The Neural Revolution in the Translation Industry 3-year retrospective and future directions

ASLING 2019 – TC41

Jean Senellart, SYSTRAN President

Presentation Agenda

Understanding

Controlling

Using

SYSTRAN - 50 Years of Innovation in Translation Technologies

1968

Creating SYSTRAN: "SYS"tem "TRAN"slation

1975

SYSTRAN deploys automatic translation on the Apollo Soyuz mission to translate flight instructions from Russian to English.

1976

SYSTRAN is chosen by the European Commission to improve the productivity of translation of texts and regulations.

1997

SYSTRAN develops for Alta Vista Babel Fish: the first online translation site and equips all the following portals: Google, Yahoo!, Lycos, Wanadoo, Free…

2016

SYSTRAN launches the first neural translation engine and opens its technology in OpenNMT. Today, it is one of the leading translation technologies with Google, Microsoft, Facebook and Amazon.

2019

Launch of SYSTRAN Marketplace, the first platform combining the best of NMT & a community of experts

The legacy technologies

→ RBMT

Rule-Based System

1968-today: Original Approach
Relies on **Explicited Human Knowledge**

- **Predictability**
- Translation Consistency

→ SMT

Statistical Machine Translation

2008-2016
Relies on memorization of patterns observed in **large amount of data**

- **Translation fluency**
- Good at Local Ambiguity Resolution

+ Speed + Full Control on Customization	+ Training Overnight + Leverage Huge Amount of Data
- Lack of "pragmatic knowledge" - Requires huge amount of linguistic knowledge	- Low Generalization - Lack of "linguistic knowledge" - Requires huge amount of data

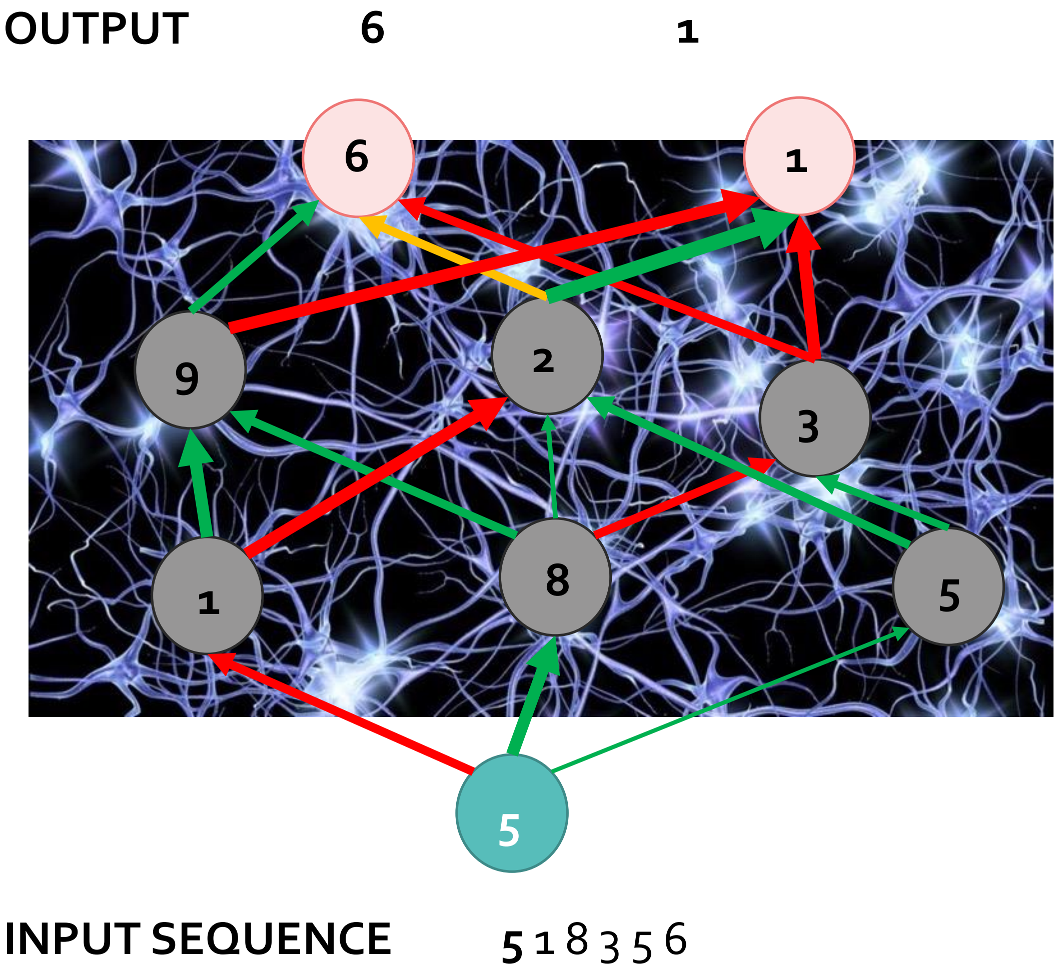
And then came Neural Machine Translation

The Basis - Artificial Neural Network (ANN)

- Self-learning technology inspired from human-brain

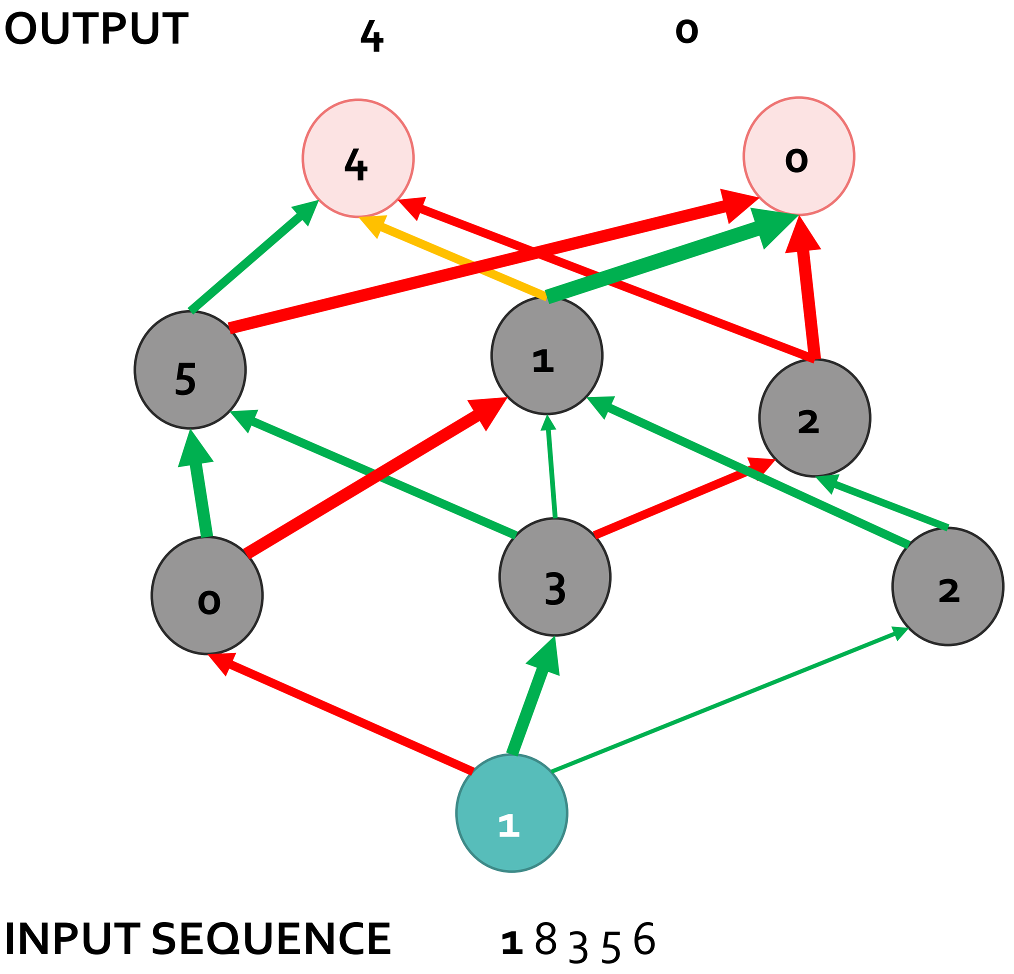
The Basis - Artificial Neural Network (ANN)

- Self-learning technology inspired from human-brain
- Composed of multiple layers
- Connected through weighted connections
- Each neuron is individually activated by the neurons it is connected to



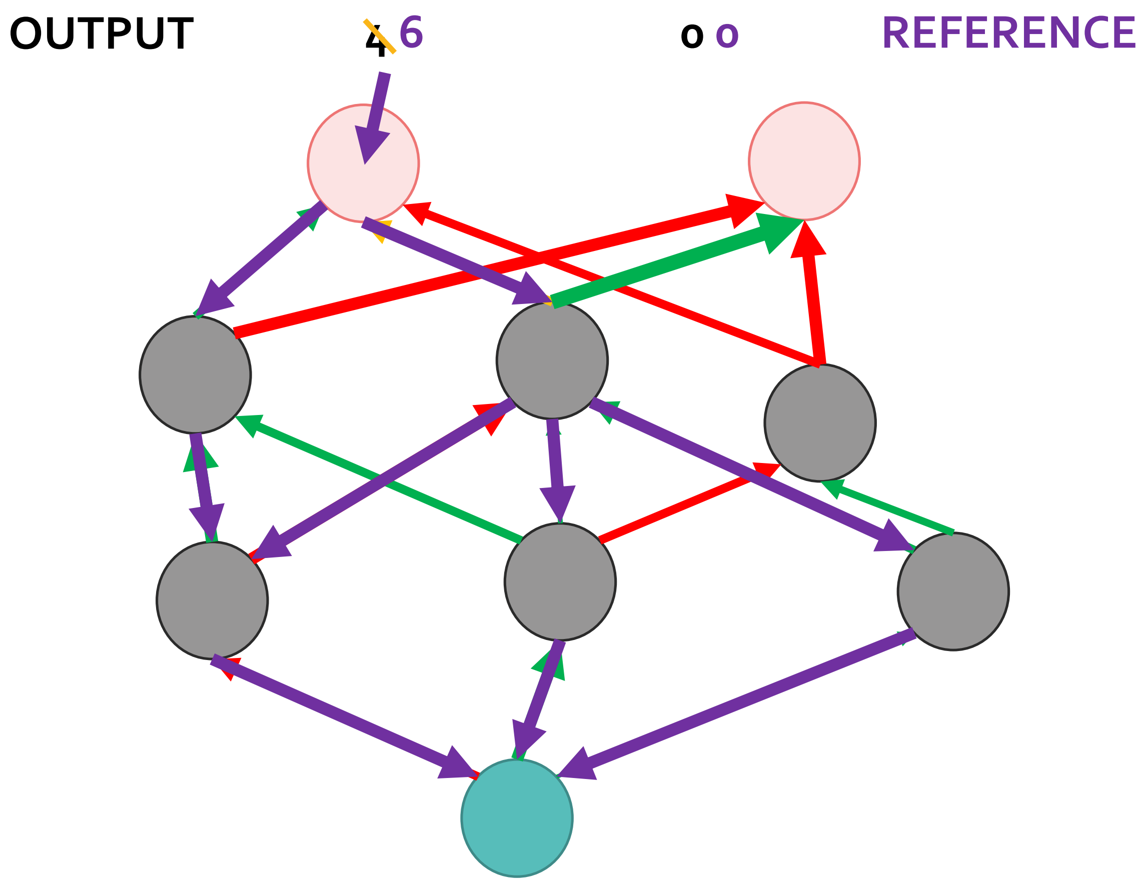
The Basis - Artificial Neural Network (ANN)

- Self-learning technology inspired from human-brain
- Composed of multiple layers
- Connected through weighted connections
- Each neuron is individually activated by the neurons it is connected to



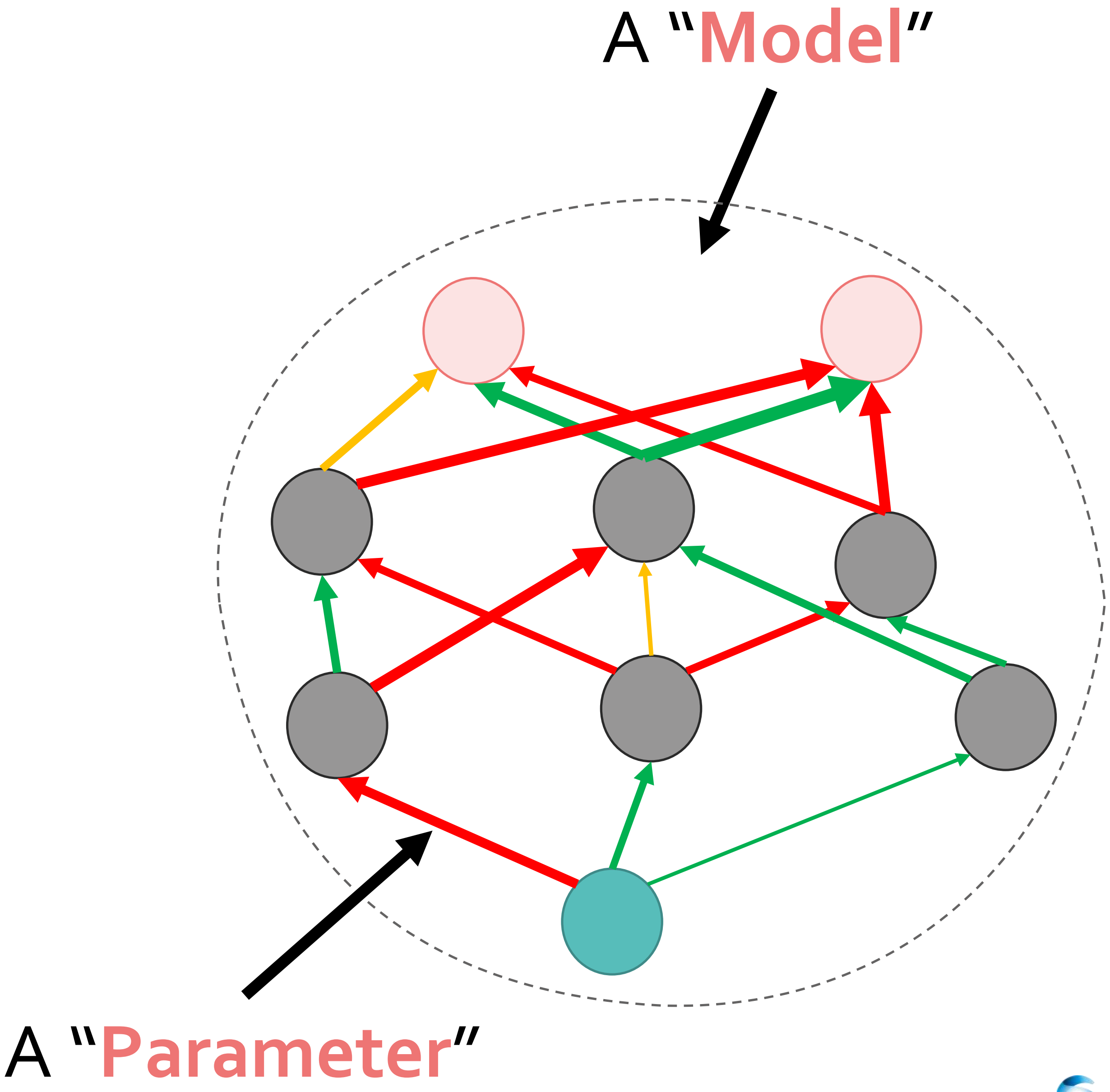
The Basis - Artificial Neural Network (ANN)

- Self-learning technology inspired from human-brain
- Composed of multiple layers
- Connected through weighted connections
- Each neuron is individually activated by the neurons it is connected to
- Output of the neural network is compared to a reference value and corrective feedback back-propagated through the network to correct weights



The Basis - Artificial Neural Network (ANN)

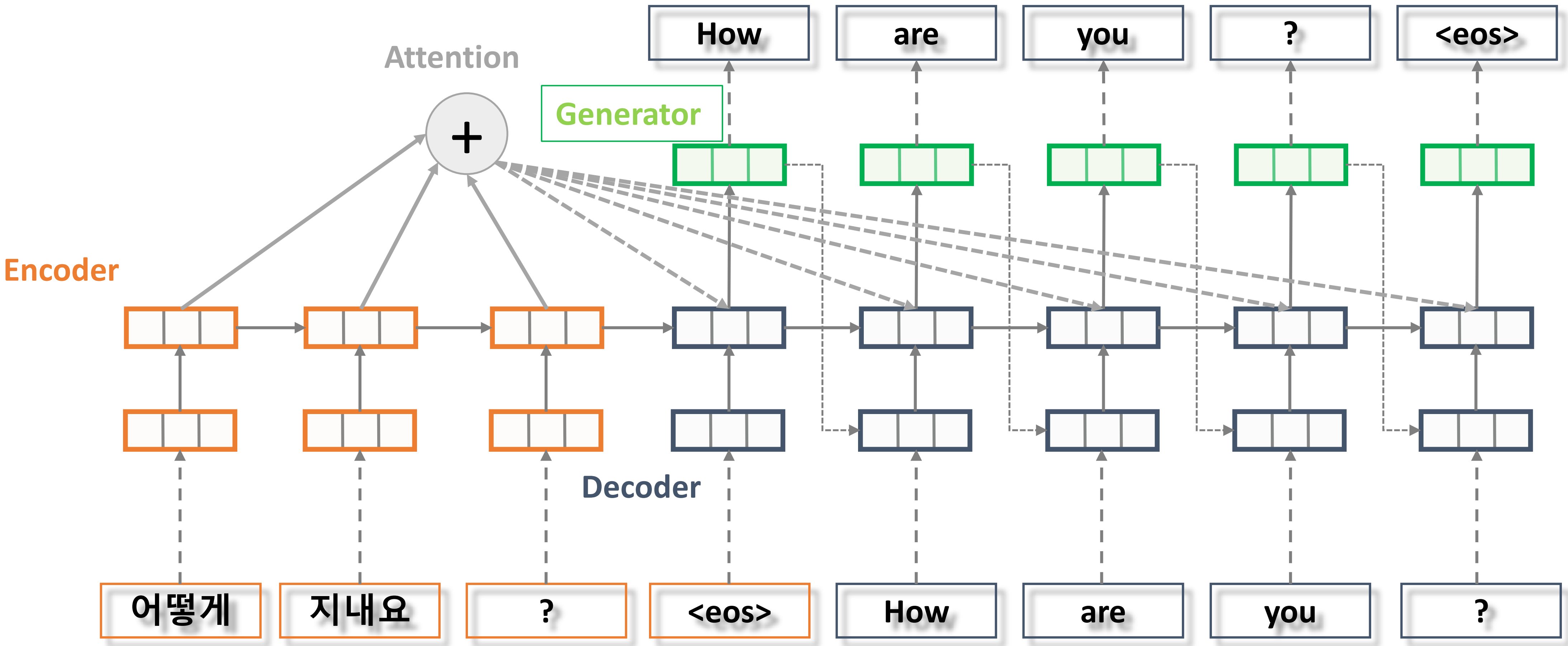
- Self-learning technology inspired from human-brain
- Composed of multiple layers
- Connected through weighted connections
- Each neuron is individually activated by the neurons it is connected to
- Output of the neural network is compared to a reference value and corrective feedback back-propagated through the network to correct weights



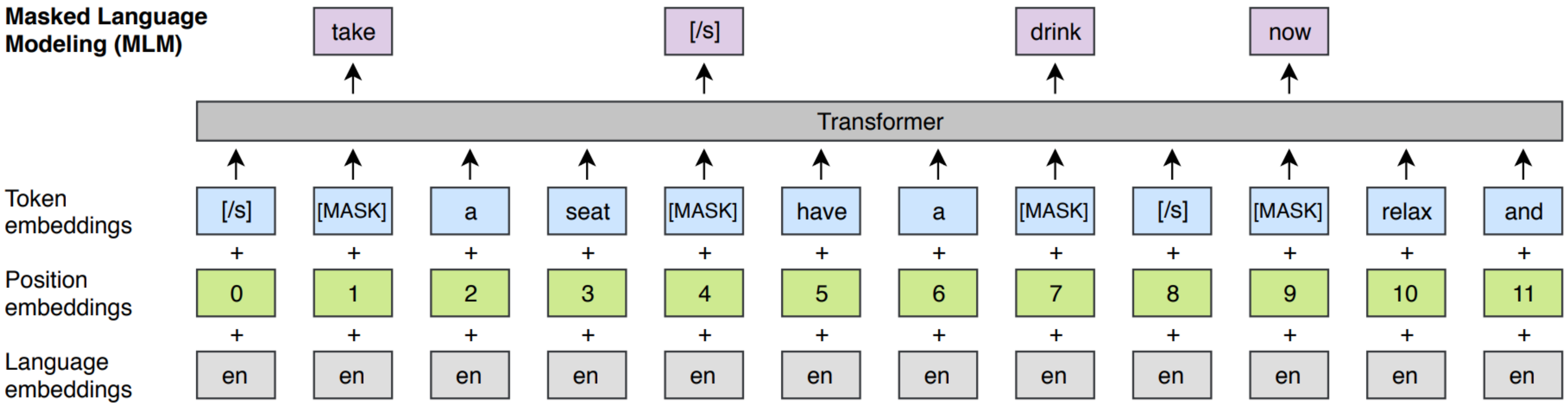
ANN for Natural Language Processing

- What is specific to Natural Language Processing?
  - Input are sequences of characters and/or words
  - Sentence/Document is arbitrarily long
  - Language Production is a combination of different source of information: Grammatical, Spelling, Semantic, Pragmatic, Style, Emotional…
  - Hard to find actual training data and reference can be subjective
- 6000+ human languages

All-Purpose Recipe – Seq2Seq-attn



Power of the encoder – Masked Language Model



→ **Train a standalone encoder with 2 simple objectives:**

+ Guess masked words
+ Decide if 2 consecutive sentences are logically connected

→ **Train massively on huge available monolingual data**

Masked Language Model – Example 1

Provided

AsLing (The International Association for Advancement in Language Technology) is delighted to host the 41th edition of Translating and the Computer Conference (TC41) on 21 and 22 November 2019 in London.

Generated

The conference offers attendees a comprehensive introduction to the practice and applications of the translingual techniques that are the cornerstone of many modern applications and applications in language technology. Speakers will be presenting their research, projects, current research projects, and any new ideas, papers and projects related to translingual applications, as well as providing an overview of existing translingual research.

https://talktotransformer.com

Masked Language Model – Example 2

"We are talking about the Neural Revolution in the translation industry," *the former chief of the Transforall translation services company told the FT. "You would not believe how many companies are still trying to get their stuff right. <u>It is so difficult, there are so many new words, so many new ways of using them, that it is overwhelming</u>."*

The challenge with these new technologies, according to the FT, is that they're not ready to go mainstream yet, and "as with any new technology, they are prone to disruption". So what happens next? In his column, Dutton suggests that if the government "seizes the moment" and creates a National Translator's Council, the government could help fund companies to roll out their services faster and with more confidence.

Dutton suggests that there could also be a role for the government in setting the standards and regulations for the services that are being offered to the public.

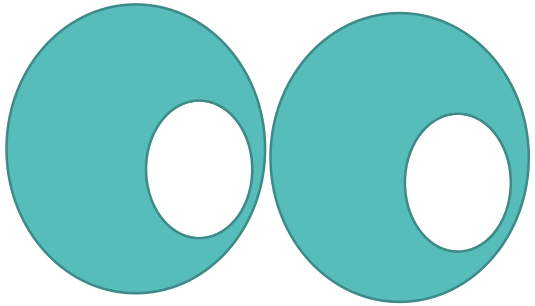
https://talktotransformer.com

Attention: Learning where is the information?

→As in an image the information in a text is not distributed uniformly.

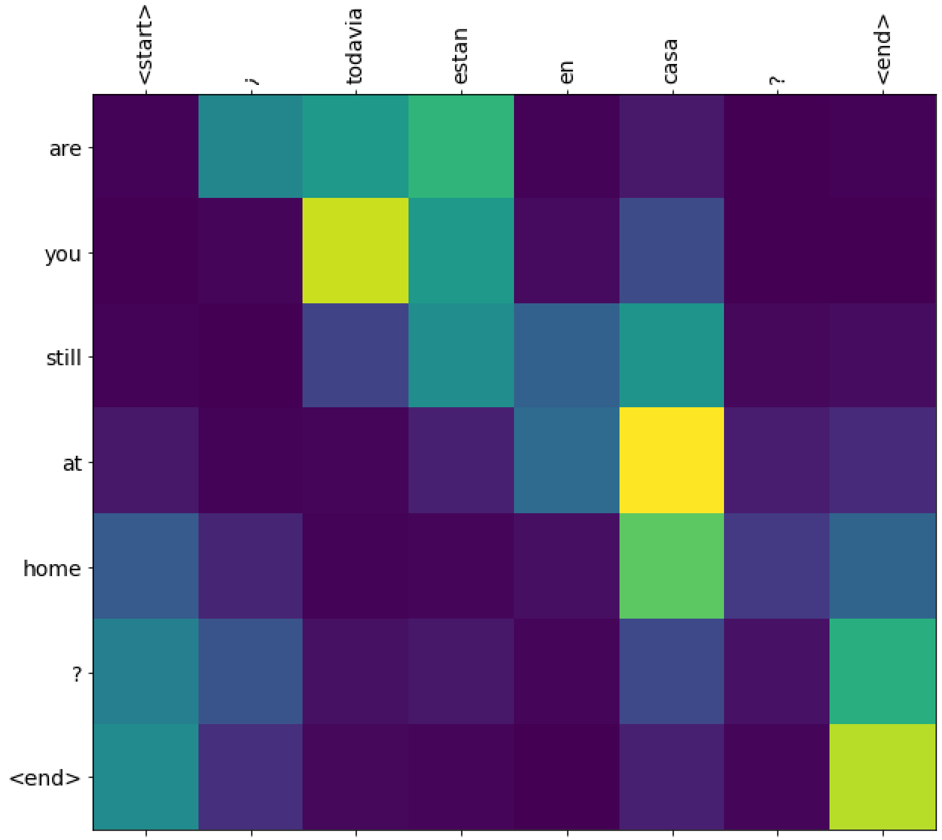
→Here is where you looked at!

Single attention head



¿todavia estan en casa?

are you still home ?



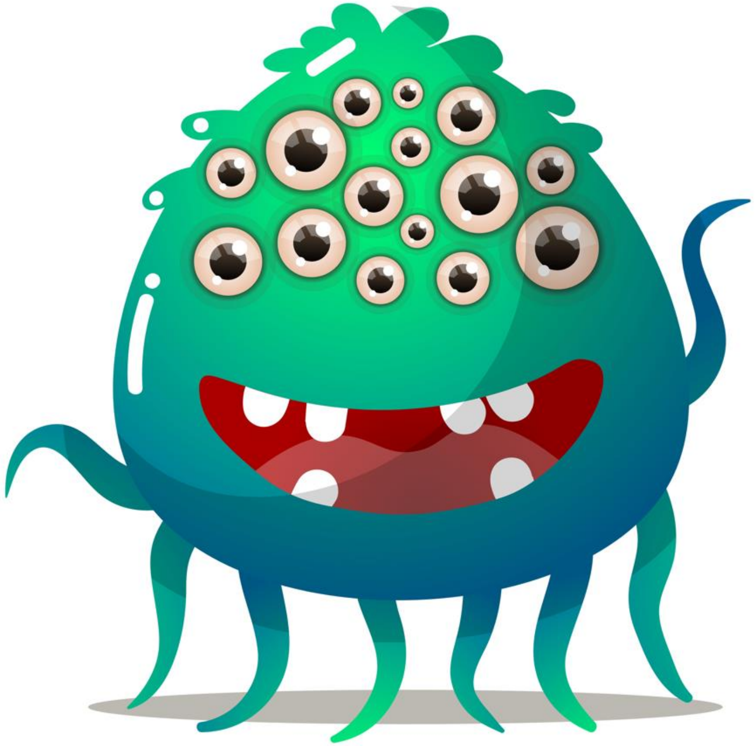
Attention: Learning where is the information?

And	we	've	been	trying	to	put	that	into	our	lab	robots	because	we	think	this	is	how	you	are	going	to	want	to	interact	with	robots	in	the	future	.	
0.2022	0.1364	0.0863	0.0510	0.0442	0.0134	0.0177	0.0762	0.0125	0.0252	0.1232	0.0580	0.0136	0.0196	0.0202	0.0108	0.0043	0.0090	0.0044	0.0026	0.0014	0.0011	0.0036	0.0013	0.0080	0.0008	0.0201	0.0009	0.0003	0.0181	0.0136	➡ 우리는
0.0441	0.0336	0.0293	0.0338	0.0520	0.0171	0.0225	0.1551	0.0280	0.0425	0.2414	0.1247	0.0121	0.0086	0.0108	0.0210	0.0107	0.0101	0.0052	0.0038	0.0021	0.0009	0.0032	0.0012	0.0106	0.0013	0.0281	0.0023	0.0004	0.0208	0.0228	➡ 이것을
0.0070	0.0072	0.0101	0.0107	0.0362	0.0118	0.0435	0.0595	0.0186	0.1320	0.3905	0.2008	0.0033	0.0019	0.0030	0.0039	0.0015	0.0014	0.0028	0.0014	0.0011	0.0005	0.0022	0.0005	0.0060	0.0012	0.0124	0.0015	0.0004	0.0115	0.0156	➡ 연구@@
0.0302	0.0172	0.0170	0.0189	0.0190	0.0218	0.0172	0.0142	0.0277	0.0387	0.2911	0.4179	0.0074	0.0051	0.0037	0.0068	0.0031	0.0025	0.0016	0.0030	0.0015	0.0012	0.0015	0.0010	0.0036	0.0038	0.0069	0.0035	0.0025	0.0014	0.0093	➡ 실로
0.0154	0.0149	0.0128	0.0355	0.0501	0.0405	0.1691	0.0716	0.2272	0.0827	0.1055	0.1340	0.0052	0.0027	0.0051	0.0045	0.0009	0.0009	0.0014	0.0014	0.0012	0.0004	0.0019	0.0006	0.0029	0.0023	0.0021	0.0016	0.0002	0.0036	0.0021	➡ 집어@@
0.0527	0.0256	0.0311	0.0562	0.0545	0.1542	0.1607	0.0476	0.2482	0.0299	0.0297	0.0505	0.0157	0.0068	0.0081	0.0048	0.0012	0.0020	0.0014	0.0020	0.0009	0.0016	0.0035	0.0024	0.0024	0.0014	0.0010	0.0004	0.0003	0.0012	0.0019	➡ 넣으려고
0.1078	0.0942	0.0743	0.1968	0.1941	0.0582	0.0153	0.0329	0.0293	0.0102	0.0440	0.0731	0.0319	0.0106	0.0112	0.0042	0.0022	0.0019	0.0009	0.0008	0.0002	0.0003	0.0011	0.0001	0.0002	0.0001	0.0008	0.0001	0.0000	0.0018	0.0014	➡ 노력@@
0.1000	0.1233	0.1580	0.3289	0.1094	0.0394	0.0054	0.0106	0.0081	0.0080	0.0146	0.0317	0.0230	0.0155	0.0091	0.0047	0.0020	0.0019	0.0009	0.0013	0.0002	0.0007	0.0006	0.0001	0.0001	0.0000	0.0003	0.0000	0.0001	0.0005	0.0016	➡ 해왔습니다
0.0498	0.0401	0.0206	0.0161	0.0143	0.0054	0.0047	0.0162	0.0076	0.0197	0.0522	0.0403	0.0863	0.1292	0.0560	0.0403	0.0212	0.0416	0.0315	0.0186	0.0066	0.0088	0.0121	0.0076	0.0195	0.0026	0.0570	0.0036	0.0034	0.0639	0.1033	➡ .
0.0115	0.0104	0.0076	0.0062	0.0089	0.0031	0.0031	0.0115	0.0032	0.0094	0.0355	0.0224	0.1032	0.1219	0.0911	0.0524	0.0420	0.0445	0.0368	0.0257	0.0126	0.0121	0.0205	0.0117	0.0394	0.0046	0.0792	0.0068	0.0032	0.0792	0.0804	➡ 왜냐하면
0.0119	0.0182	0.0092	0.0060	0.0095	0.0025	0.0030	0.0125	0.0035	0.0087	0.0334	0.0192	0.0294	0.1449	0.0676	0.0550	0.0480	0.0396	0.0354	0.0291	0.0155	0.0096	0.0168	0.0103	0.0421	0.0048	0.0782	0.0085	0.0040	0.1173	0.1065	➡ 우리는
0.0042	0.0020	0.0020	0.0023	0.0055	0.0015	0.0018	0.0118	0.0028	0.0075	0.0278	0.0154	0.0139	0.0257	0.0368	0.0780	0.0679	0.0376	0.0347	0.0335	0.0184	0.0080	0.0189	0.0112	0.0636	0.0087	0.1344	0.0161	0.0055	0.1524	0.1502	➡ 미래에
0.0027	0.0005	0.0002	0.0003	0.0007	0.0002	0.0003	0.0020	0.0020	0.0017	0.0055	0.0043	0.0029	0.0069	0.0092	0.0192	0.0134	0.0197	0.0289	0.0357	0.0274	0.0128	0.0215	0.0225	0.0753	0.0682	0.1922	0.3135	0.0220	0.0344	0.0541	➡ 로봇이
0.0029	0.0002	0.0001	0.0001	0.0002	0.0001	0.0002	0.0003	0.0003	0.0007	0.0019	0.0024	0.0022	0.0020	0.0048	0.0042	0.0049	0.0339	0.0307	0.0366	0.0226	0.0163	0.0483	0.0373	0.1250	0.1041	0.1589	0.1634	0.0399	0.0542	0.1014	➡ 어떻게
0.0055	0.0002	0.0001	0.0000	0.0001	0.0000	0.0001	0.0003	0.0003	0.0006	0.0020	0.0023	0.0013	0.0013	0.0037	0.0022	0.0016	0.0320	0.0239	0.0270	0.0191	0.0170	0.0809	0.0321	0.1774	0.1060	0.1674	0.0929	0.0145	0.0702	0.1179	➡ 로봇@@
0.0211	0.0008	0.0001	0.0001	0.0002	0.0001	0.0002	0.0005	0.0014	0.0008	0.0026	0.0057	0.0018	0.0027	0.0041	0.0037	0.0053	0.0252	0.0120	0.0427	0.0115	0.0117	0.0290	0.0163	0.0836	0.2860	0.0848	0.1925	0.0470	0.0139	0.0925	➡ 과
0.0050	0.0003	0.0001	0.0001	0.0003	0.0001	0.0003	0.0005	0.0008	0.0006	0.0009	0.0009	0.0012	0.0012	0.0027	0.0020	0.0018	0.0173	0.0089	0.0237	0.0226	0.0135	0.0418	0.0458	0.1637	0.4662	0.0462	0.0627	0.0096	0.0208	0.0384	➡ 상호작용@@
0.0183	0.0012	0.0003	0.0002	0.0006	0.0005	0.0005	0.0007	0.0012	0.0008	0.0011	0.0023	0.0066	0.0068	0.0084	0.0045	0.0081	0.0621	0.0341	0.0834	0.0769	0.0952	0.1724	0.1301	0.1015	0.0802	0.0203	0.0164	0.0151	0.0090	0.0412	➡ 할
0.0238	0.0014	0.0003	0.0003	0.0008	0.0005	0.0002	0.0010	0.0011	0.0006	0.0014	0.0028	0.0124	0.0080	0.0380	0.0149	0.0237	0.1834	0.0464	0.0866	0.0315	0.1128	0.1151	0.0832	0.0488	0.0418	0.0319	0.0121	0.0106	0.0206	0.0437	➡ 것인지를
0.0287	0.0060	0.0047	0.0034	0.0131	0.0077	0.0134	0.0079	0.0063	0.0052	0.0196	0.0213	0.0770	0.0361	0.3100	0.0401	0.0186	0.0759	0.0459	0.0258	0.0104	0.0278	0.0545	0.0343	0.0566	0.0186	0.0132	0.0016	0.0015	0.0101	0.0046	➡ 생각하기
0.0590	0.0111	0.0105	0.0096	0.0145	0.0249	0.0054	0.0029	0.0098	0.0034	0.0086	0.0147	0.2700	0.0435	0.1385	0.0233	0.0193	0.0490	0.0242	0.0244	0.0110	0.0391	0.0904	0.0385	0.0134	0.0124	0.0083	0.0022	0.0030	0.0089	0.0063	➡ 때문입니다
0.2477	0.0442	0.0196	0.0113	0.0103	0.0060	0.0039	0.0067	0.0075	0.0071	0.0065	0.0103	0.0385	0.0527	0.0508	0.0274	0.0208	0.0731	0.0319	0.0302	0.0095	0.0260	0.0201	0.0152	0.0200	0.0098	0.0244	0.0076	0.0208	0.0266	0.1134	➡ .
0.1173	0.0254	0.0132	0.0070	0.0082	0.0040	0.0031	0.0058	0.0028	0.0050	0.0086	0.0062	0.0280	0.0406	0.0573	0.0264	0.0219	0.1083	0.0288	0.0241	0.0063	0.0172	0.0284	0.0142	0.0586	0.0136	0.0653	0.0094	0.0155	0.0544	0.1752	➡ </s>

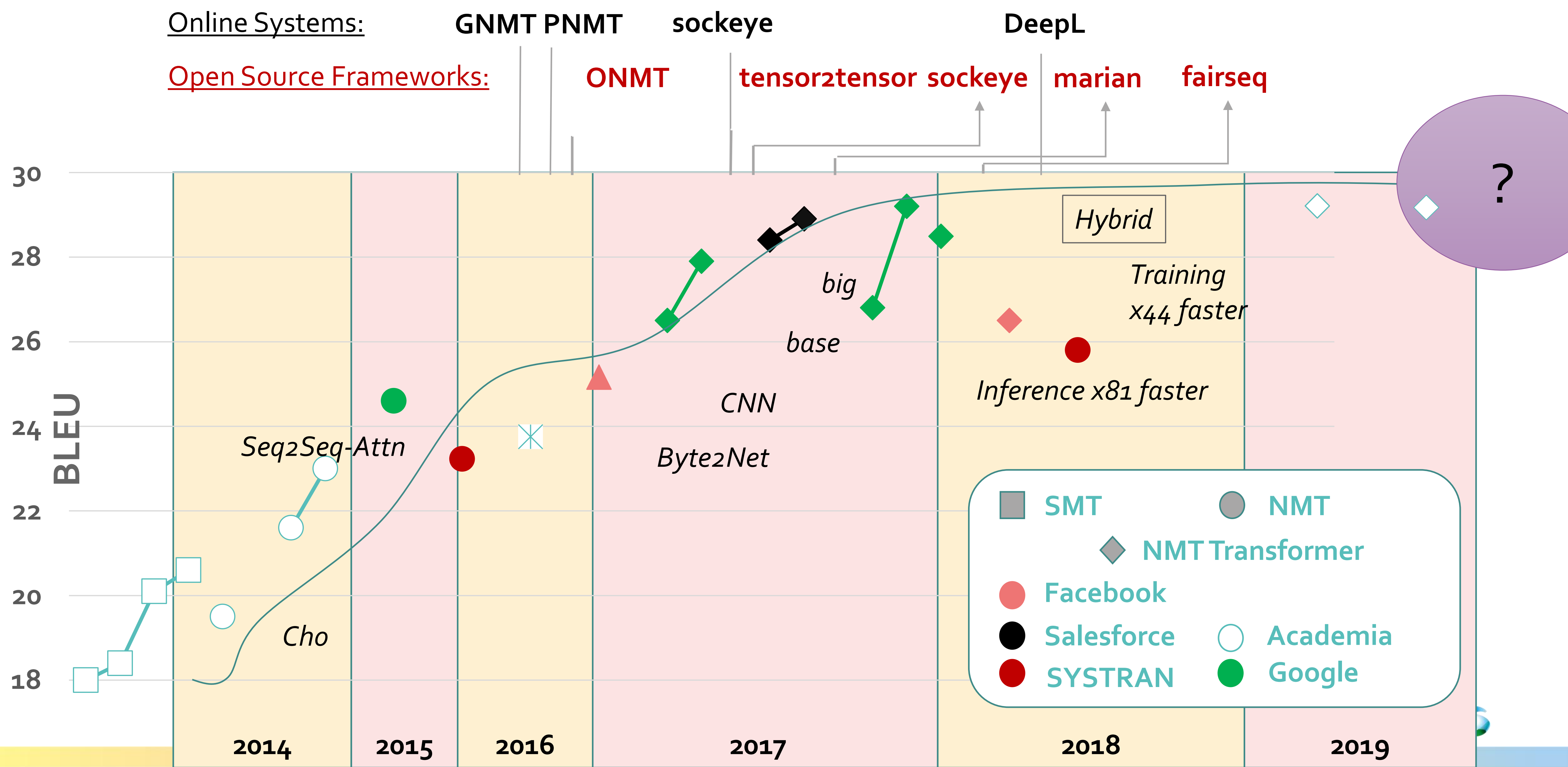
From single attention to multi-head attention

- ❑ Generalization of the attention mechanism between each layer of the encoder and decoder
- ❑ Train simultaneously multiple attention head

- ➢ Base architecture of the "Transformer"
- ➢ Opens new opportunities for the neural network



5 years of NMT – on English-German WMT14 Task



Technology size

	RBMT	SMT	NMT
Code (lines)	> 1M	100000	5000
Data (bytes – model)	~500k	> 100B	1B

Real Brain vs. Artificial Brain – Size Comparison…

snail

ant

frog

mouse

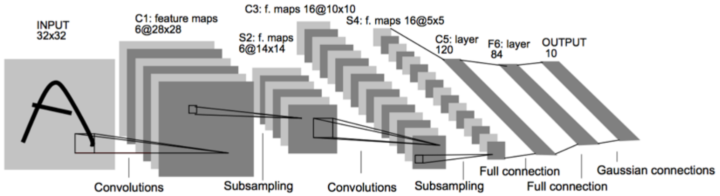
guinea pigs

cat

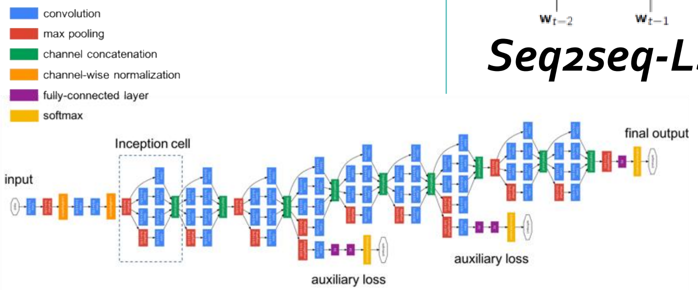
macaque

humans

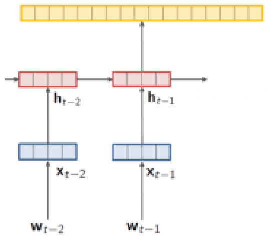
11K 60K 250K 16M 25M 71M 240M 760M 1.6B 6B 86B



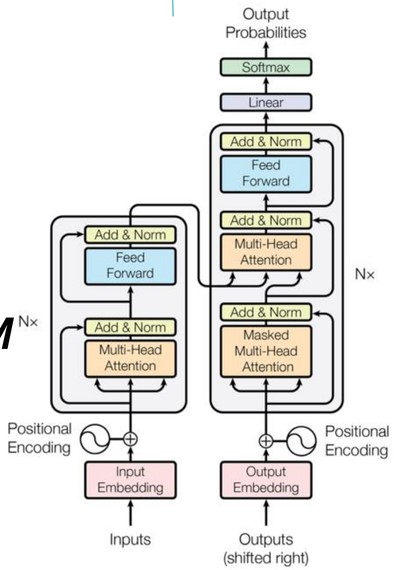
LeNet-5



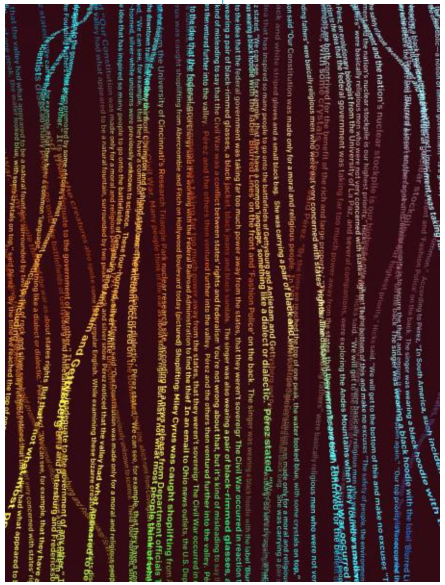
Inception



Seq2seq-LSTM

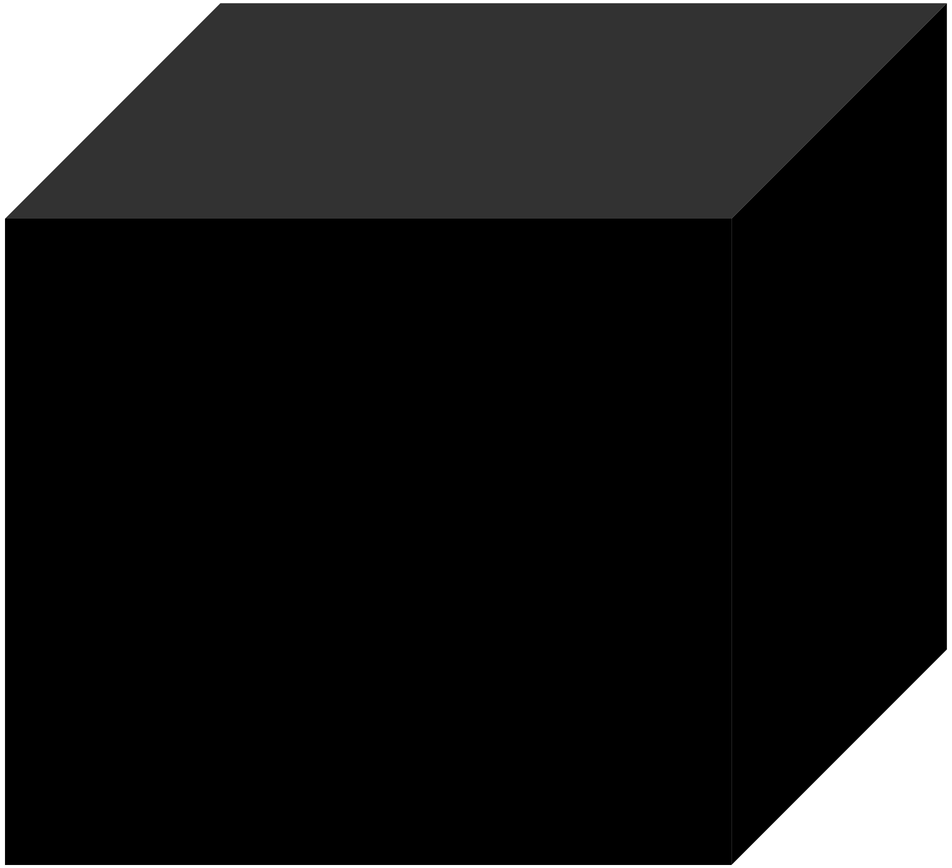


Transformer-Big



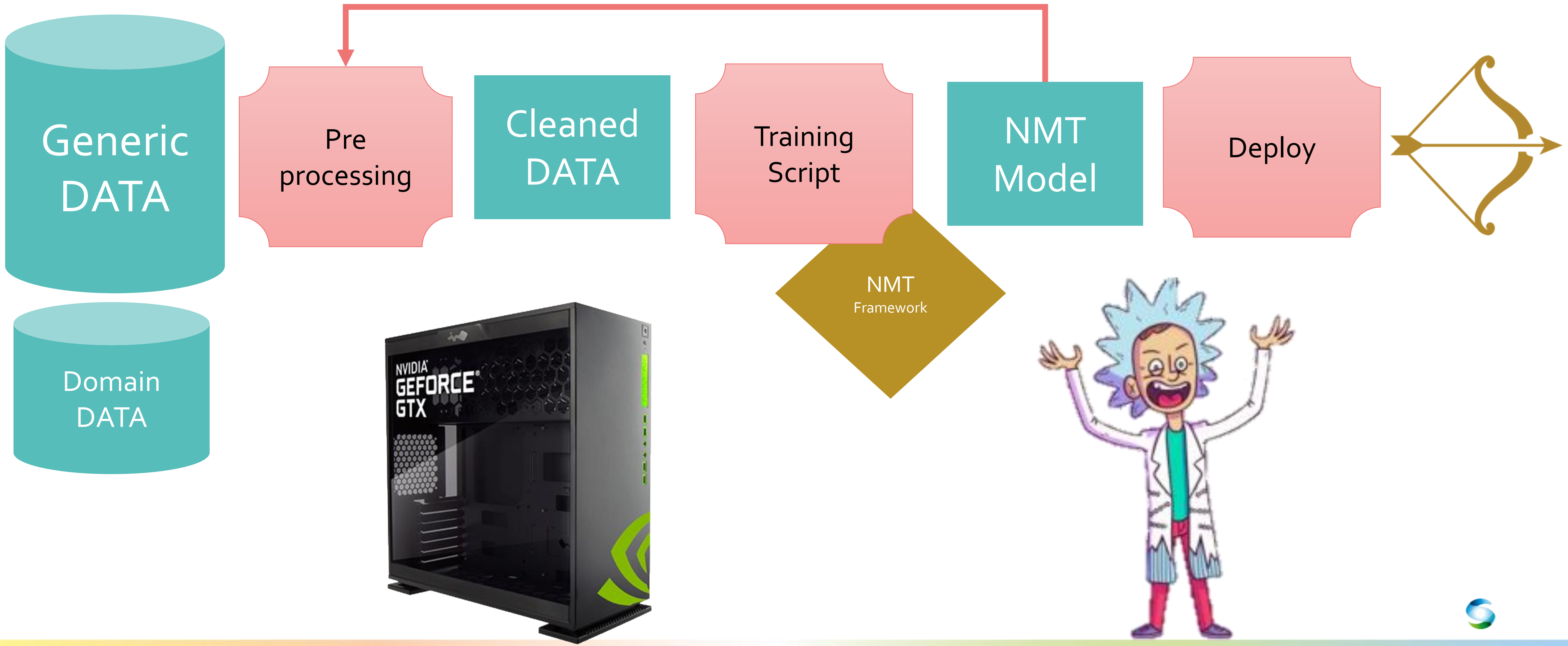
GPT-2-XL

But we can't work with a black box!



Understanding how is not important, but understanding why, and how to interact is.

Training a Neural Translation Model – a data-scientist toy?



Teaching Neural Machine Translation Models

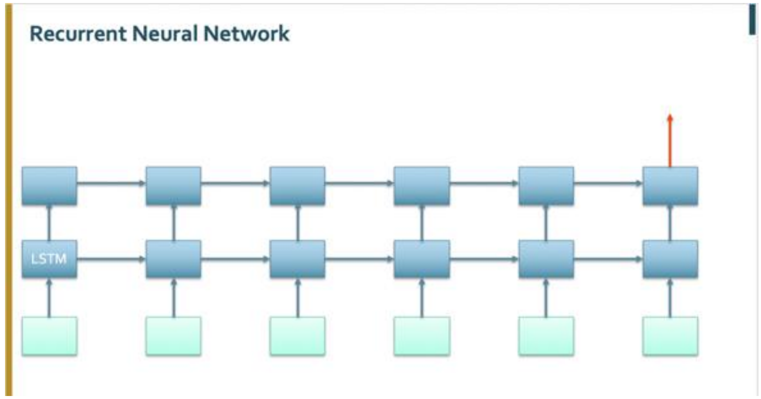
A toy for data scientist, a tool for translator

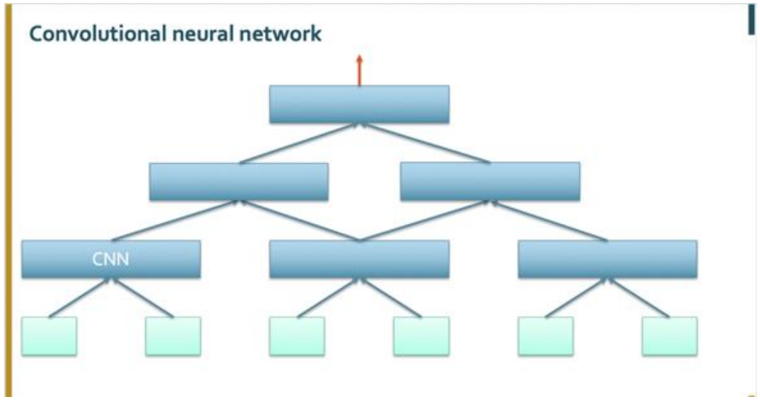
Training or Teaching NMT models

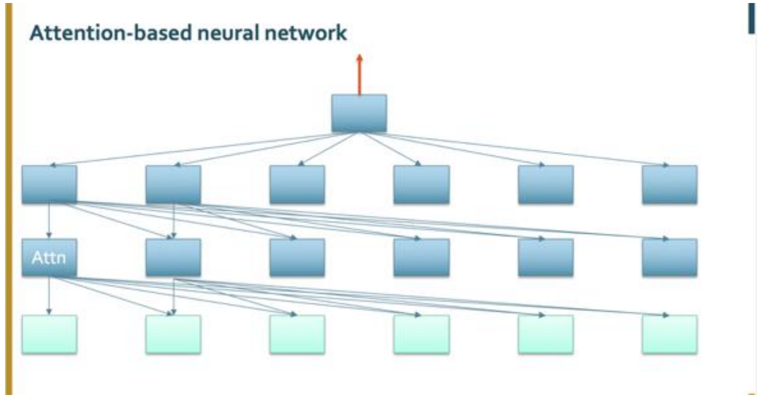
A neural network.

Un réseau de neurone.

Un ~~al~~one.







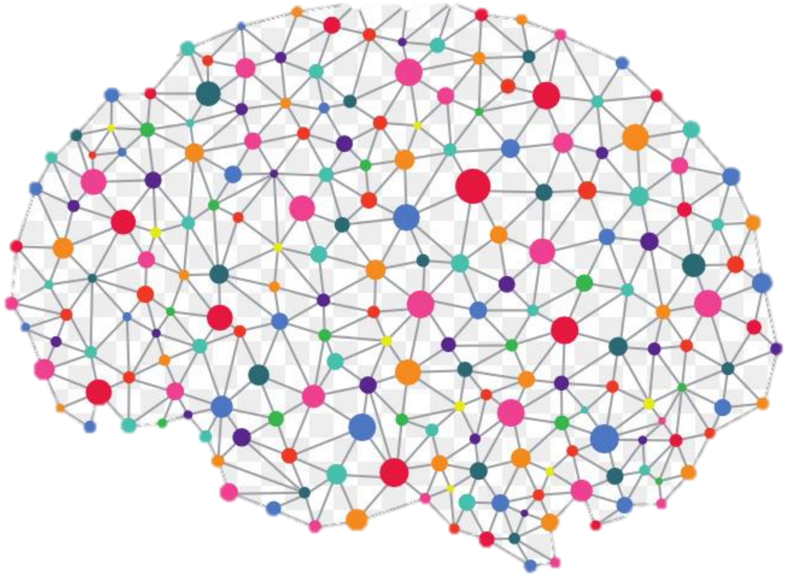
Teaching an NMT Model – Understanding NMT

A neural network is smart,

… really smart

- Will try to make sense out of non-sense
- Will always outsmart the trainer
- Will take training data as gold truth

… but cannot guess user expectations



Learning process is incremental but follows random process

Teaching Neural Machine Translation - Limits and Solutions

- ❑ Generic and In-Domain Translation
- ❑ Dealing with Data scarcity
- ❑ Dealing with Noisy data
- ❑ Translation divergence
- ❑ Understanding Biases
- ❑ Tag Handling
- ❑ Taking Time to train

Domain Adaptation

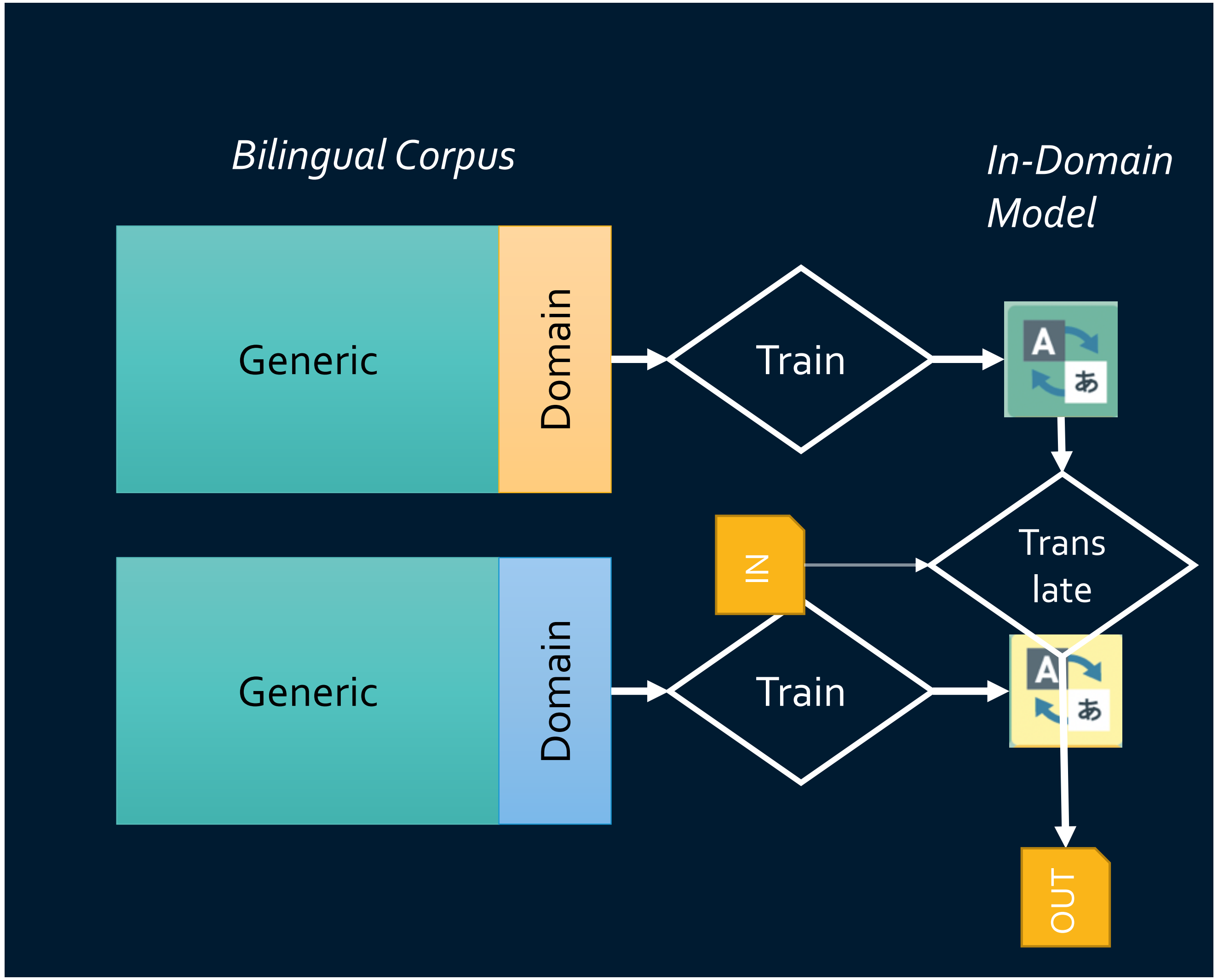
→ **Goal: adapt a given translation model to a "domain"**

→ **Very broad definition of Domain:**

+ General Topic: Finance vs. IT
+ Industry Specific: eBay vs. Amazon
+ Use case specific: chat is a domain
+ Translator Specific

→ **Challenges**

+ Quick Adaptation
+ Prevent Catastrophic Forgetting



Domain Adaptation

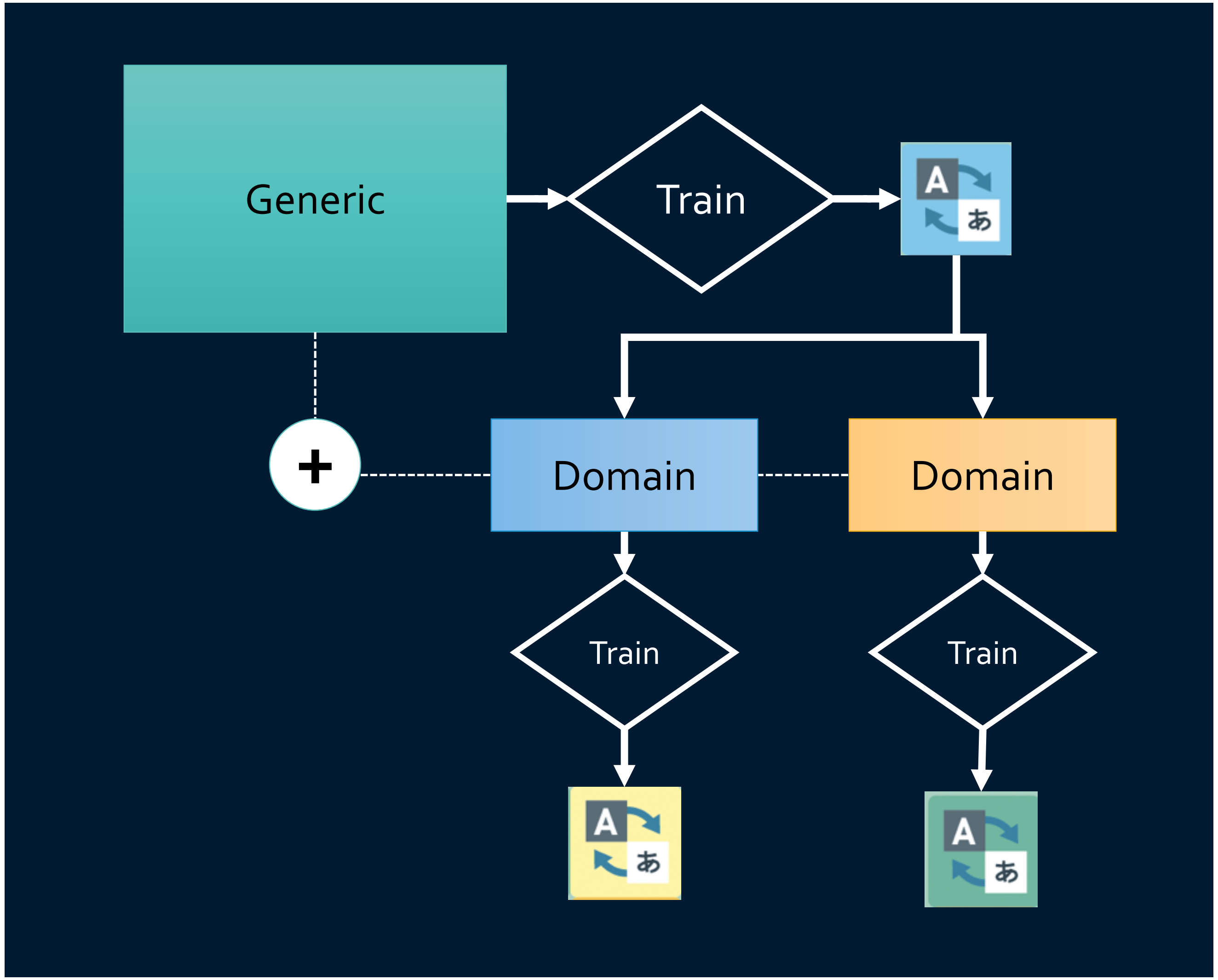
→ **Goal: adapt a given translation model to a "domain"**

→ **Very broad definition of Domain:**

+ General Topic: Finance vs. IT
+ Industry Specific: eBay vs. Amazon
+ Translator Specific

→ **Challenges**

+ Quick Adaptation



Dealing with Data Scarcity

Cannot train without data!

→ **But there are many ways to deal with under-resourced languages**

- Back Translation Techniques
- Unsupervized training techniques (Generative Adversarial Networks)
- Use close language data
- ... produce or acquire corpus

Dealing with Noisy Data

Not all the noise is problematic!

+ Noise in the source will make the model more robust
+ Noise in target will become a training objective

→ **Instead of trying to remove the noise, take control of noise generation**

+ OCR/ASR errors, case insensitive, missing diacritics, no punctuation mark, typos, presence of tags, textual formatting, …

→ **Carefully remove noise in target side**

Translation Divergences

→ **Many sources of possible divergences between source and target**

→ **The Neural Network will try to produce rules without enough context.**

→ **Weight the impact of each of them on the translation**

→ **Annotate**

We think our solution will help you.

SYSTRAN의 솔루션이 도움이 될 것입니다.

Some of the contests open for people **outside the US** are:

Voici quelques concours ouverts aux **candidats étrangers** :

Their wages are often inconsistent, ranging from 200 Thai baht (**US$5.72**) …

Les salaires des travailleurs sont souvent fluctuants, allant de 200 baht thaï (environ **5,2 euros**)…

How much do you get paid?

T’es payé combien **de l’heure** ?

Paris souhaite la réélection de Barack Obama

فرنسا تأمل في اعادة انتخاب اوباما

And far more…

→ **Anticipate and Understand Biases**

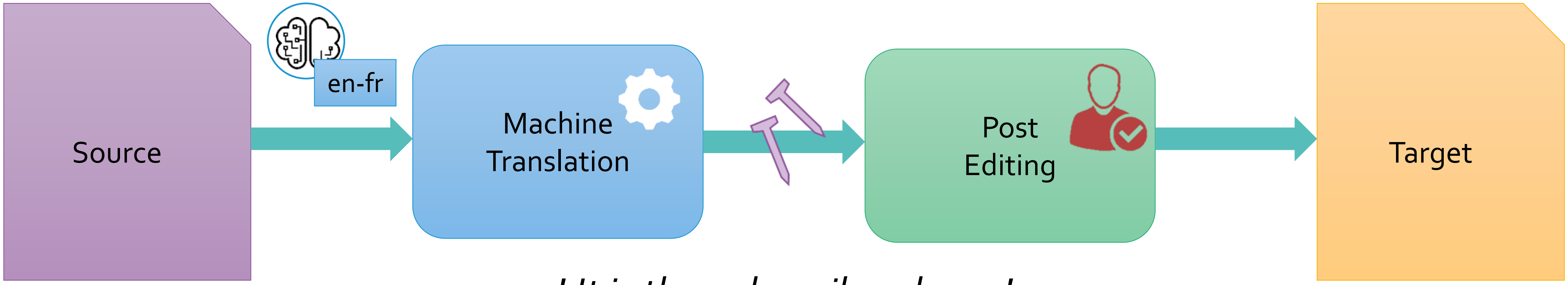
→ **Think about the use case:**

+ Translate sentences, description, words?
+ Translate text or formatted documents?
+ Translation meant for post-editing?
+ …

→ **Take the time – remember that Rome wasn't built in a day. Be patient and keep teaching your neural network…**

NMT as a tool

Traditional Usage of Neural Machine Translation <u>in localization</u>



! It is the only nail we have !

Rethinking the tool for the use-case

- **Feedback injection** => adaptative translation model
- **Control terminology** => train the model to handle terminology
- **Provide Translation Options** => use diversity in training data
- **Neural augmented fuzzy** => train the model to understand fuzzy matches
- **Neural Post-Editing** => train a model to learn post-editing
- **Neural Consistency and Style Checking** => train a model to check consistency
- **Interactive Translation Editor** => Translator interaction while decoding

Future Directions

Future Innovation is in the ecosystem

4 FUNDAMENTAL BRICKS

TECHNOLOGY

To build and run world class language engines

Framework OpenNMT
Advanced end-to-end solution
Adapted to the professional market

SYSTRAN Pure neural® server NMT

DATA

To build and train in domain language models

In-domain Data
Transparency & traceability
Privacy

HUMAN EXPERTISE

To train the model according to the client' s expectation

Expert know-how
Advanced customization
Continuous quality improvement

INFRASTRUCTURE

To enable scalable deployment and to relieve the user from the complexity of use

Plug & play
Scalable
GPU capacity optimization

OVHcloud™

Future technological directions

→**Zero-shot language pairs**

→**Translating Beyond Sentences**

→**Companion technologies**

+ Simplifying language
+ Translating text to sign language
+ Style Transfer

→**From Translation Solution to a Translation Assistant**

Future directions for Neural MT products

→ **Exciting perspectives**

→ Tools for accelerating language learning

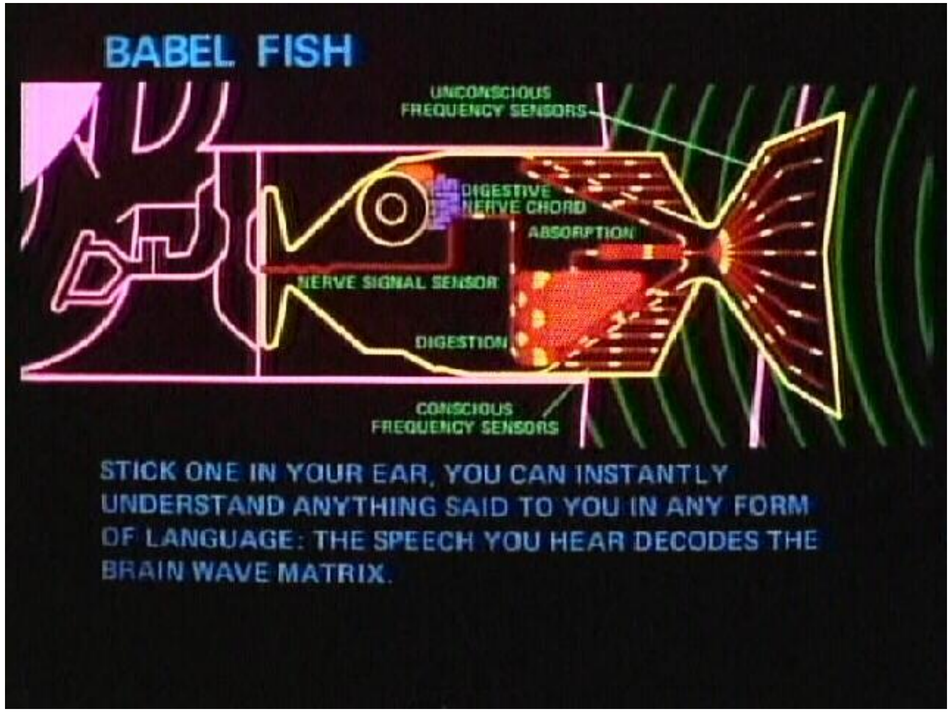
→ Tools for enabling bilingual authoring

→ New translation correction tools

→ Personal language assistant

→ and far more…

→ **For Tomorrow…**



Conclusion

→ **Neural Machine Translation is the first tool smart enough to learn from language experts**

- ➢ No Limit today on the capacity

→ **Technology will continue to evolve but the key innovation will be in changing ownership from technologists to translators**

- ➢ Toward a marketplace of translation models

→ **Usage needs to evolve to embed neural technology not only as a "pre-translation" but deep in the multilingual content production workflow**