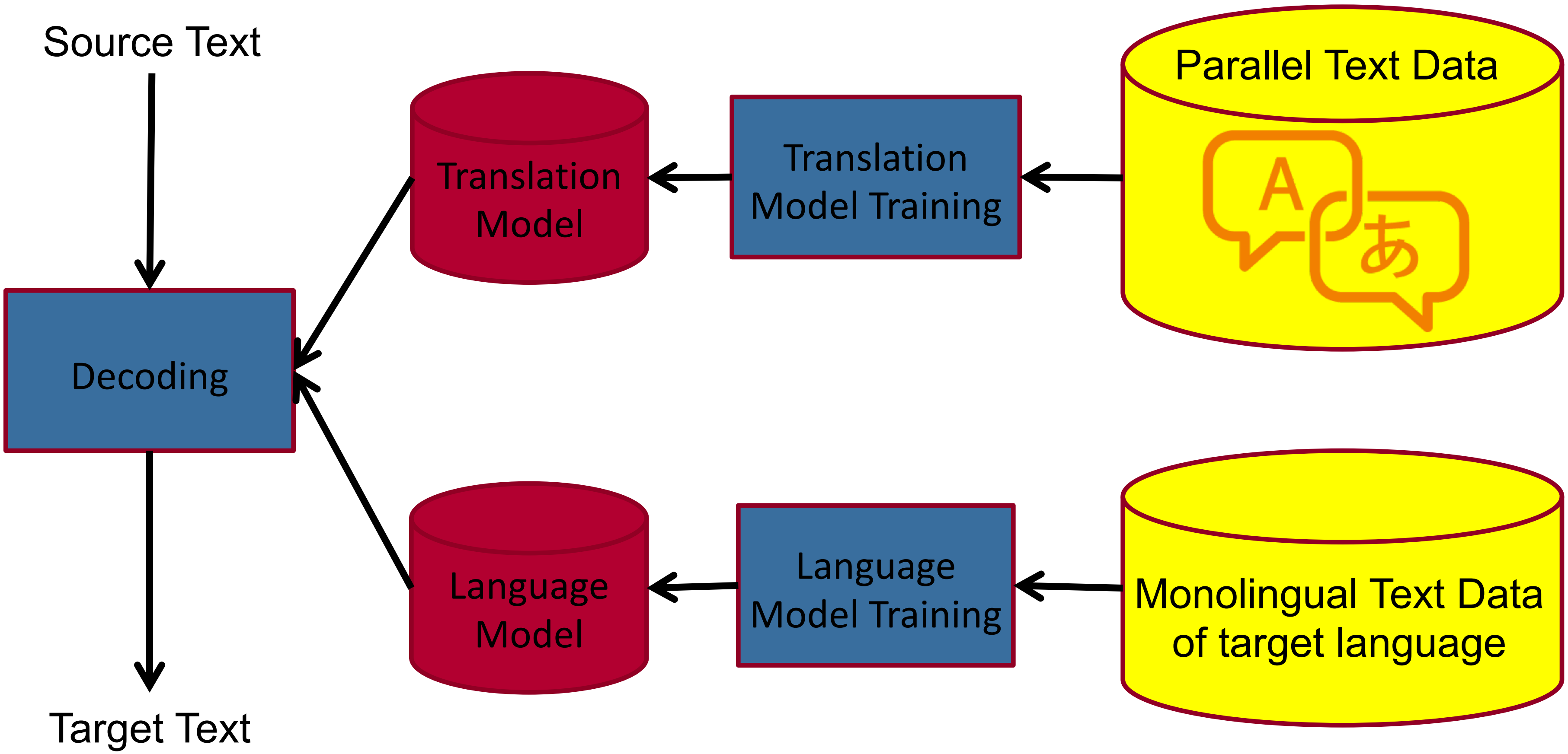
- *Much* better than word-based SMT!
- Standard technology: global Localisation & Translation industry

- Moses open-source PB-SMT toolkit
- Most widely used SMT platform
- Research funded by EC
- Used by EC DGT's MT@EC

Good Quality Language & Translation Models

- Any statistical approach to MT requires the availability of aligned bilingual corpora which are:
  - large;
  - good-quality;
  - representative.

SMT Flow



Source Text
Parallel Text Data
Translation Model
Translation Model Training
A
あ
Decoding
Language Model
Language Model Training
Monolingual Text Data of target language
Target Text

1990s-2010s: Statistical Machine Translation

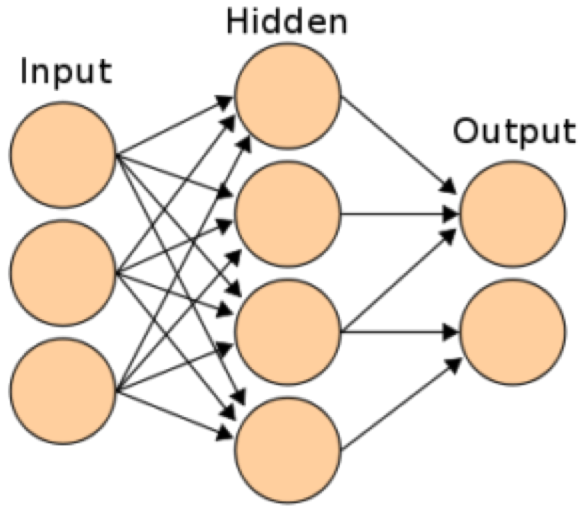
- SMT is a huge research field
- The best systems are extremely complex
  - Hundreds of important details we haven't mentioned here
  - Systems have many separately-designed subcomponents
  - Lots of feature engineering
    - Need to design features to capture particular language phenomena
  - Require compiling and maintaining extra resources
    - Like tables of equivalent phrases
  - Lots of human effort to maintain
    - Repeated effort for each language pair!

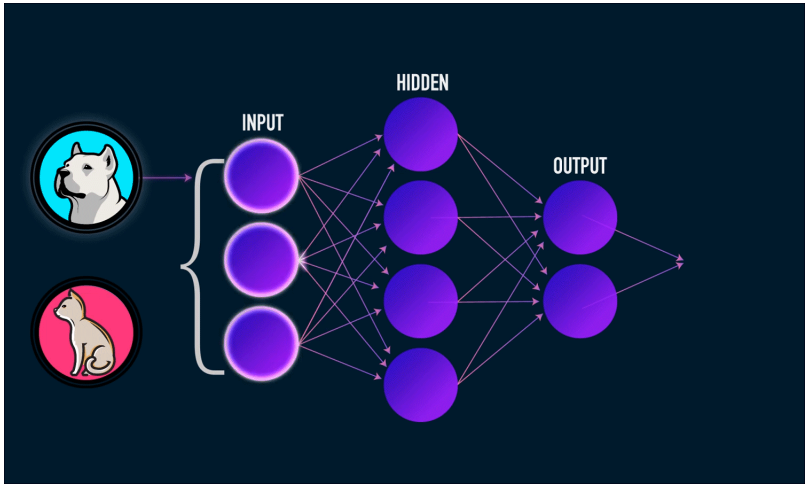
Discussion

71

Neural MT

- Paradigm Shift: Machine Learning → Deep Learning
- Hardware: CPU → GPU
- Open-Source Tools:
  - *Theano*
  - *Tensorflow*
  - *Pytorch*
  - *mxnet*
- Slower Training & Decoding





Neural MT

- End-to-end translation, not word-to-word (or phrase-to-phrase)
- Promising results compared to PB-SMT:
  - Good generalisation capability
  - Good predictive performance
- But issues with sentence length, vocab size

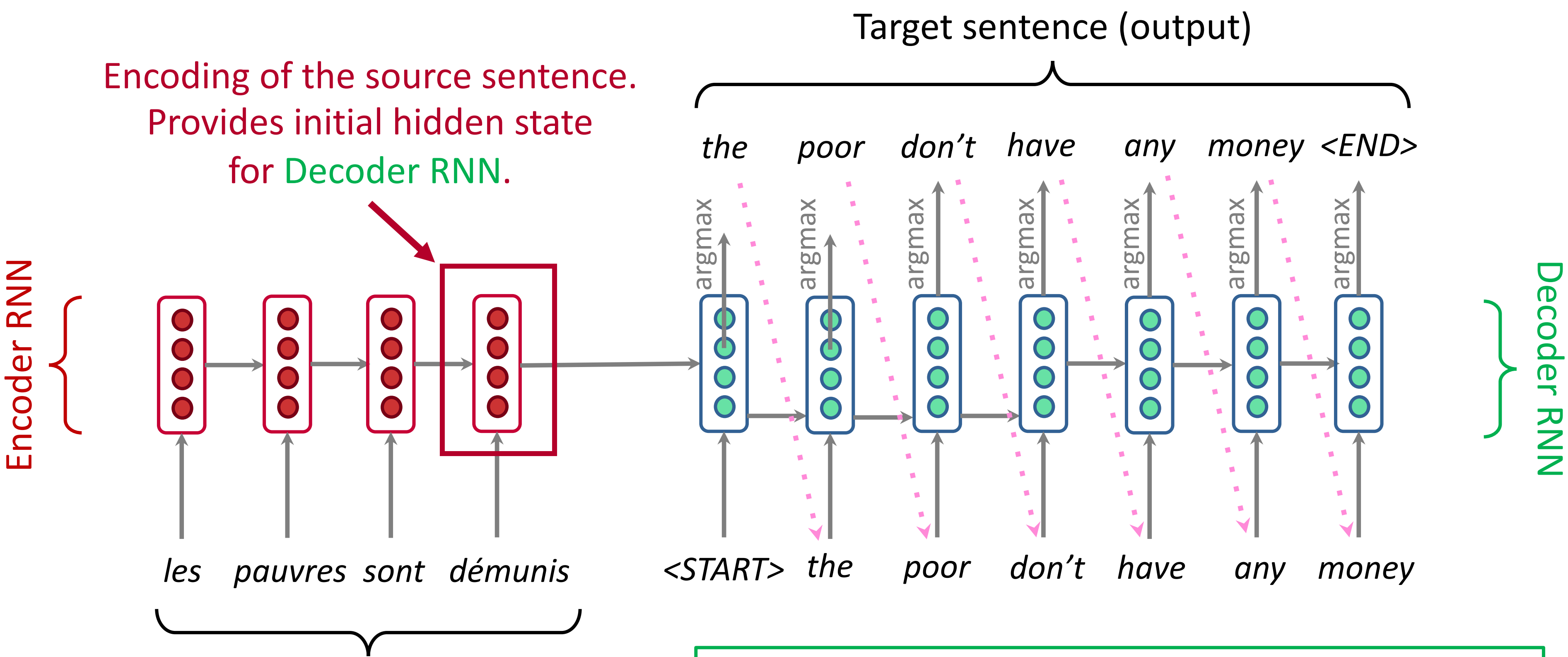
What is Neural Machine Translation?

- Neural Machine Translation (NMT) is a way to do Machine Translation with a *single neural network*

- The neural network architecture is called sequence-to-sequence (aka seq2seq) and it involves *two* RNNs.

Neural Machine Translation (NMT)

The sequence-to-sequence model



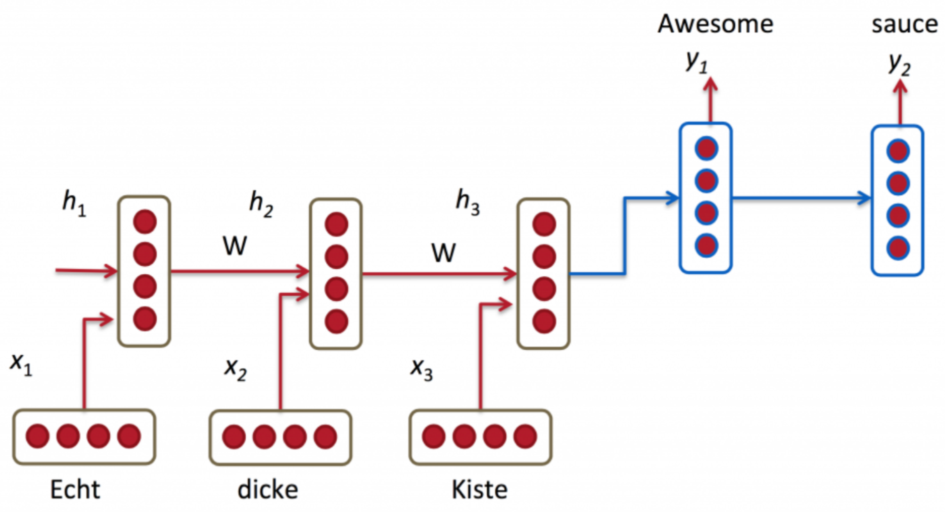
Decoder RNN is a Language Model that generates target sentence conditioned on encoding.

Encoder RNN produces an encoding of the source sentence.

Note: This diagram shows **test time** behavior: decoder output is fed in → as next step's input

Note there's an embedding layer for each CNN

You have two separate sets of word embeddings, for the two languages



RNN for Machine Translation. Image Source: http://cs224d.stanford.edu/lectures/CS224d-Lecture8.pdf

Neural Machine Translation (NMT)

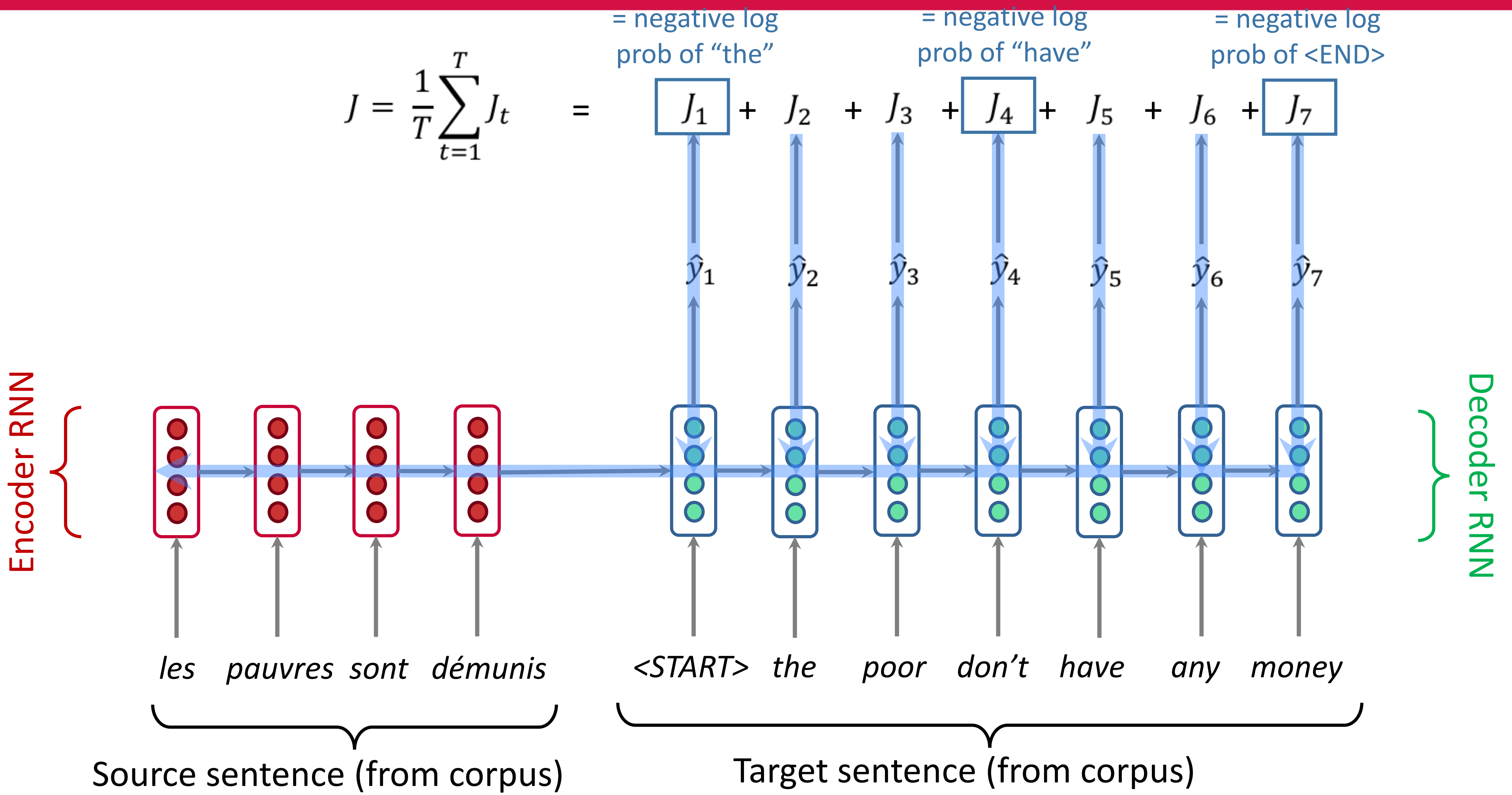
- The sequence-to-sequence model is an example of a **Conditional Language Model**.
  - **Language Model** because the decoder is predicting the next word of the target sentence *y*
  - **Conditional** because its predictions are *also* conditioned on the source sentence *x*

- NMT directly calculates $P(y|x)$:

$$P(y|x) = P(y_1|x)\, P(y_2|y_1, x)\, P(y_3|y_1, y_2, x) \ldots , \underbrace{P(y_T|y_1, \ldots , y_{T-1}, x)}_{\text{Probability of next target word, given target words so far and source sentence } x}$$

- NMT directly estimates p(y|x), whereas SMT modularized into translation model and language model
- **Question**: How to train a NMT system?
- **Answer**: Get a big parallel corpus…

Training a Neural Machine Translation system



Seq2seq is optimized as a **single system.**
Backpropagation operates *"end to end"*.

Advantages of NMT

Compared to SMT, NMT has many advantages:

- Better performance
  - More fluent
  - Better use of context
  - Better use of phrase similarities
- A single neural network to be optimized end-to-end
  - No subcomponents to be individually optimized
- Requires much less human engineering effort
  - No feature engineering
  - Same method for all language pairs

Disadvantages of NMT?

Compared to SMT:

- NMT is less interpretable
  - Hard to debug
- NMT is difficult to control
  - For example, can't easily specify rules or guidelines for translation
  - Safety concerns!

Disadvantages of NMT?

Compared to SMT:

- NMT is less interpretable
  - Hard to debug
- NMT is difficult to control
  - For example, can't easily specify rules or guidelines for translation
  - Safety concerns!

SMT is still very much in use!

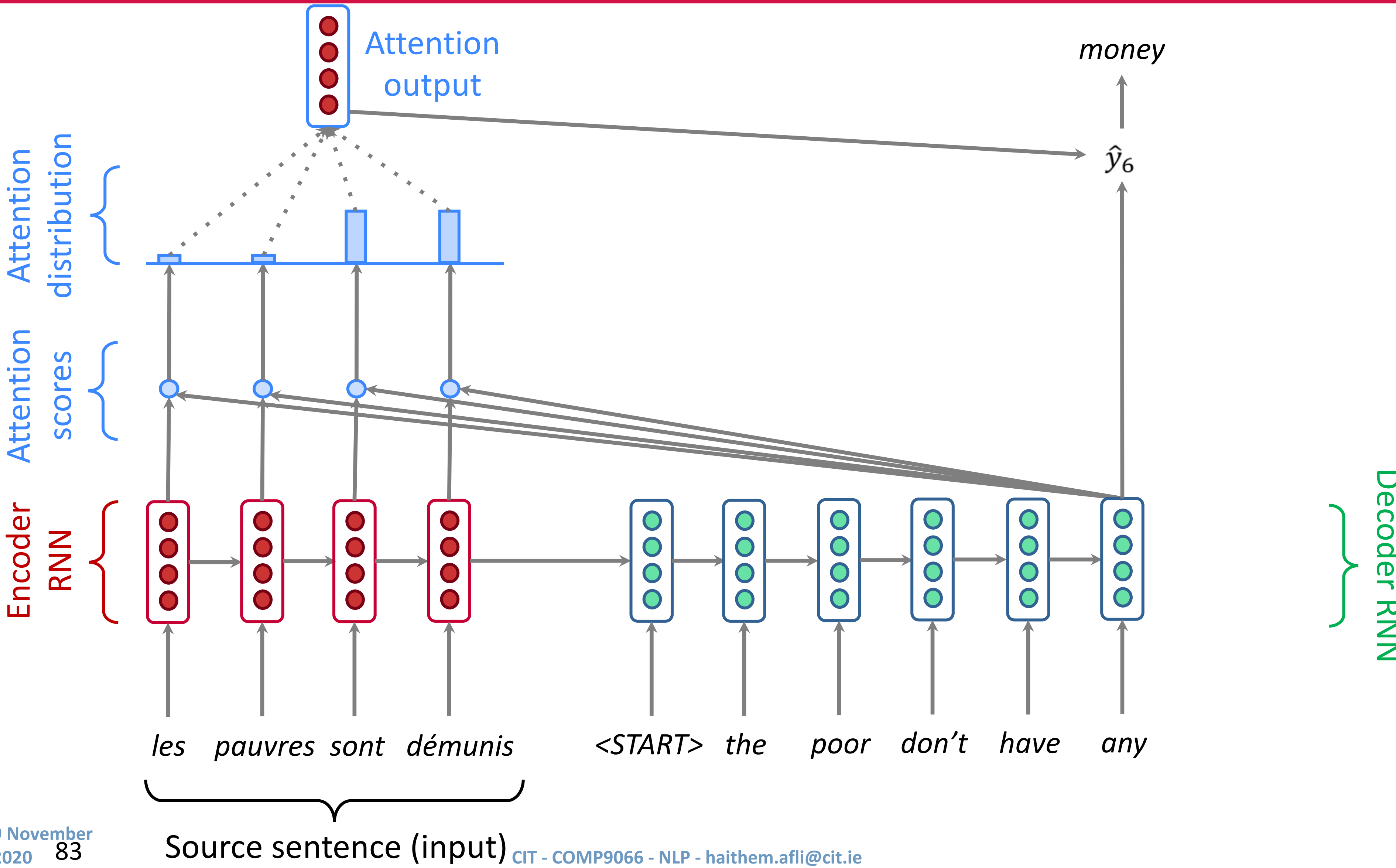
NMT research continues …

NMT is the **flagship task** for NLP Deep Learning

- NMT research has pioneered many of the recent innovations of NLP Deep Learning
- In **2018**: NMT research continues to thrive
  - Researchers have found *many, many* improvements to the "vanilla" seq2seq NMT system we've presented today
  - But one improvement is so integral that it is the new vanilla…

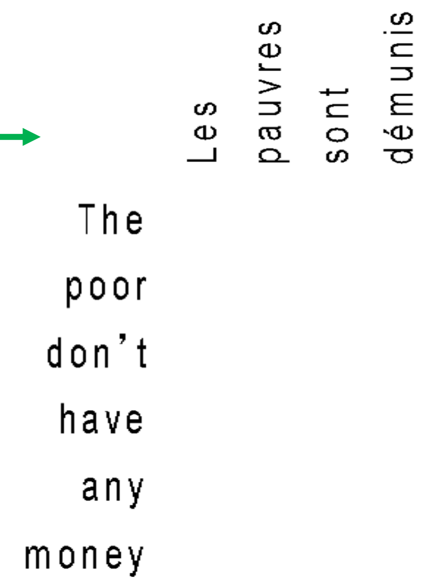
ATTENTION

Sequence-to-sequence with attention



Attention is great

- Attention significantly improves NMT performance
  - It’s very useful to allow decoder to focus on certain parts of the source
- Attention solves the bottleneck problem
  - Attention allows decoder to look directly at source; bypass bottleneck
- Attention provides some interpretability
  - By inspecting attention distribution, we can see what the decoder was focusing on
  - We get alignment for free!
  - This is cool because we never explicitly trained an alignment system
  - The network just learned alignment by itself



How Much Attention Do You Need? ..

You May Not Need Attention

Ofir Press[♠] **Noah A. Smith**[♠♣]
[♠]Paul G. Allen School of Computer Science & Engineering, University of Washington
[♣]Allen Institute for Artificial Intelligence
{ofirp,nasmith}@cs.washington.edu

0.13409v1 [cs.CL] 31 Oct 2018

Abstract

In NMT, how far can we get without attention and without separate encoding and decoding? To answer that question, we introduce a recurrent neural translation model that does not use attention and does not have a separate encoder and decoder. Our **eager translation model** is low-latency, writing target tokens as soon as it reads the first source token, and uses constant memory during decoding. It performs on par with the standard attention-based model of Bahdanau et al. (2014), and better on long sentences.[1]

1 Introduction

Nearly all actively-researched NMT models have the following properties:

- The decoder uses an attention mechanism over the source sequence representations. (Bahdanau et al., 2014; Luong et al., 2015; Vaswani et al., 2017).
- The encoder and decoder are two different

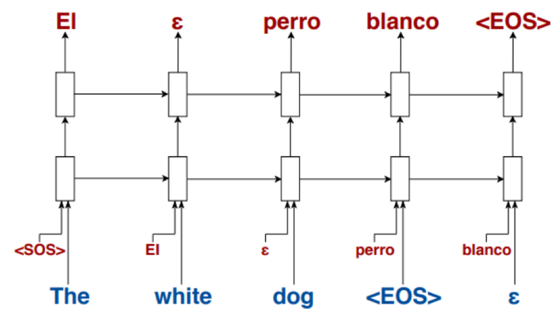


Figure 1: The eager model translating the sentence "The white dog" into Spanish. Source (target) tokens are in blue (red). ε is the padding token, which is removed during postprocessing. The diagram presents an eager translation model with two LSTM layers.

and can finish translating soon after the last source word is read.

The **eager translation model** uses a constant amount of memory, since it needs to use only one previous hidden state (rather than all previous hidden states) at every timestep. Instead of "cram-

Next Week

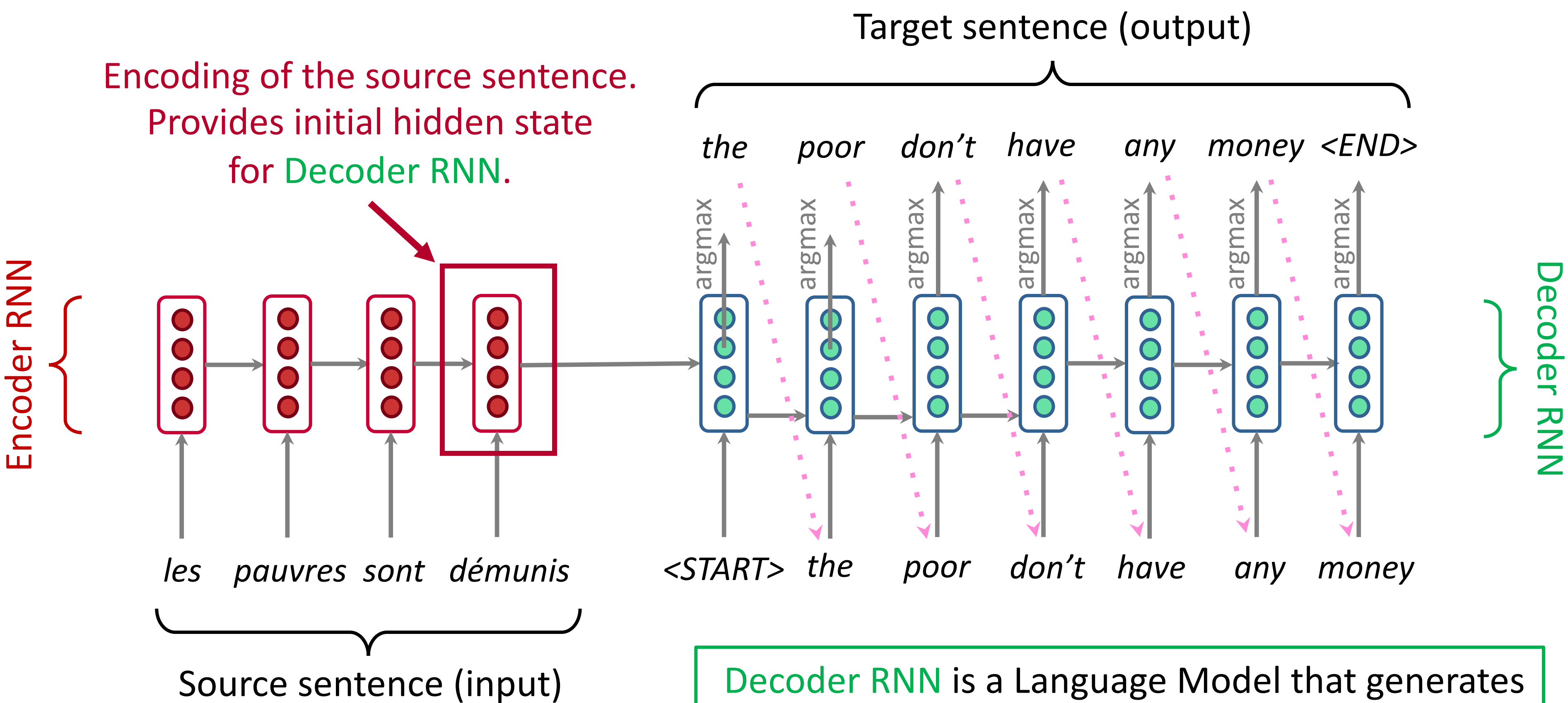
Machine Translation Evaluation

Discussion

87

Neural Machine Translation (NMT)

The sequence-to-sequence model

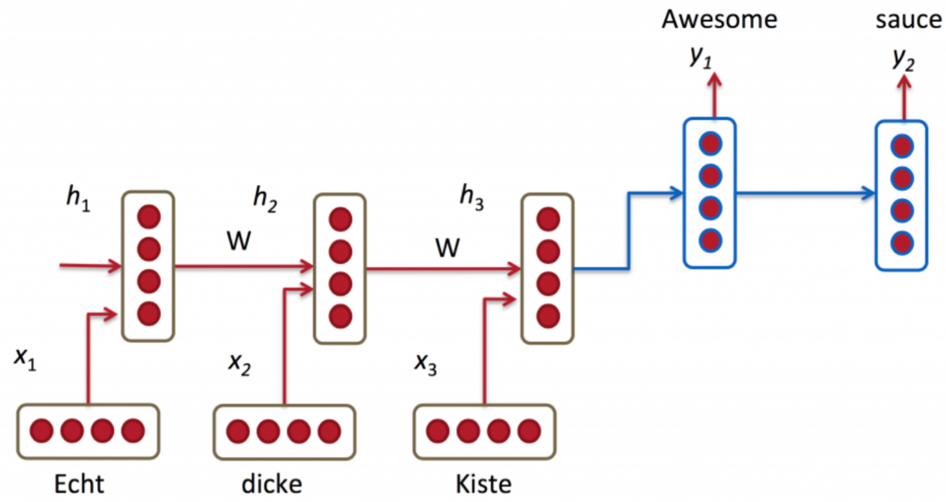


Decoder RNN is a Language Model that generates target sentence conditioned on encoding.

Encoder RNN produces an encoding of the source sentence.

Note: This diagram shows **test time** behavior: decoder output is fed in ···➔ as next step's input

Note there's an embedding layer for each CNN
You have two separate sets of word embeddings, for the two languages

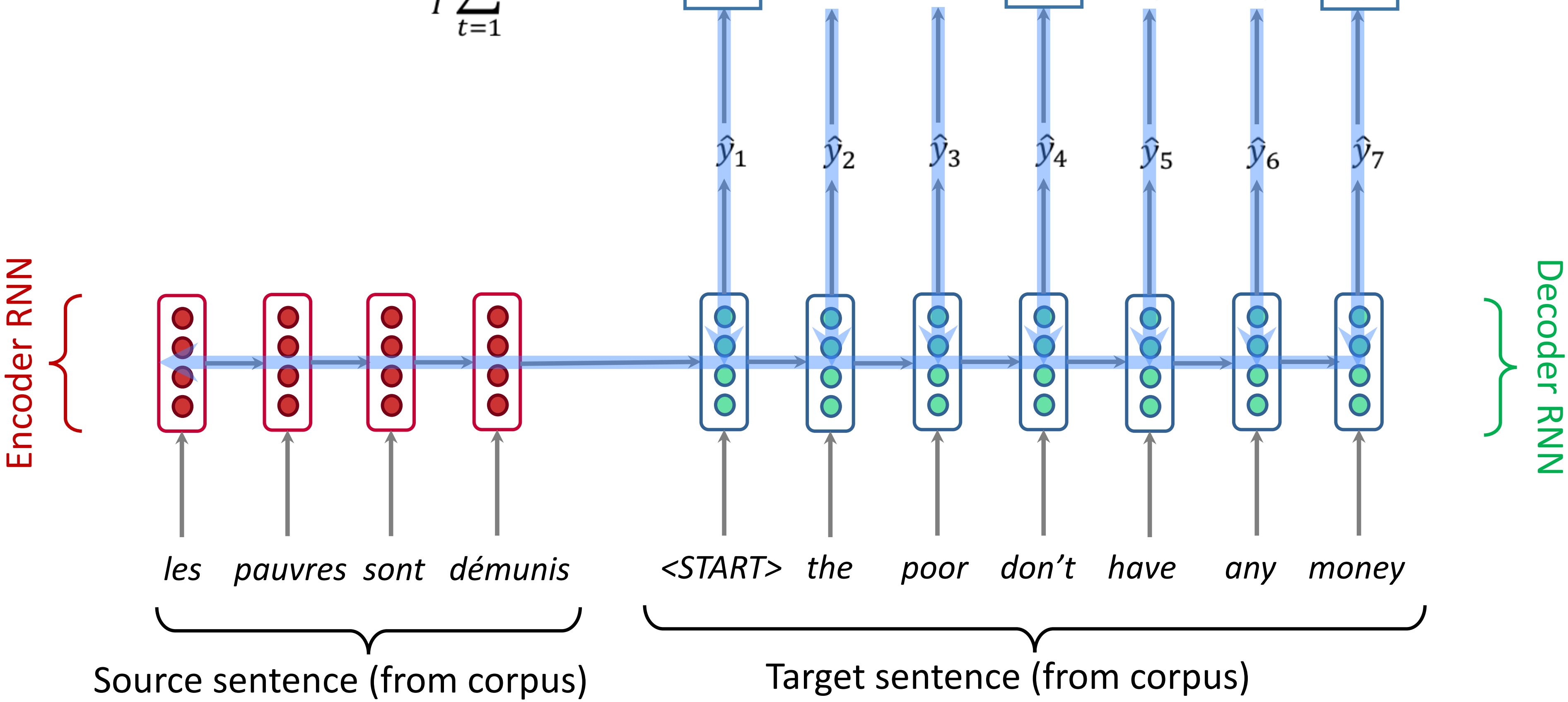


RNN for Machine Translation. Image Source: http://cs224d.stanford.edu/lectures/CS224d-Lecture8.pdf

Training a Neural Machine Translation system

= negative log prob of “the”   = negative log prob of “have”   = negative log prob of <END>

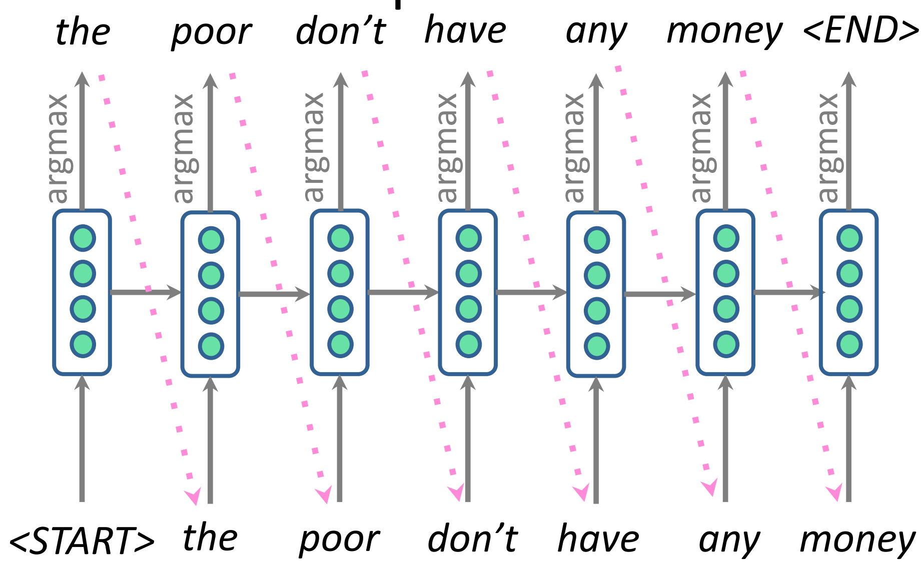
$$J = \frac{1}{T}\sum_{t=1}^{T} J_t \quad = \quad \boxed{J_1} + J_2 + J_3 + \boxed{J_4} + J_5 + J_6 + \boxed{J_7}$$



Seq2seq is optimized as a **single system.**
Backpropagation operates *“end-to-end”*.

Better-than-greedy decoding?

- We showed how to generate (or "decode") the target sentence by taking argmax on each step of the decoder



- This is greedy decoding (take most probable word on each step)
- **Problems?**

Better-than-greedy decoding?

- Greedy decoding has no way to undo decisions!
  - *les pauvres sont démunis (the poor don't have any money)*
  - → *the ____*
  - → *the poor ____*
  - → *the poor are ____*

- Better option: use beam search (a search algorithm) to explore *several* hypotheses and select the best one

Beam search decoding

- $P(y|x) = P(y_1|x)\, P(y_2|y_1, x)\, P(y_3|y_1, y_2, x) \ldots, P(y_T|y_1, \ldots, y_{T-1}, x)$

- We could try enumerating all *y* → too expensive!
  - Complexity $O(V^T)$ where *V* is vocab size and *T* is target sequence length

- Beam search: On each step of decoder, keep track of the *k* most probable partial translations
  - *k* is the beam size (in practice around 5 to 10)
  - Not guaranteed to find optimal solution
  - But much more efficient!

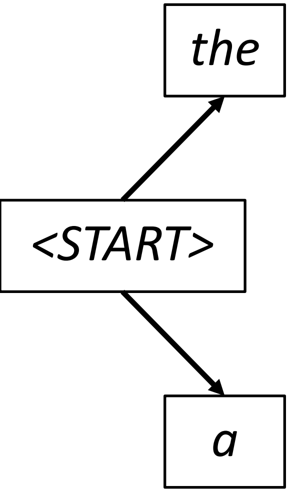
Beam search decoding: example

Beam size = 2

<START>

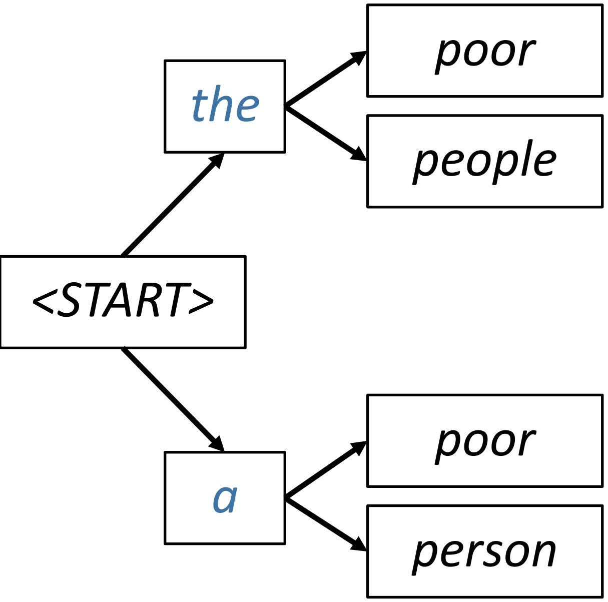
Beam search decoding: example

Beam size = 2



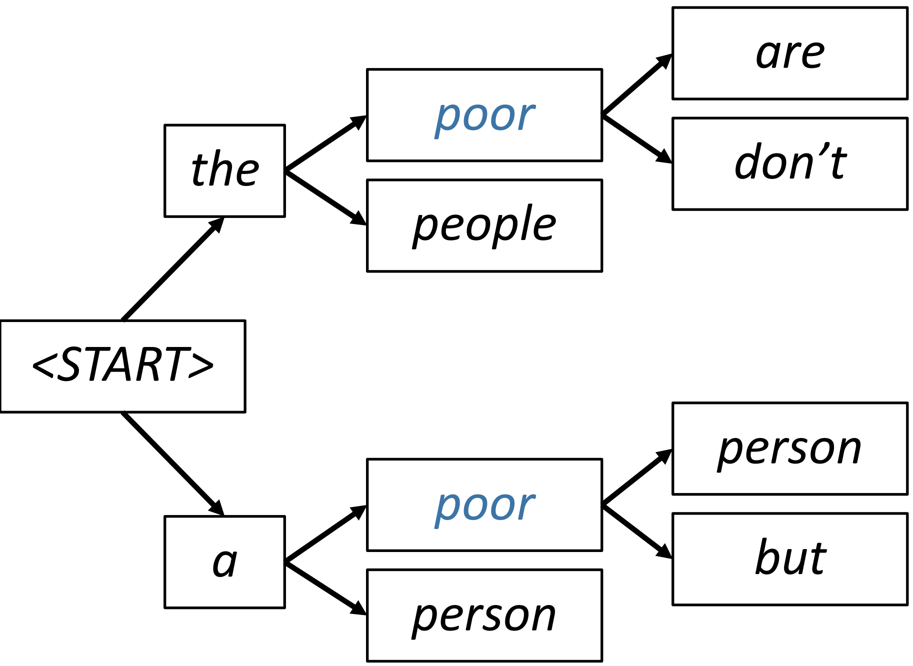
Beam search decoding: example

Beam size = 2



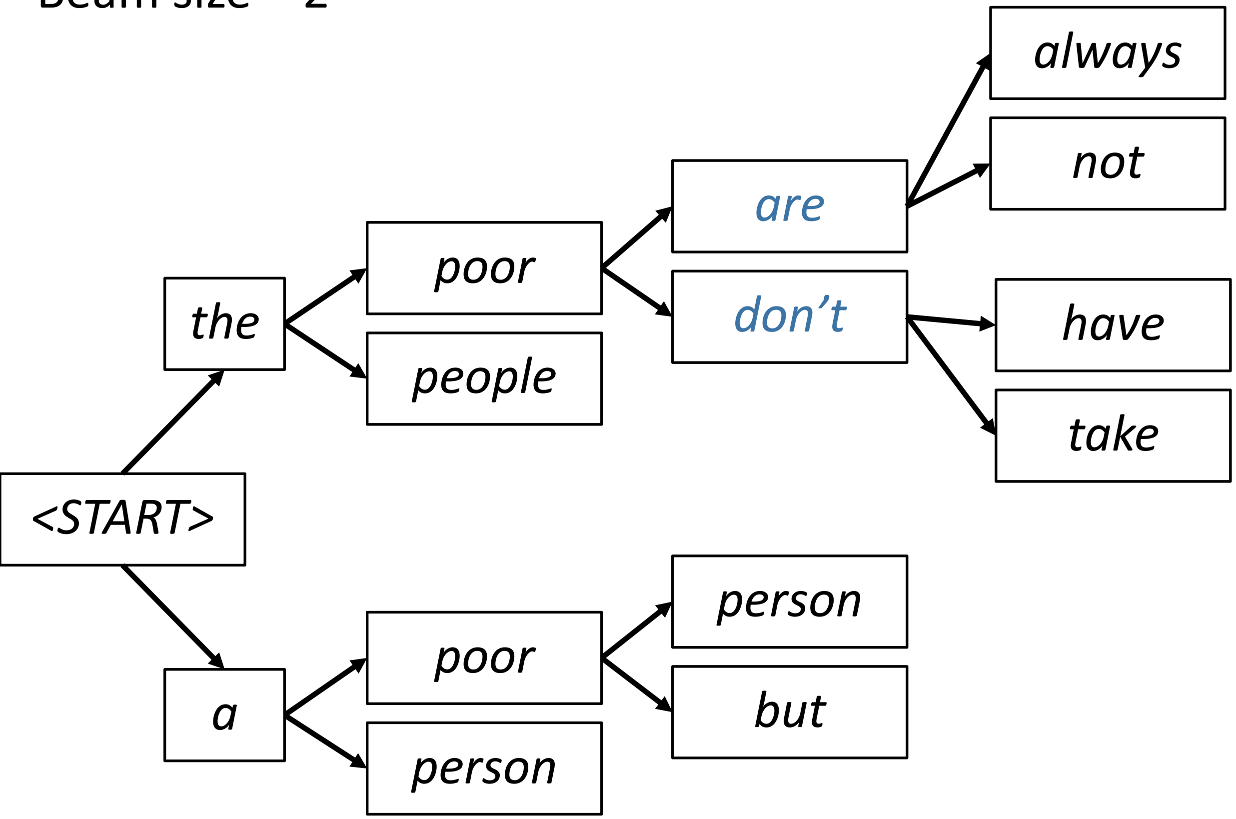
Beam search decoding: example

Beam size = 2



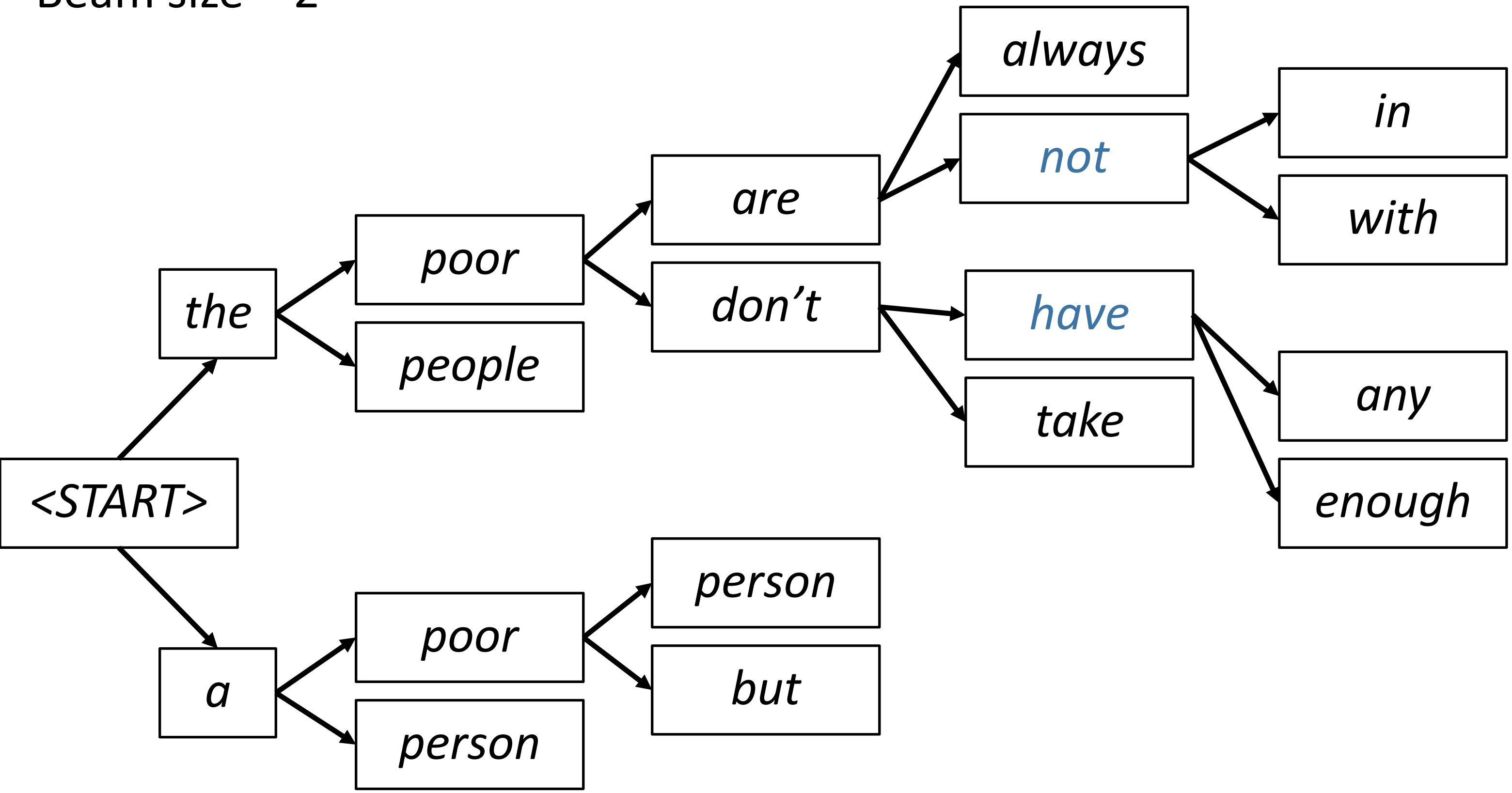
Beam search decoding: example

Beam size = 2



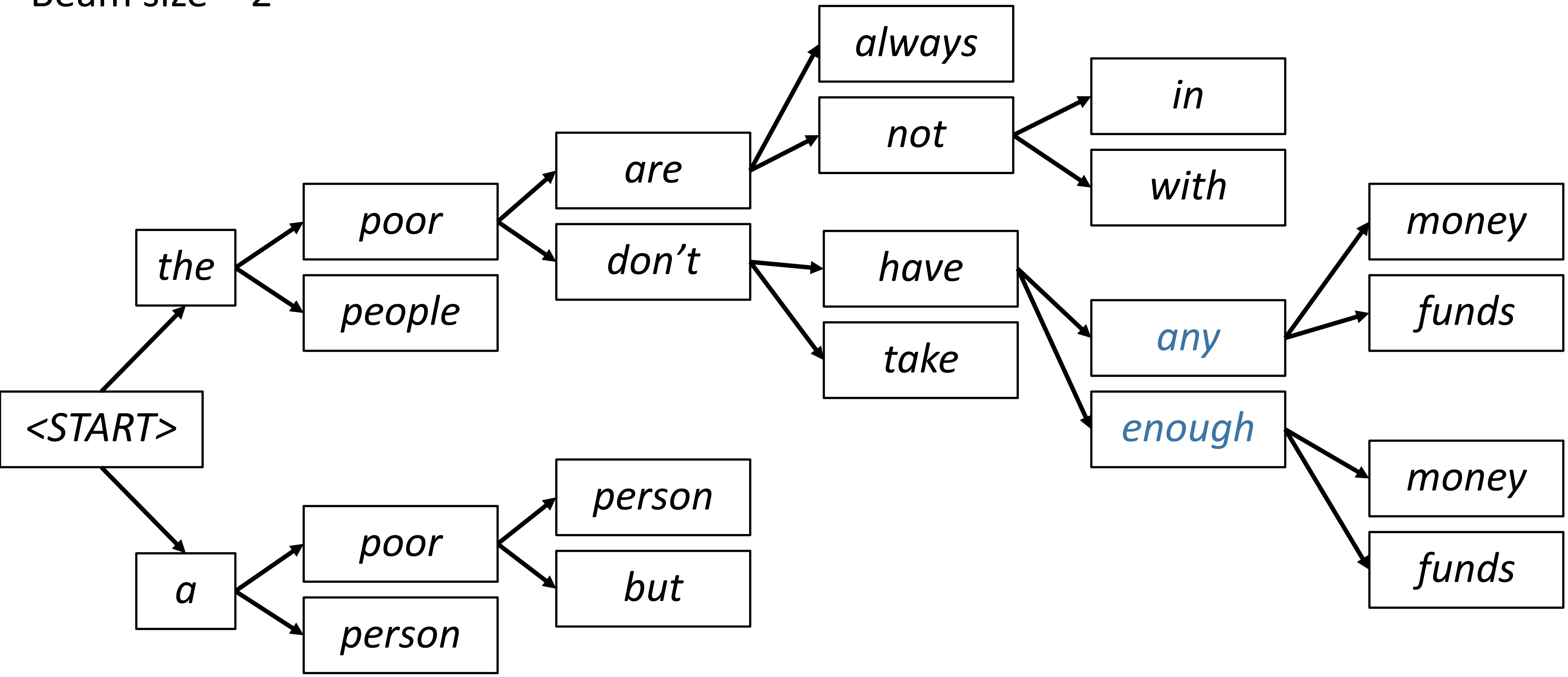
Beam search decoding: example

Beam size = 2



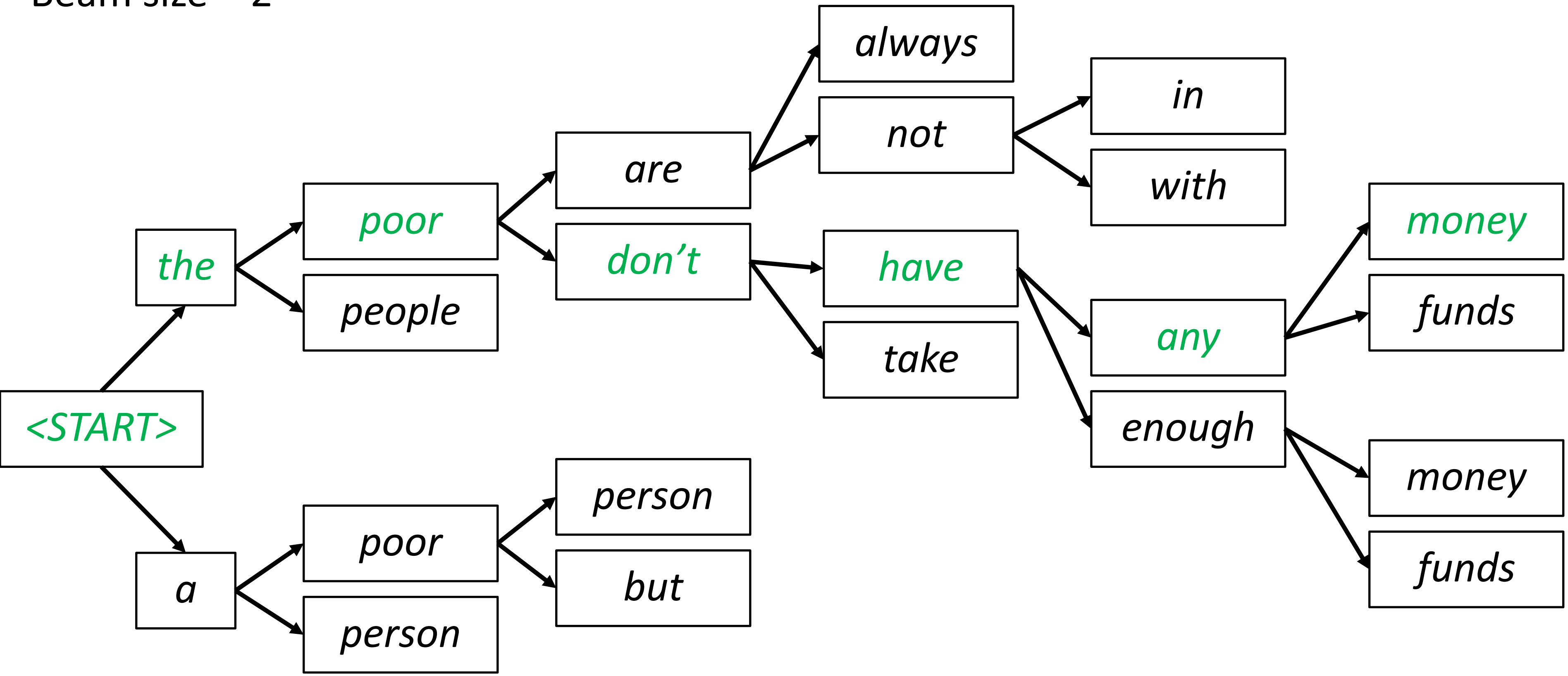
Beam search decoding: example

Beam size = 2



Beam search decoding: example

Beam size = 2



CORK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

Computer Science

http://www.cit.ie

Thank you

Haithem. afli@cit.ie

@AfliHaithem