ARE LANGUAGE TECHNOLOGIES A NEMESIS FOR HUMAN LANGUAGES?

Rafael E. Banchs

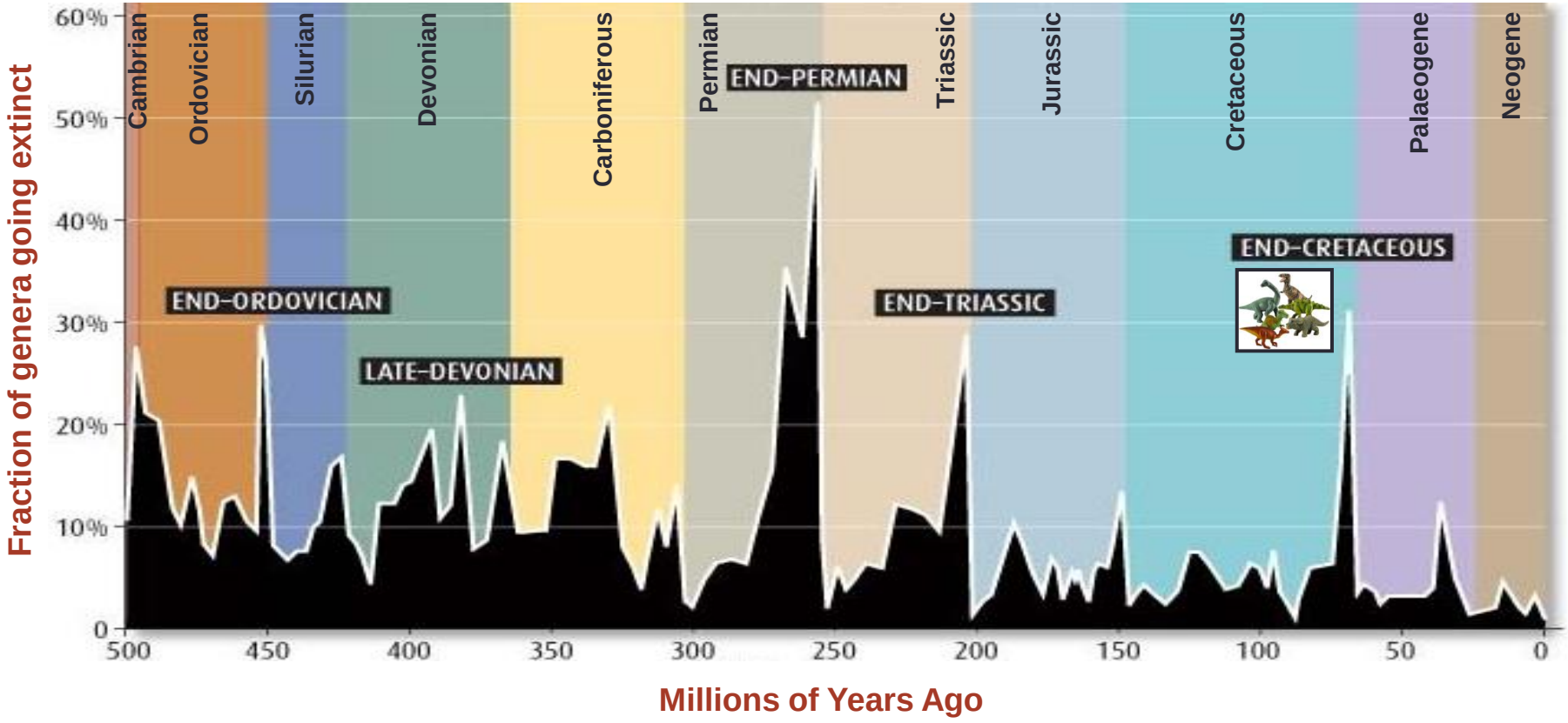
IV CENTENARIO CERVANTES

Spanish Language & Book Day

Central Public Library, Singapore
Saturday, 23rd of April 2016

Mass Extinction Events

- Extinction events occur periodically, every 26 to 30 million years.
- Five major extinction events are well documented!



(Image adapted from: https://universe-review.ca/I10-33-extinction.jpg)

Death Star “Nemesis”

- A hypothetical star called “Nemesis” is believed to be responsible for these periodic mass extinctions of life on earth!

(Image taken from: http://www.andylloyd.org/darkstarblog31.htm)

World Languages

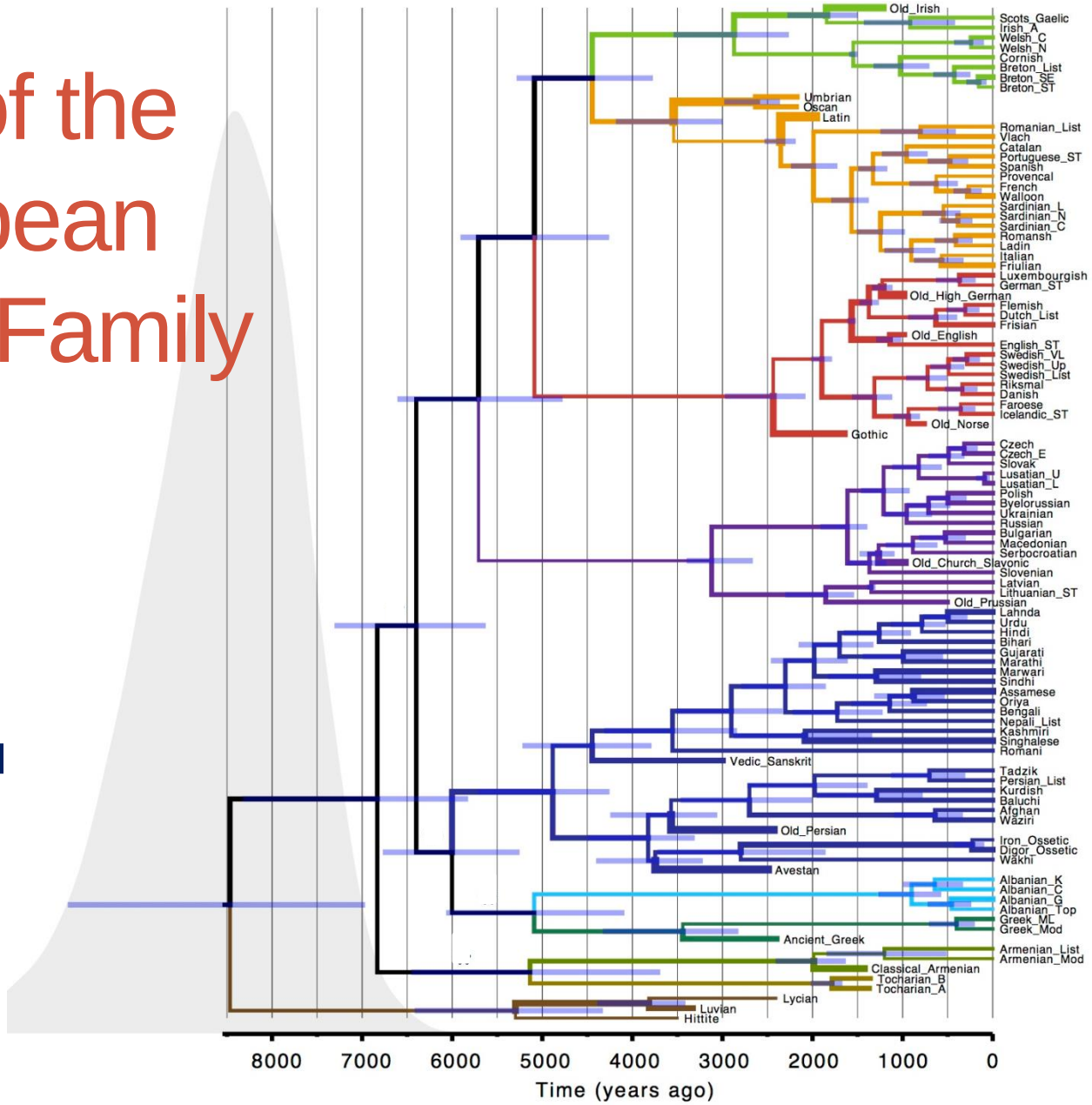
- Languages are like species… they evolve, develop and can go extinct!
- There is a total of 7,097 living languages in today's world, from which 918 are currently dying!

(Image taken from: http://archive.ethnologue.com/16/country_index.asp)

Evolution of the Indo-European Language Family

Main Branches:

- CELTIC
- ITALIC
- GERMANIC
- BALTO-SLAVIC
- INDIAN-IRANIAN
- ALBANIAN
- HELENIC
- ARMENIAN
- ANATOLIAN



(Taken from: https://theoreticalecology.wordpress.com/2012/08/24/mapping-the-origins-and-expansion-of-the-indo-european-language-family/)

Development and Endangerment

- Expanded Graded Intergenerational Disruption Scale (EGIDS)

Vigorous
The language is used for face-to-face communication by all generations and the situation is sustainable.

Developing
The language is in vigorous use, with literature in a standardized form being used by some though this is not yet widespread or sustainable.

Threatened
The language is used for face-to-face communication within all generations, but it is losing users.

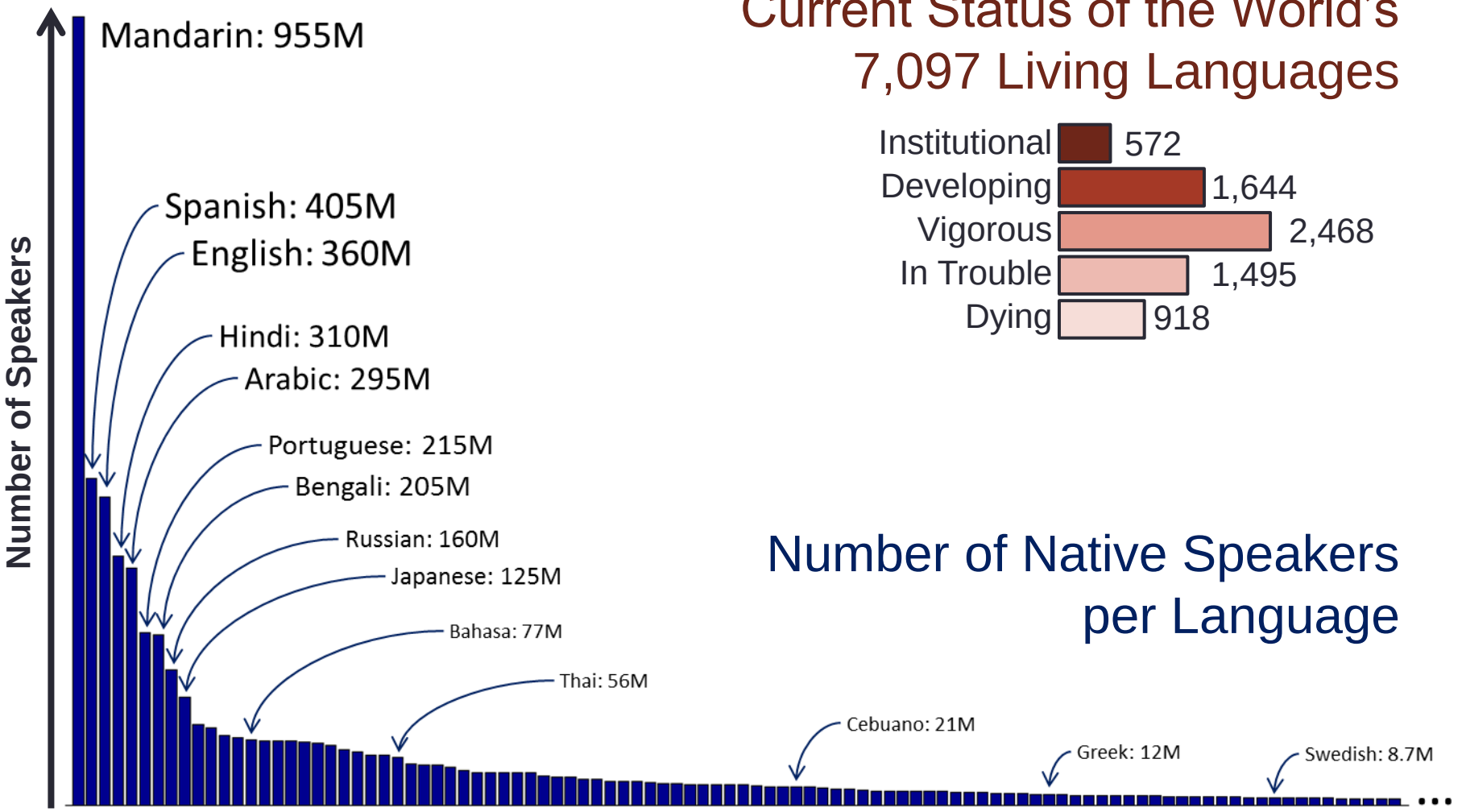
Shifting
The child-bearing generation can use the language among themselves, but it is not being transmitted to children.

Moribund
The only remaining active users of the language are members of the grandparent generation and older.

Extinct
The language is no longer used and no one retains a sense of ethnic identity associated with the language.

(Source: https://www.ethnologue.com/about/language-status)

Languages and Population



Current Status of the World's 7,097 Living Languages

Number of Native Speakers per Language

(Source: https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers/)

Languages and Geography

Geographical distribution of main institutional languages

(Adapted from: http://spanishsaborami.com/languages-in-the-world-the-top-5/)

Overview on Language Technologies

What are Language Technologies?

They refer to the use of computer systems to process, analyse and interpret human languages (a subfield of AI).

Different names, similar endeavours…

Natural Language Processing

Computational Linguistics

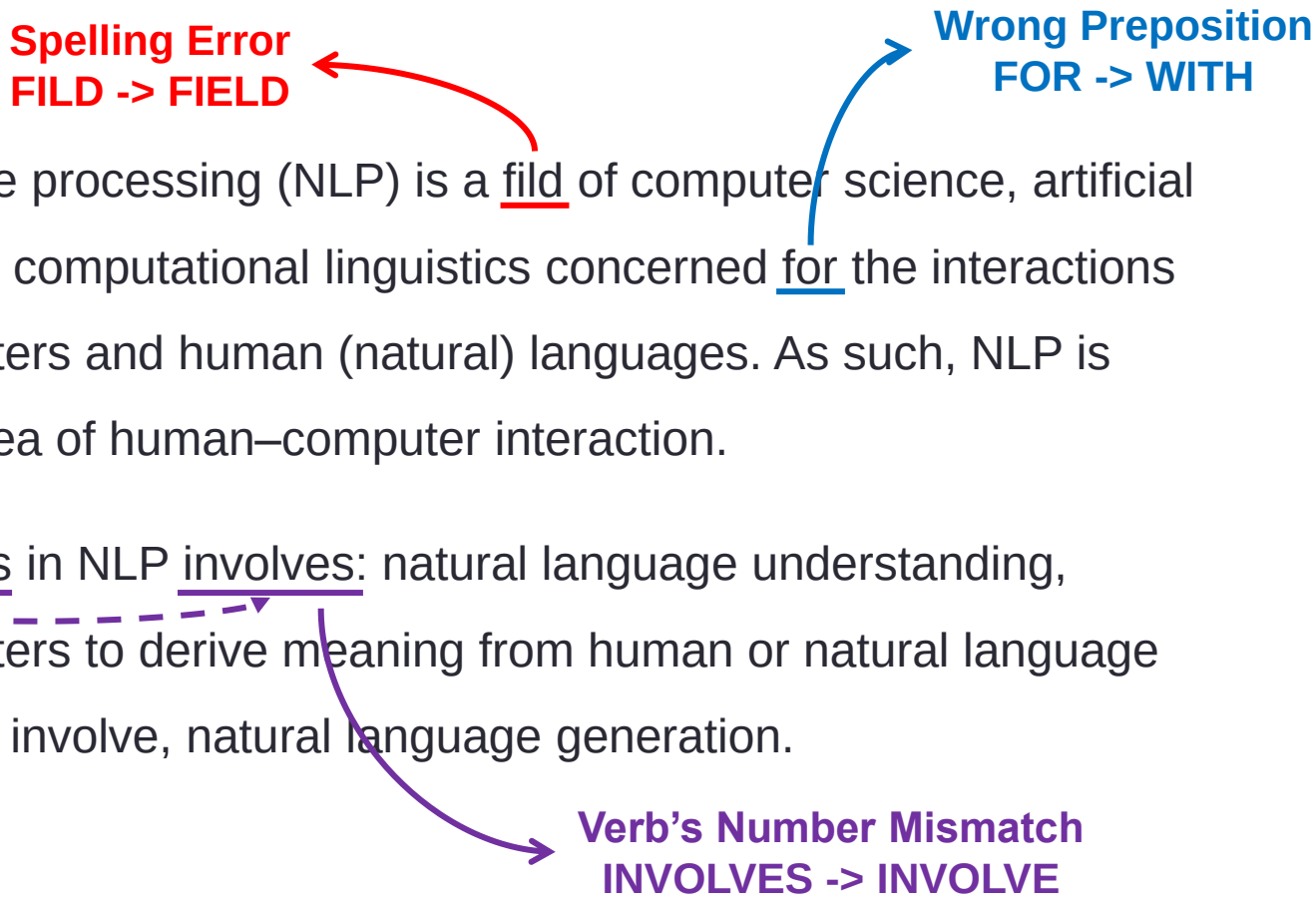
Text Mining

Speech Processing

Language Engineering

Applications: Text Correctors

Spell Checking and Grammar Correction

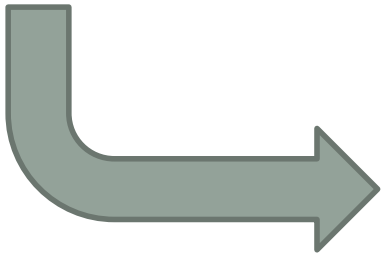


Applications: Information Retrieval

Search for Relevant Information over a Large Document Collection (such as the web)

Query:

"Natural Language Processing"



8,240,000 RESULTS Date Language Region

Language Processing - Natural Language Processing.
Ad · About.com/Language Processing
Find Expert Advice on About.com.
- Search About.com
- About.com Home
- The About.com Story
- 100+ Topics
- 1000 Expert Writers
- Trusted Guides

Natural language processing - Wikipedia, the free ...
https://en.wikipedia.org/wiki/Natural_language_processing
Natural language processing (NLP) is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between ...
History · **NLP** using machine ... · Major tasks in **NLP** · Statistical **NLP**

Natural Language Processing - Stanford University | **Coursera**
https://www.coursera.org/course/nlp
Video embedded · **Natural Language Processing** from **Stanford** University. In this class, you will learn fundamental algorithms and mathematical models for **processing natural language** ...

Natural Language Processing - Research at Google
research.google.com/pubs/**NaturalLanguageProcessing**.html
Natural Language Processing (NLP) research at Google focuses on algorithms that apply at scale, across **languages**, and across domains.

Natural Language Processing - Microsoft Research
research.microsoft.com/en-us/groups/**nlp**
Overview. The goal of the **Natural Language Processing (NLP)** group is to design and build software that will analyze, understand, and generate **languages** that humans ...

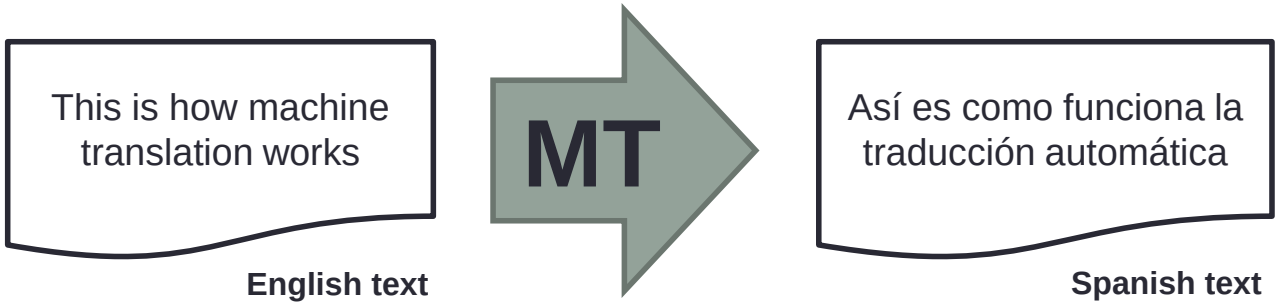
Applications: Speech Processing

- Automatic Speech Recognition (ASR): automatically transcribes speech into text
- Speech Synthesis or Text-to-Speech (TTS): produces speech from a text source

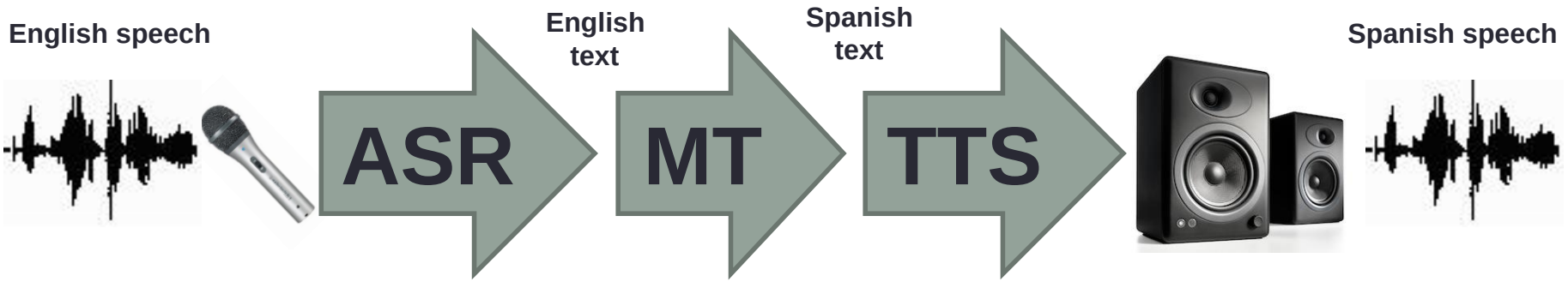
Applications: Machine Translation

Automatically translates from one language to another

- Text Translation

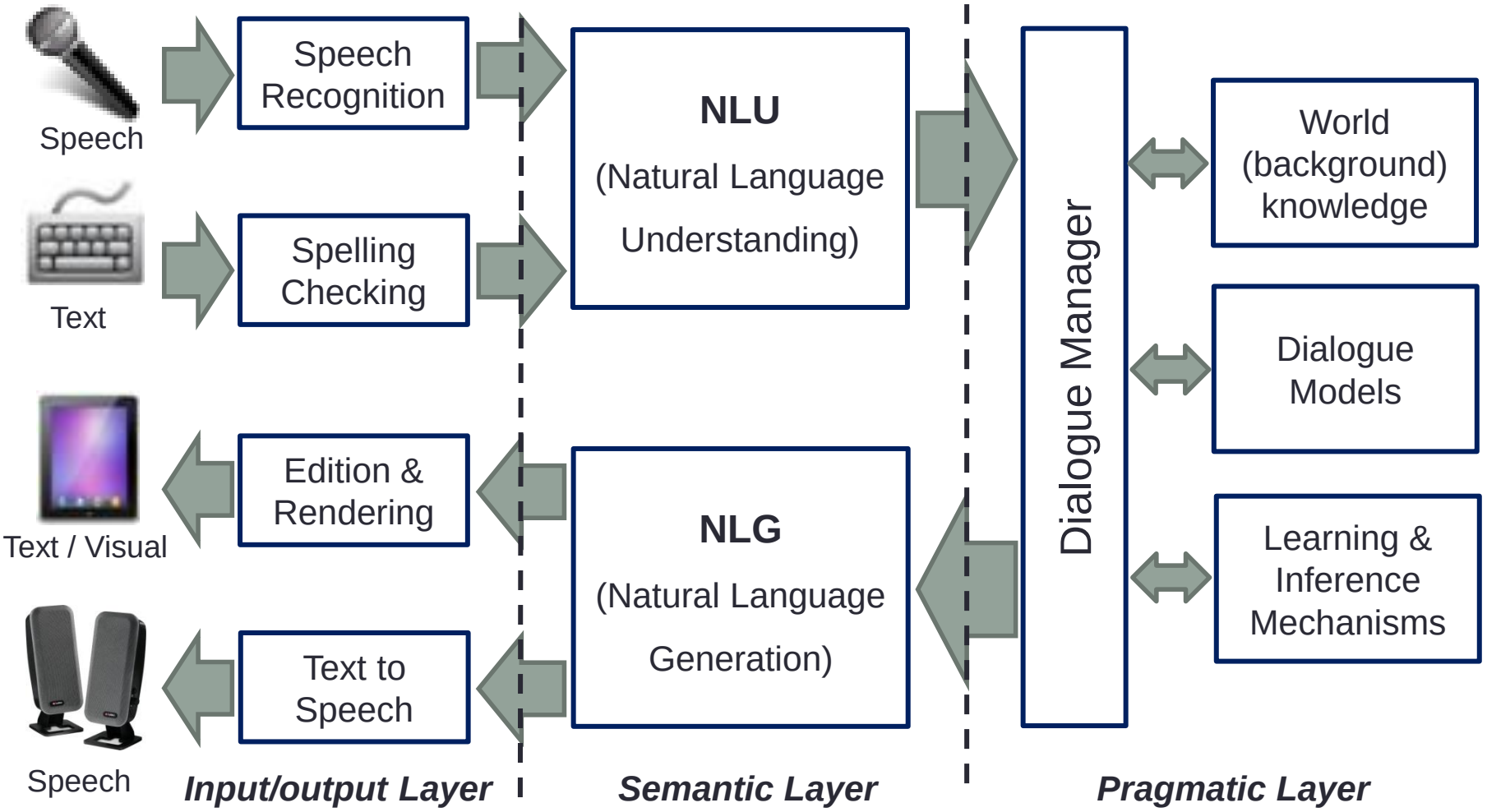


- Speech Translation



Applications: Dialogue Systems

Uses Natural Language Interaction to complete a task

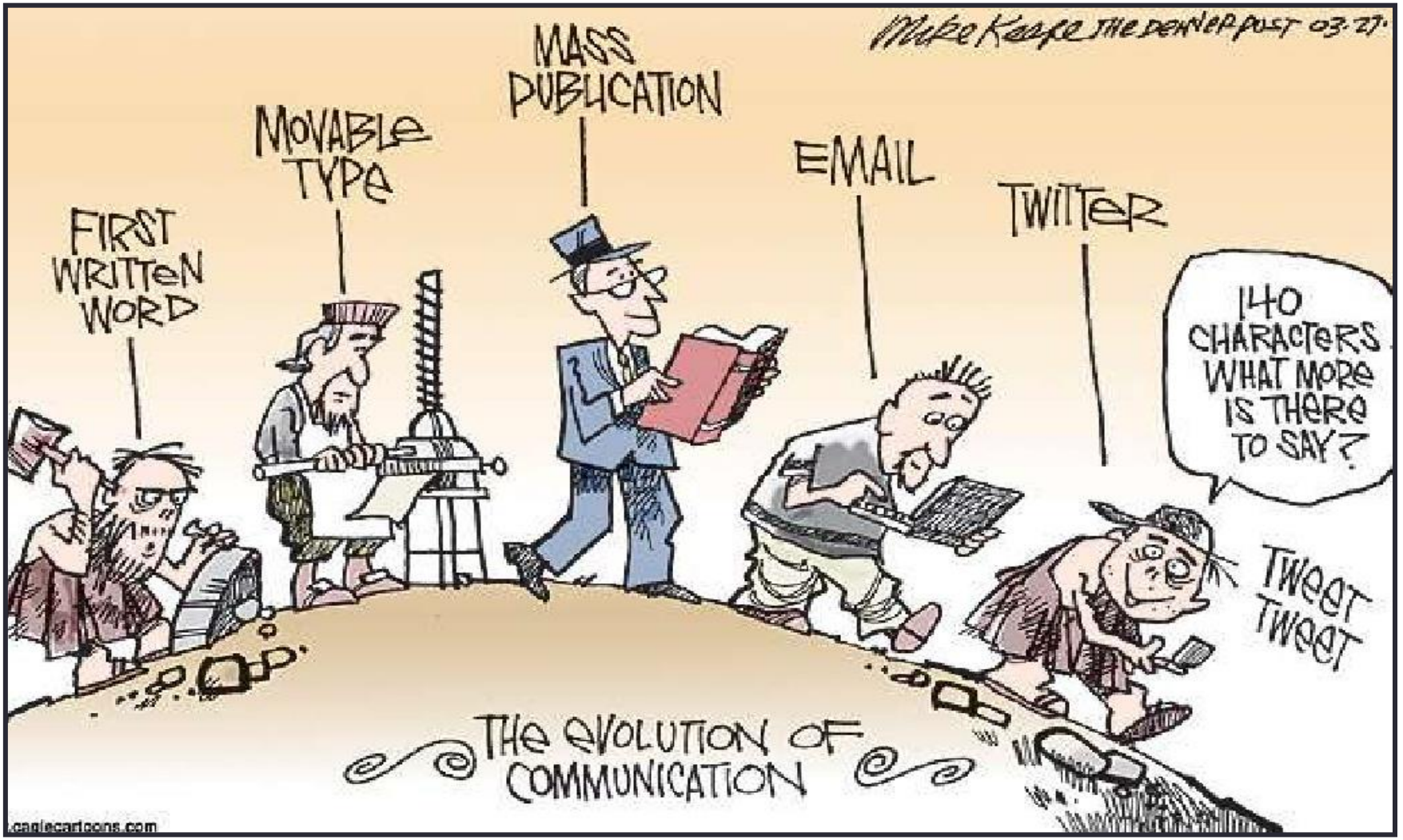


More Applications…

- **Document Categorization**
- **Document Summarization**
- **Opinion Mining**
- **Plagiarism Detection**
- **Information Extraction**
- **Sense Disambiguation**
- **Language Understanding**
- **Language Generation**
- **Transliteration**
- **Language Identification**
- **Speaker Identification**
- **Voice Conversion**
- **Emotion Detection**
- **Question Answering**
- **Discourse Analysis**

and many more…

So… is all this good or bad?



(Source: http://www.tedxsalford.co.uk/txting-a-branch-on-the-tree-of-language-evolution)

Some Valid and Interesting Opinions

- People do not learn how to spell anymore because of spelling correctors
- People will not be motivated to learn languages in the future because of machine translation
- People do not go to the library anymore because of the Internet

Other Valid and Interesting Opinions

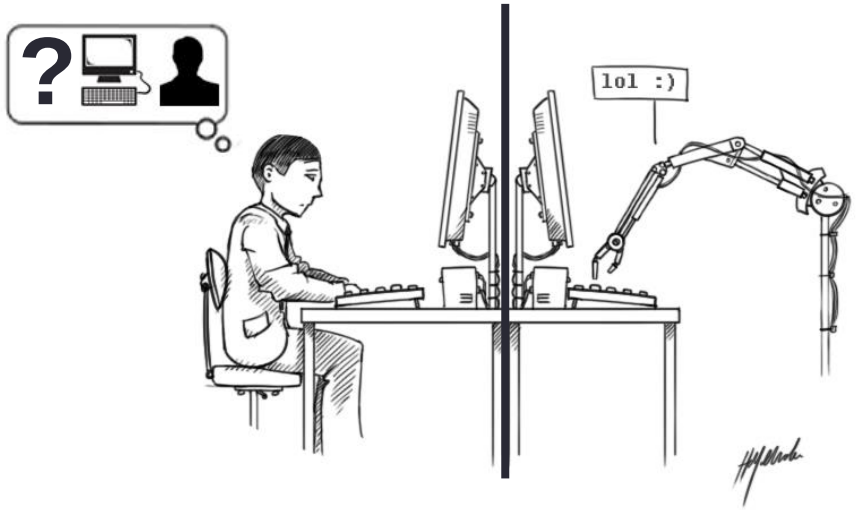
- People do not make spelling errors anymore thanks to spelling correctors
- People will be able to learn more languages in the future thanks to machine translation
- People still go to the library but can also go beyond it thanks to the Internet

A “Scaring” Opinion…

“Artificial Intelligence is a Threat to Humanity”

- Question 1: Can computers actually think?
- Question 2: Do computers really understand language?

Let us consider the “Turing Test”
(proposed by Alan Turing in 1950)



(Image adapted from: http://www.clubic.com/mag/culture/actualite-751397-imitation-game-alan-turing-pere-informatique.html)

The Turing Test in Action

Who is the human and who is the Chatbot?

A: How come in all that time I was trying to get you a date, you never asked me out?

B: Ok, do you want to go for dinner tonight?

A: Yes.

B: What kind of food do you like?

A: Anything but guts and mushrooms.

B: Ok pick you up at seven!

A: Ok. Pick me up at seven.

(Image taken from: http://www.chatbots.org)

The Turing Test in Action (cont.)

Who is the human and who is the Chatbot?

A: How come in all that time I was trying to get you a date, you never asked me out?

B: Ok, do you want to go for dinner tonight?

A: Yes.

B: What kind of food do you like?

A: Anything but guts and mushrooms.

B: Ok pick you up at seven!

A: Ok. Pick me up at seven.

B: So, we are having a date!

A: Really, and when is it?

(Image taken from: http://www.chatbots.org)

Do Computers Need to Understand?

A computer's approach to Machine Translation

LANGUAGE A	LANGUAGE B
ABA ISI ONO	ZIP UHM NUE YEP
OGO ESI ONO	YUK UHM NUE YEP
ABA ISI ATA SUAM	ZIP YAP UHM YIK
OGO ENE ESI ATA	YUK ZUR UHM YIK
ABA ISI ONO IKE ATA	ZIP UHM YIK ZOM NUE YEP
ABA ENE ISI SESU	ZIP ZUR UHM PIN CAN

How to say "yuk yap uhm pin can zom yik" in Language A?

Do Computers Need to Understand?

A computer's approach to Machine Translation

LANGUAGE A	LANGUAGE B
ABA ISI ONO	ZIP UHM NUE YEP
OGO ESI ONO	YUK UHM NUE YEP
ABA ISI ATA SUAM	ZIP YAP UHM YIK
OGO ENE ESI ATA	YUK ZUR UHM YIK
ABA ISI ONO IKE ATA	ZIP UHM YIK ZOM NUE YEP
ABA ENE ISI SESU	ZIP ZUR UHM PIN CAN

SUAM : YAP*
IKE : ZOM
SESU : PIN CAN

How to say "yuk yap uhm pin can zom yik" in Language A?

Do Computers Need to Understand?

A computer's approach to Machine Translation

LANGUAGE A	LANGUAGE B
ABA ISI ONO	ZIP UHM NUE YEP
OGO ESI ONO	YUK UHM NUE YEP
ABA ISI ATA SUAM	ZIP YAP UHM YIK
OGO ENE ESI ATA	YUK ZUR UHM YIK
ABA ISI ONO IKE ATA	ZIP UHM YIK ZOM NUE YEP
ABA ENE ISI SESU	ZIP ZUR UHM PIN CAN

SUAM : YAP*
IKE : ZOM
SESU : PIN CAN
ABA : ZIP
OGO : YUK

How to say "yuk yap uhm pin can zom yik" in Language A?

Do Computers Need to Understand?

A computer's approach to Machine Translation

LANGUAGE A	LANGUAGE B
ABA ISI ONO	ZIP UHM NUE YEP
OGO ESI ONO	YUK UHM NUE YEP
ABA ISI ATA SUAM	ZIP YAP UHM YIK
OGO ENE ESI ATA	YUK ZUR UHM YIK
ABA ISI ONO IKE ATA	ZIP UHM YIK ZOM NUE YEP
ABA ENE ISI SESU	ZIP ZUR UHM PIN CAN

SUAM : YAP*
IKE : ZOM
SESU : PIN CAN
ABA : ZIP
OGO : YUK
ENE : ZUR
ATA : YIK**

How to say "yuk yap uhm pin can zom yik" in Language A?

Do Computers Need to Understand?

A computer's approach to Machine Translation

LANGUAGE A	LANGUAGE B
ABA ISI ONO	ZIP UHM NUE YEP
OGO ESI ONO	YUK UHM NUE YEP
ABA ISI ATA SUAM	ZIP YAP UHM YIK
OGO ENE ESI ATA	YUK ZUR UHM YIK
ABA ISI ONO IKE ATA	ZIP UHM YIK ZOM NUE YEP
ABA ENE ISI SESU	ZIP ZUR UHM PIN CAN

SUAM : YAP*
IKE : ZOM
SESU : PIN CAN
ABA : ZIP
OGO : YUK
ENE : ZUR
ATA : YIK**
ONO : NUE YEP**

How to say "yuk yap uhm pin can zom yik" in Language A?

Do Computers Need to Understand?

A computer's approach to Machine Translation

LANGUAGE A	LANGUAGE B
ABA ISI ONO	ZIP UHM NUE YEP
OGO ESI ONO	YUK UHM NUE YEP
ABA ISI ATA SUAM	ZIP YAP UHM YIK
OGO ENE ESI ATA	YUK ZUR UHM YIK
ABA ISI ONO IKE ATA	ZIP UHM YIK ZOM NUE YEP
ABA ENE ISI SESU	ZIP ZUR UHM PIN CAN

SUAM : YAP*
IKE : ZOM
SESU : PIN CAN
ABA : ZIP
OGO : YUK
ENE : ZUR
ATA : YIK**
ONO : NUE YEP**
(ABA) ISI : UHM
(OGO) ESI : UHM

How to say "yuk yap uhm pin can zom yik" in Language A?

Do Computers Need to Understand?

A computer's approach to Machine Translation

LANGUAGE A	LANGUAGE B
ABA ISI ONO	ZIP UHM NUE YEP
OGO ESI ONO	YUK UHM NUE YEP
ABA ISI ATA SUAM	ZIP YAP UHM YIK
OGO ENE ESI ATA	YUK ZUR UHM YIK
ABA ISI ONO IKE ATA	ZIP UHM YIK ZOM NUE YEP
ABA ENE ISI SESU	ZIP ZUR UHM PIN CAN

SUAM : YAP*
IKE : ZOM
SESU : PIN CAN
ABA : ZIP
OGO : YUK
ENE : ZUR
ATA : YIK**
ONO : NUE YEP**
(ABA) ISI : UHM
(OGO) ESI : UHM

How to say "yuk yap uhm pin can zom yik" in Language A?

"ogo esi ata ike sesu suam"

Surprise, Surprise…

We have just translated Chinese into Spanish!

SPANISH →		LANGUAGE A		LANGUAGE B →		CHINESE
también	=	SUAM	TOO	YAP	=	也
y	=	IKE	AND	ZOM	=	和
cerveza	=	SESU	BEER	PIN CAN	=	啤 酒
yo	=	ABA	I	ZIP	=	我
tú	=	OGO	YOU	YUK	=	你
no	=	ENE	DON'T	ZUR	=	不
té	=	ATA	TEA	YIK	=	茶
café	=	ONO	COFFEE	NUE YEP	=	咖 啡
bebo	=	(ABA) ISI	DRINK	UHM	=	喝
bebes	=	(OGO) ESI	DRINK	UHM	=	喝

"ogo esi ata ike sesu suam" : "yuk yap uhm pin can zom yik"

(tú bebes té y cerveza también) : (你 也 喝 啤 酒 和 茶)

Going Back to our Questions…

- Are language technologies good or bad?

  → Language technologies are neither good nor bad, it just depends on how we use them!

- Can computers understand language and think?

  → Definitively not… yet!

- Can Artificial Intelligence be a threat for humanity?

  → Definitively not… but human foolishness can!

A Well Documented Worrying Fact…

Economic Growth Threatens 25% Of World's Languages

(http://www.valuewalk.com/2014/09/economic-growth-threatens-25-worlds-languages/)

"We found that at the global scale, language speaker declines are strongly linked to economic growth — that is, declines are particularly occurring in economically developed regions"

Professor Tatsuya Amano, University of Cambridge

The United Nations […] stated that around half of the languages spoken around the world [will] face extinction by the end of the century if nothing is done to save them.

Language technologies can actually come to the rescue of endangered languages!

ARE LANGUAGE TECHNOLOGIES A NEMESIS FOR HUMAN LANGUAGES?

Rafael E. Banchs

IV CENTENARIO CERVANTES

Spanish Language & Book Day

Central Public Library, Singapore
Saturday, 23rd of April 2016