From Phrase-Based to Sentence-Based MT

SYSTRAN PNTM® Machine Translation contribution to quality improvement

Stéphanie LABROUE, Expert Consultant, SYSTRAN

Date: February 2nd, 2018
SYSTRAN Today

+140 Language combinations

+25% Revenue invested in R&D

Research Ecosystem

Seoul Paris – R&D San Diego

200 Employees

Leader in machine translation and natural language processing

SYSTRAN’s MT technologies coverage
Translation Quality evolution across the last decades
Strengths 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 bleue féminin
Rule-Based Machine Translation (1960’s)

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

  ![Diagram showing part-of-speech tagging, morphological analysis, semantic analysis, constituent analysis, and dependency analysis.]

  - **Part of speech tagging**
  - **Morphological analysis** (“plays” \(\rightarrow\) inflected third person present form of the verb “play”)
  - **Semantic analysis**: (“violin” \(\rightarrow\) instrument)
  - **Constituent analysis**: (“the smart mouse” \(\rightarrow\) 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

```plaintext
<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> → <violin.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:

\[ \text{La le DET} \quad \text{souris NOUN singular animal} \quad \text{intelligente ADJ} \quad \text{joue du} \quad \text{violin NOUN singular instrument} \quad . \quad \text{PUNCT} \]

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

- **July 2016**: POC ZH<>EN (12 language pairs)
- **October 2016**: 32 language pairs, OpenNMT creation
- **December 2016**: A growing and active OpenNMT community (1500 followers)
- **March 2017**: 114 language pairs in production
- **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 continuous 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

How are you?

encoder

attention

decoder

SYSTRAN

©2018 ALL RIGHTS RESERVED
2nd PNMT® ingredient: The Attention Mechanism 3/3
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**

<table>
<thead>
<tr>
<th>Source</th>
<th>Back Translated (EN&gt;FR&gt;EN)</th>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
<td>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.</td>
<td>+ Cognition transferred + Simplified lexicon + Simplified grammatical structure + More explicit and understandable</td>
</tr>
<tr>
<td>There is a possibility of prior Board approval of these investments.</td>
<td>The Commission may approve these investments.</td>
<td>+ Simplified + More explicit and understandable</td>
</tr>
<tr>
<td>Both experience and books are very important about living.</td>
<td>Experience and books are very important for life.</td>
<td>+ Corrected about living → for life</td>
</tr>
<tr>
<td>That’s why I like to make travel by using my own car</td>
<td>That’s why I like to travel using my own car.</td>
<td>+ Corrected to make travel → to travel by using my own → using my own</td>
</tr>
</tbody>
</table>
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