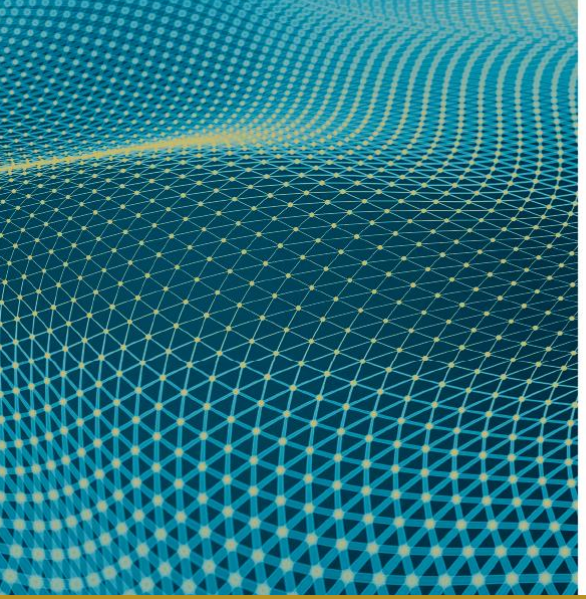


SYSTRAN

beyond language

TAILORED AND SECURED

From Phrase-Based to Sentence-Based MT

SYSTRAN PNTM® Machine Translation contribution to quality improvement

Université de Lille — UFR LEA – Langues Etrangères Appliquées

Stéphanie LABROUE, Expert Consultant, SYSTRAN

Date: February 2nd, 2018

SYSTRAN Today

+140
Language combinations

+25%
Revenue invested in R&D

Seoul
Paris – R&D
San Diego

Research Ecosystem

HARVARD UNIVERSITY

cnrs

icar

RWTH AACHEN UNIVERSITY

ETRI

Limsi

200
Employees

cnrs

Leader in machine translation and natural language processing

SYSTRAN

Pure neural MT™

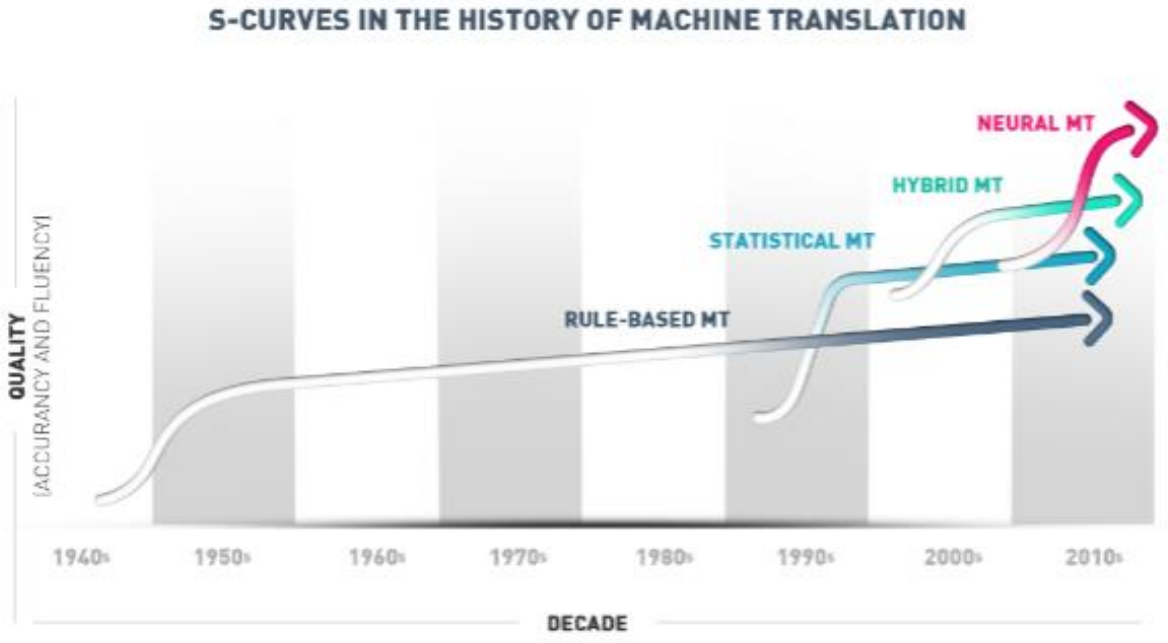
RB

SMT

NMT

SYSTRAN's MT technologies coverage

Translation Quality evolution across the last decades



Strenghts and Weaknesses of RB | Hybrid & SMT | NMT systems

Rule-based

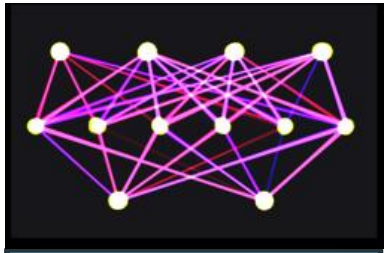
\+
Linguistic knowledge
+Translation consistency
+Speed
+Full control of customization

\-
Costly to develop

Hybrid / SMT

Memorizing patterns observed
+Fluency
+Leverage massive amounts of data

Limited control over the patterns learnt



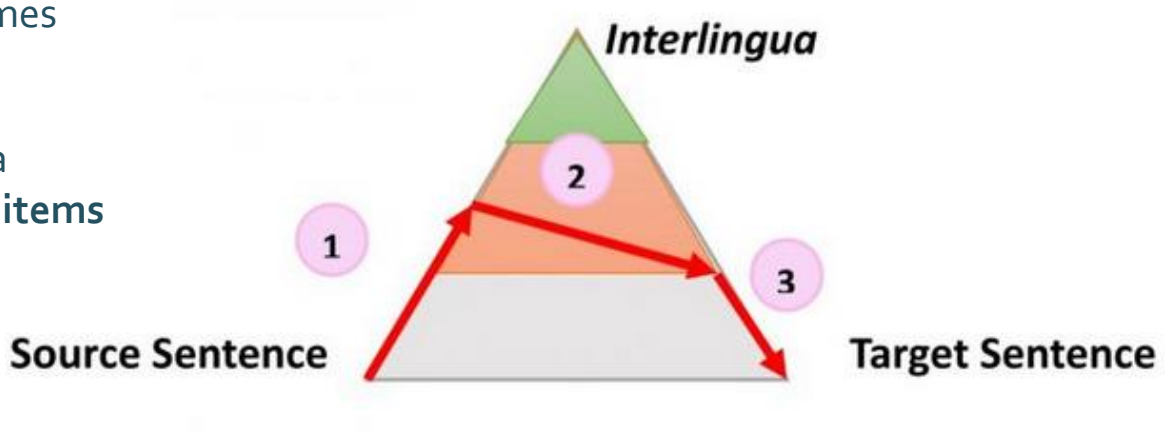
Neural

Fine-grained analysis of the language
+Better contextualization
+Fluency

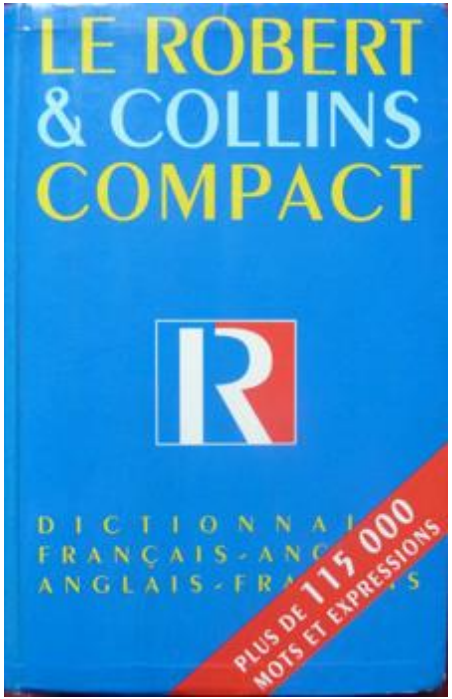
Requires more processing power at runtime

Rule-Based Machine Translation (1960's)

- Process strictly follows the **Vauquois triangle** and the analysis side is often very advanced, while the generation part is sometimes reduced to the minimum
- All 3 steps of the process use a **database of rules and lexical items on which the rules apply**
- These rules and lexical items are « readable » and **can be modified by linguist/lexicographer**



Rule-Based Machine Translation (1960's)



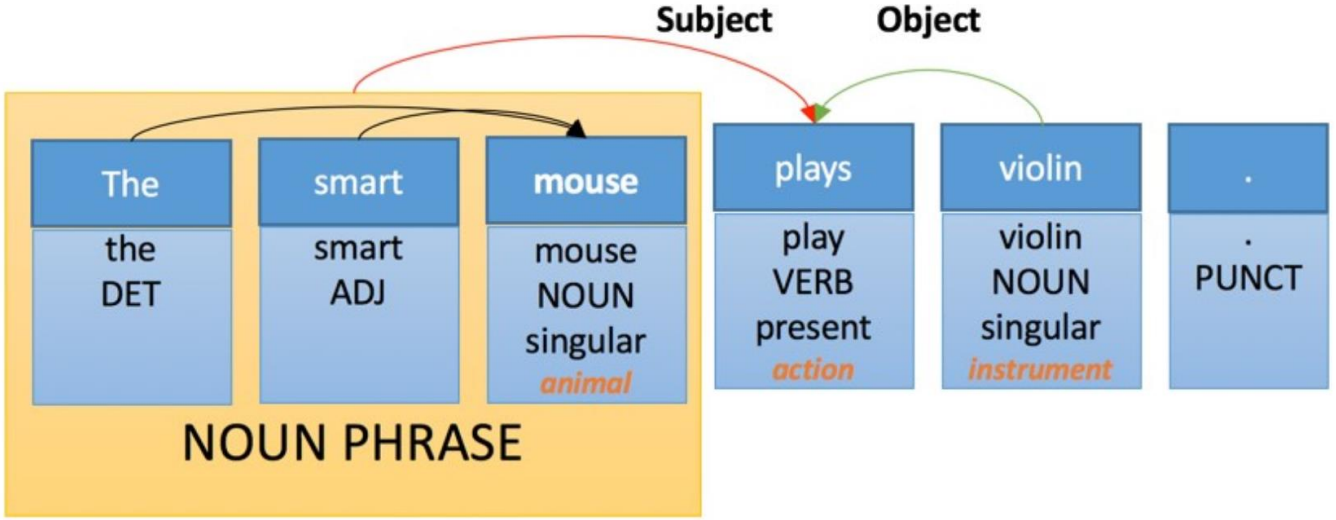
The blue car

Le- La voiture bleu e

féminin

Rule-Based Machine Translation (1960's)

- **E.g, internal representation of a sentence:**



- **Part of speech tagging**
- **Morphological analysis** ("plays" → inflected third person present form of the verb "play")
- **Semantic analysis**: ("violin" → instrument)
- **Constituent analysis**: ("the smart mouse" → noun phrase)
- **Dependency analysis**: words and phrases are connected with "links", here we identify the subject and the object of the main verb "play"

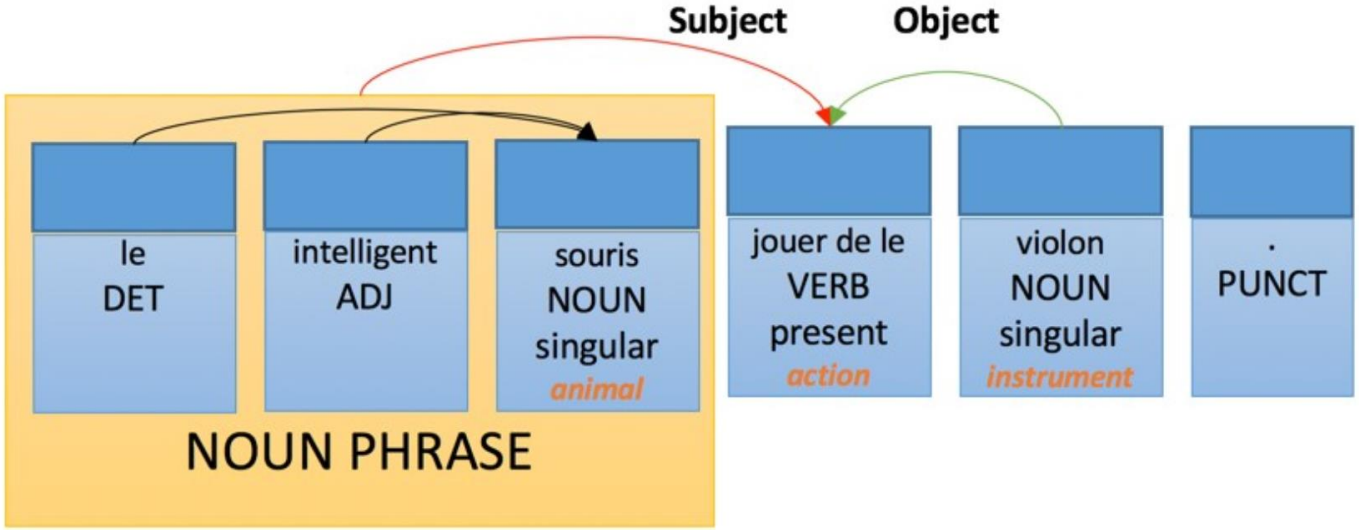
Rule-Based Machine Translation (1960's)

- Transfer of such a structure will use rules and lexical transformations

<the.DET>→<le.DET>
<smart.ADJ> modifying <NOUN+animated> → <intelligent.ADJ>
<mouse.NOUN> → <souris.NOUN>
<play.VERB>(subject: S, object: O <NOUN+instrument>) → <jouer de le.VERB>(S,O)
<violin.NOUN> → <violon.NOUN>
<NOUNPHRASE> → <NOUN PHRASE>

Rule-Based Machine Translation (1960's)

- **Application of these rules on the previous example will generate the target language representation of the sentence:**

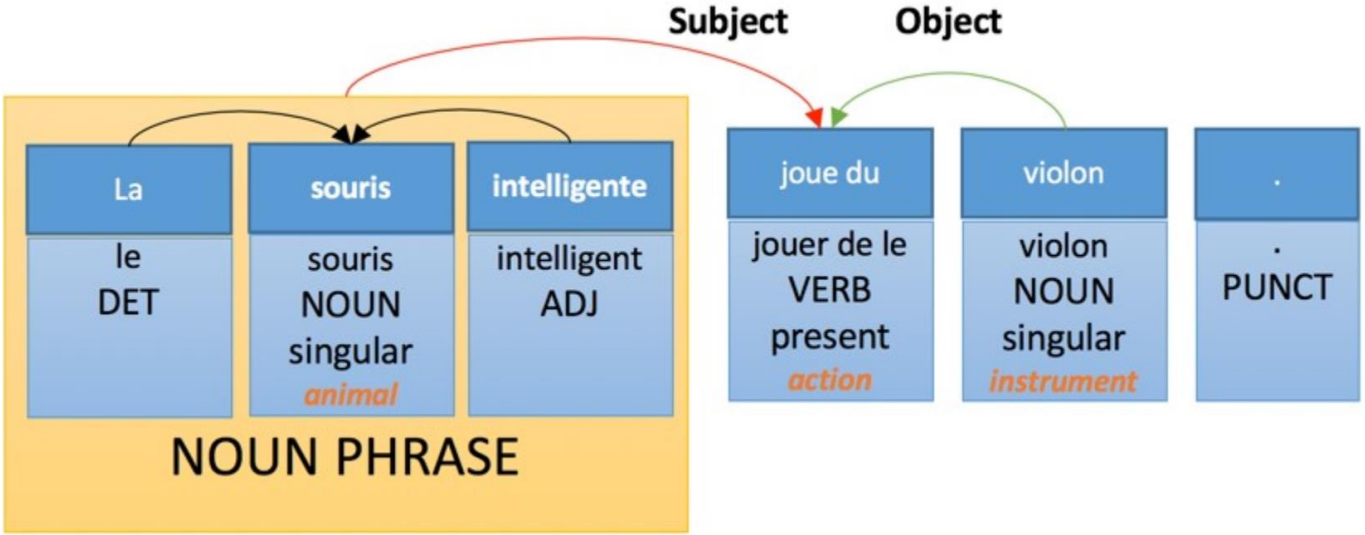


French generation rules will define:

- The adjective in a noun phrase follows the noun
- A determiner agrees in number and gender with the noun it modifies
- An adjective agrees in number and gender with the noun it modifies
- The verb agrees with the subject

Rule-Based Machine Translation (1960's)

- Application of these rules on the previous example will generate the target language representation of the sentence:

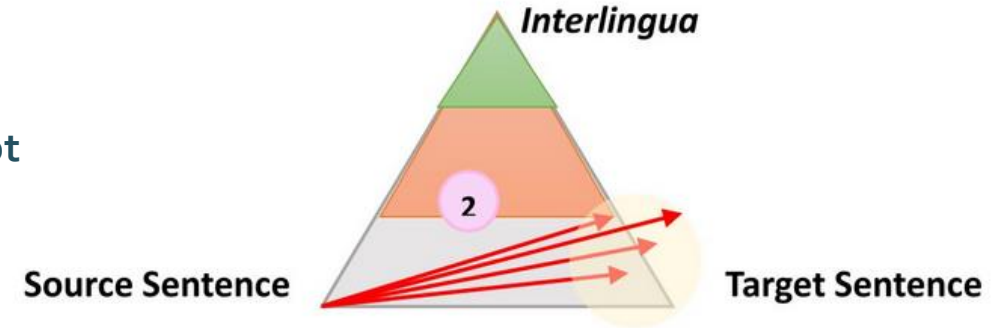


French generation rules will define:

- The adjective in a noun phrase follows the noun
- A determiner agrees in number and gender with the noun it modifies
- An adjective agrees in number and gender with the noun it modifies
- The verb agrees with the subject

Phrase-Based Machine Translation

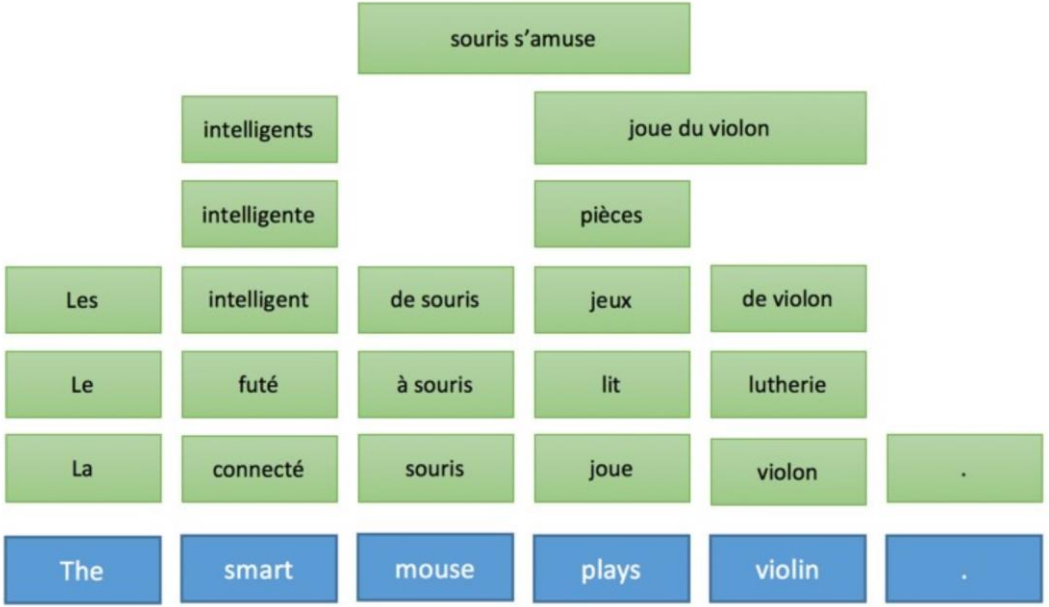
- Simplest and most popular version of **Statistical Machine Translation**
- Technically-speaking, phrase-based machine translation **does not follow the process described by Vauquois**
- Not only is there **no analysis or generation**, but more importantly the **transfer part is not deterministic**
  → the engine can generate multiple translations for one source sentence
  → strength of the approach resides in its ability to select the best one



Phrase-Based Machine Translation

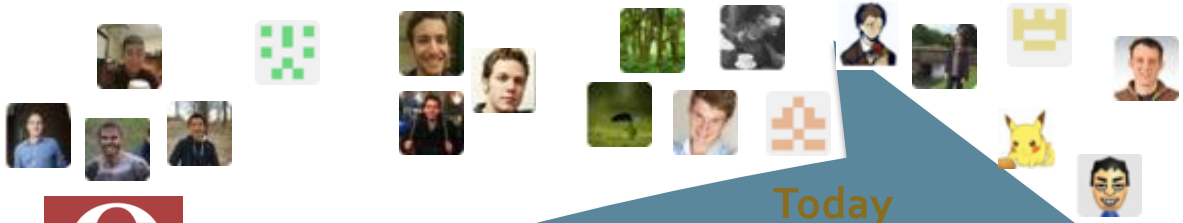
The model is based on 3 main resources:

- A **phrase-table** which produces translation option and their probabilities for "phrases" (sequences of words)
- A **reordering table** indicating how words can be reordered when transferred from source language to target language
- A **language model** which gives probability for each possible *word sequence* in the target language



- **Smart probability calculations** and smarter search algorithms → only the **most likely translation will be explored and the best one kept**.

Neural MT at SYSTRAN



NMT

Pure neural MT™

July 2016

POC
ZH<>EN

October 2016

12 language pairs

December 2016

32 language pairs

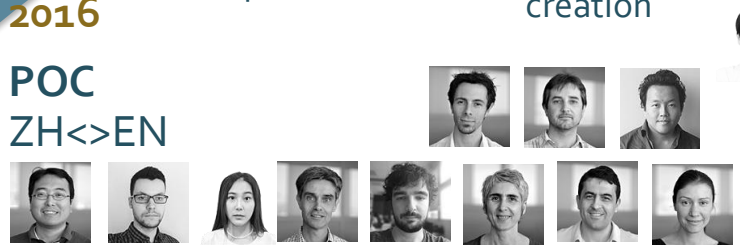
OpenNMT creation

March 2017

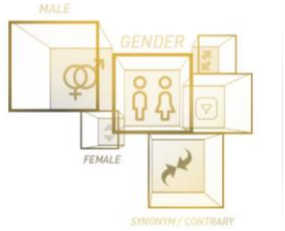
A **growing** and **active** OpenNMT community (**1500** followers)

Today

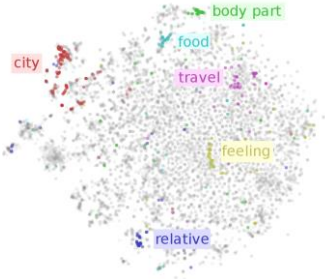
114 language pairs in production



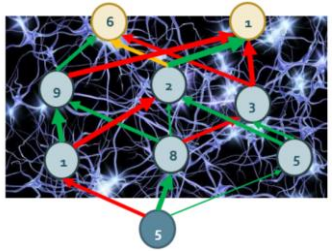
PNMT - 3 main ingredients



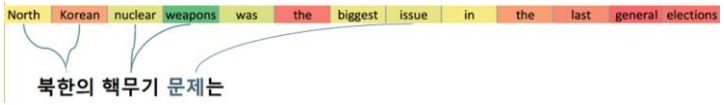
- **Word Embeddings**
- Powerful language map
- Words are grouped by meaning, grammar or semantic commonality



- **Recurrent Neural Networks**
- Contextual knowledge
- As the human brain would do, brings consistency and fluency



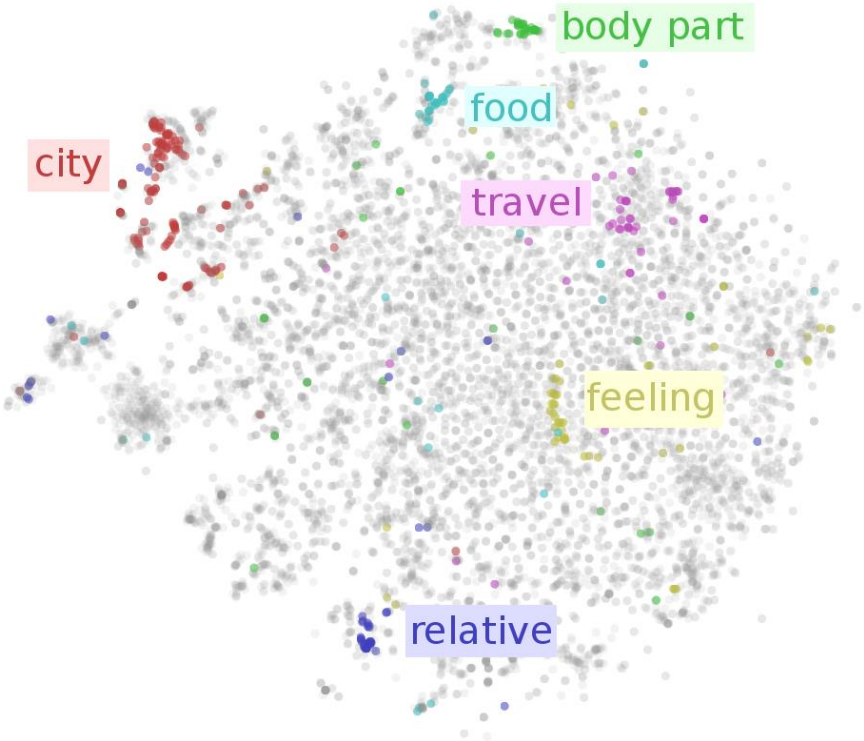
- **Attention Model**
- Attention capacity
- First focus on the key words at a given stage of translation



1st PNMT® ingredient: The Word Embeddings

- Words are "sparse" → we need a **continous representation**

- "**Word embedding**" force the representation of words into a relatively small area of a multi-dimensionnal space (e.g 1000) where similar words are close

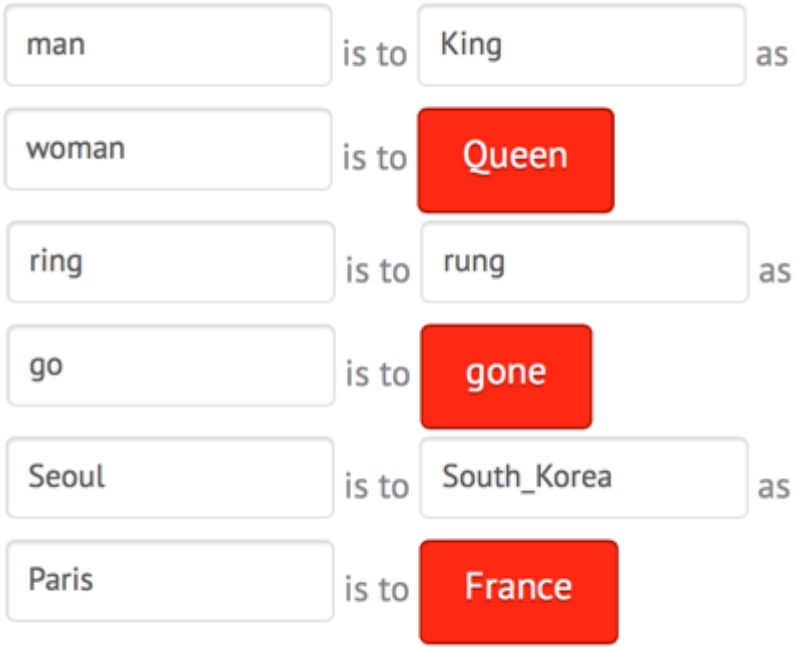


The power of Word Embeddings

This internal representation covers vast knowledge such as:

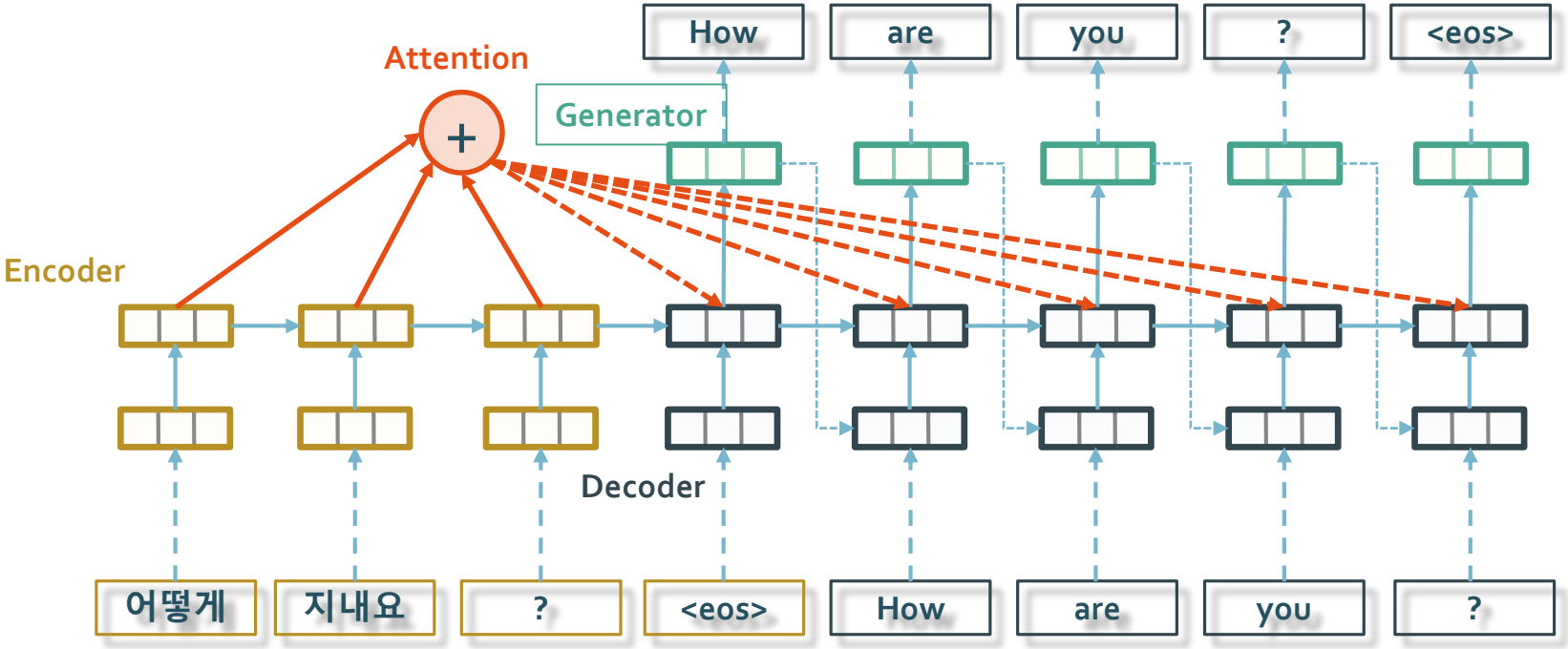
- Semantic
- Morphologic
- Pragmatic
- Grammatical

… and many others

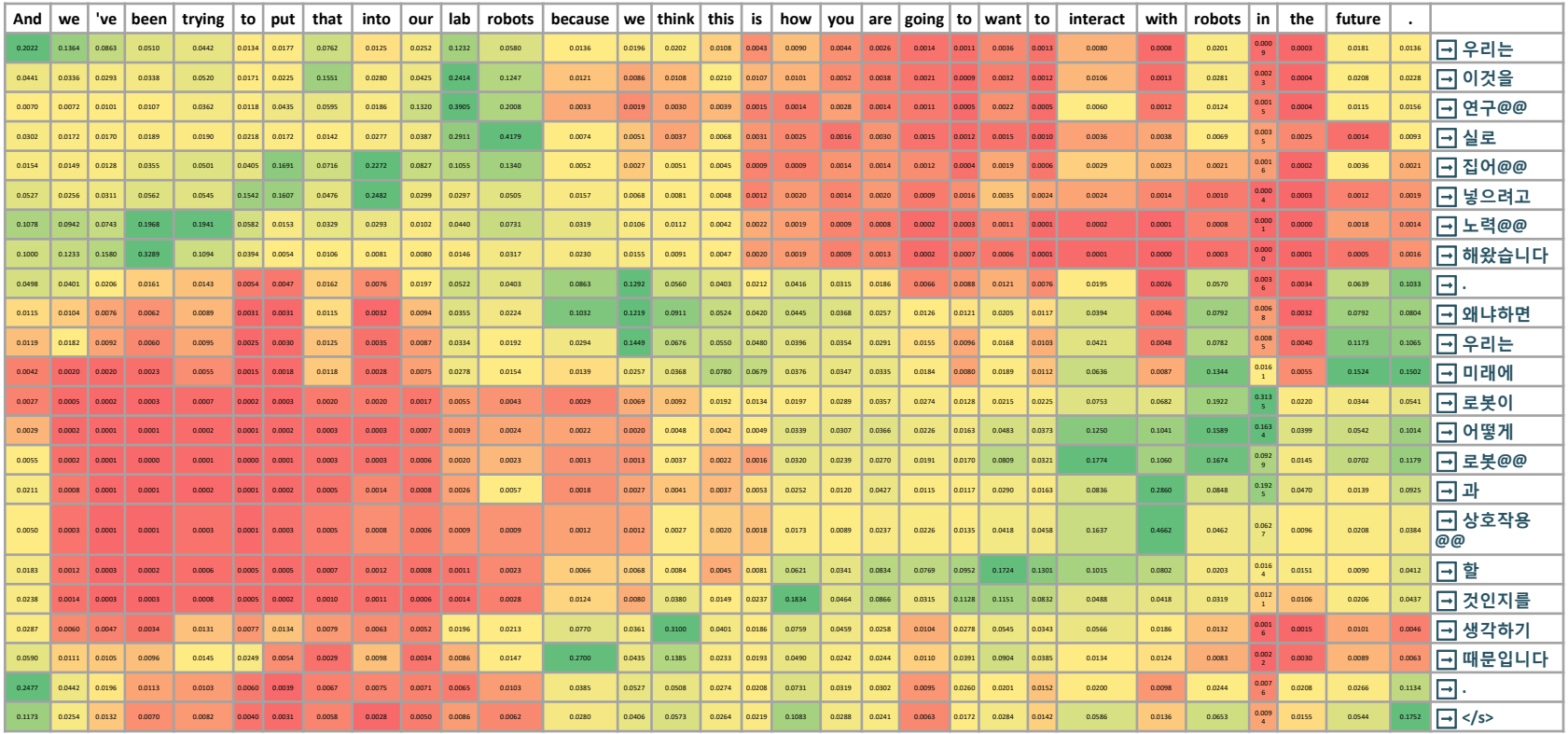


2nd PNMT® ingredient: The Attention Mechanism 1/3

2nd PNMT® ingredient: The Attention Mechanism 2/3



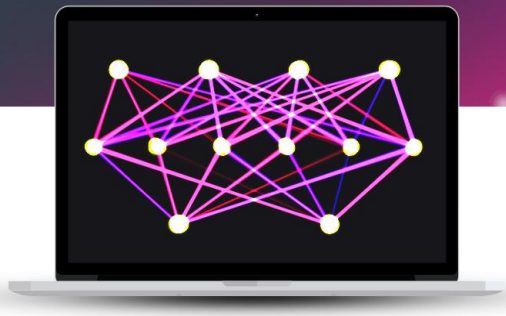
2nd PNMT® ingredient: The Attention Mechanism 3/3



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>

3rd ingredient: Neural Networks

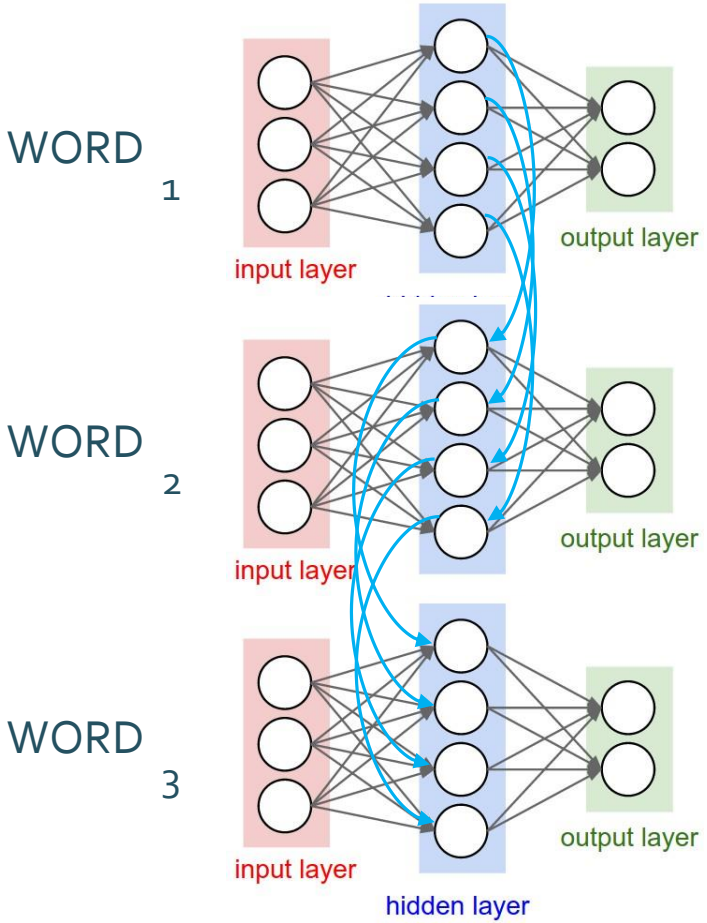
AN ARTIFICIAL NEURAL NETWORK (ANN) IS COMPOSED OF LAYERS OF ARTIFICIAL NEURONS



- The Layers are connected together with **weights** known as parameters
- The Neural Network is unique: the **training phase** allows to correct the parameters
- By sending **corrective feedback** to the engine based on the generated output and the expected output (reference)
- Neural Networks used for Natural Language Processing have 8 to 20 layers: called **Deep Neural Network** (DNN)

Recurrent Neural Networks (RNNs)

- **Basic neural networks** do not have any memory
- **RNNs** extend neural networks to allow them to store some information (context of a sentence as an example, or previous word(s), previous character…)
- **Weights** assigned to each connection (synapse)
- Output is compared with human reference translations during the learning process, and **errors are fed back in the system to recalculate weights → Learning process**



PNMT – Simplified and Corrected with Back Translation

Source	Back Translated (EN>FR>EN)	Benefits
Is often used by laypeople to evaluate a machine translation system, or to test whether a text is suitable for MT when they are unfamiliar with the target language. Because the resulting text can often differ substantially from the original, RTT can also be a source of entertainment.	Is often used by profanes to evaluate an automatic translation system or to check if a text is appropriate for MT when they are not familiar with the target language. Because the resulting text may often differ considerably from the original, the road can also be a source of entertainment.	+ Cognition transferred + Simplified lexicon + Simplified grammatical structure + More explicit and understandable
There is a possibility of prior Board approval of these investments.	The Commission may approve these investments.	+ Simplified + More explicit and understandable
Both experience and books are very important about living.	Experience and books are very important for life.	+ Corrected ~~about living~~ → for life
That's why I like to make travel by using my own car	That's why I like to travel using my own car.	+ Corrected ~~to make travel~~ → to travel ~~by using my own~~ → using my own

Application perspectives for this new technology

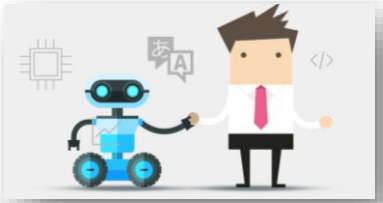
The same approach can be applied to all "Sequence to Sequence" type tasks

- Automatic text simplification
- Spellchecking
- Chatbot
- Information extraction…

Chatbot

- Q: When do you sleep?
  A: I don't sleep.
- Q: You actually did care for her, didn't you?
  A: Of course I did.
- Q: Am I talking to myself?
  A: I'm talking to you.
- Q: Want to hear what seung-moon is lamenting about?
  A: No, I don't want to hear it.
- Q: What do you know about him?
  A: I've seen him around.
- Q: Do you know what he does?
  A: He's a writer.
- Q: NOW, HOW MANY HAVE BEEN INJURED?
  A: I DON'T KNOW.
- Q: What are you doing?
  A: I'm trying to get you out of here.
- Q: Where did you come from?
  A: I've been looking for you.

Specialization and Infinite Training

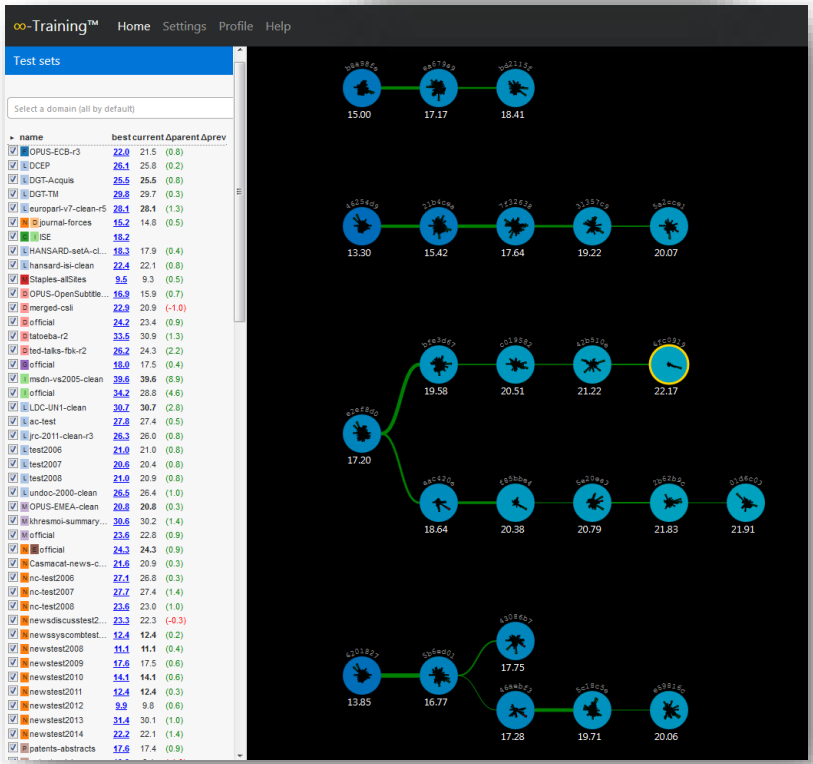


Neural introduced a quantum leap in translation quality, but raised new challenges

- How to customize an engine for a specific task?
- How to speed up the training process?
- How to analyze and correct a mistake?
- How to prevent an AI system to forget?
- …

Towards a new approach: raising AIs

- Training a model is comparable to teaching a student
  - Repetition, Diversity, Incremental complexity
  - Polyvalence (=> several languages, domains)



Thank you!

www.systrangroup.com