

SYSTRAN
beyond language

TAILORED AND SECURED

Open Source NMT The Race Is On

Presenters:

Jean Senellart

Date: 2018/05/17

www.systrangroup.com

Overview

- **SYSTRAN, OpenNMT and a short history of MT**
- **The rising of NMT and the battle of Open Source**
- **What is really at stake here?**
- **Are you we done with MT?**
 - The new perspectives

SYSTRAN: 50 years of Machine Translation



+140
Language combinations



+25%
Revenue invested
in R&D



Seoul
Paris – R&D
San Diego

Israel – Japan – China
Germany

Research Ecosystem



200
Employees

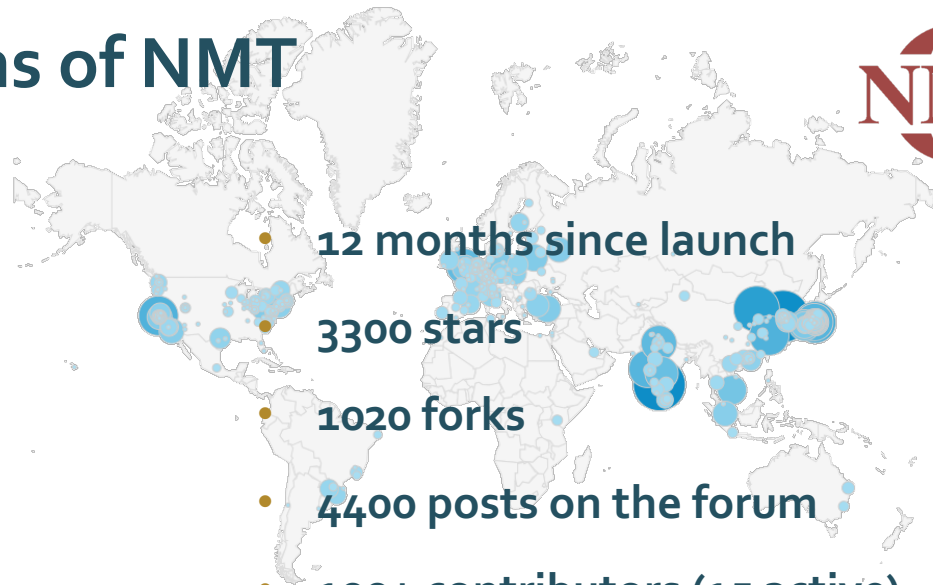


Leader in machine
translation and natural
language processing

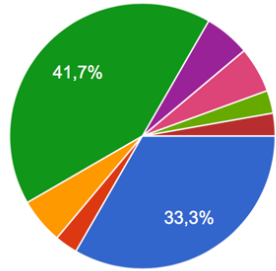
OpenNMT: ... 14 months of NMT



[ubiquus]



- 12 months since launch
- 3300 stars
- 1020 forks
- 4400 posts on the forum
- 100+ contributors (15 active)
- 18 major releases
- 6 complete code refactoring
- 600 unit tests



- Developer
- Linguist, Language specialist
- Translator, or Translation Project Manager
- Researcher / Academics
- Independent Expert
- Executive
- Hobbyist
- First two and last
- Software architect (researcher + de...)

Only ... 5000 lines of code

A Short History of Machine Translation

Neural Machine Translation

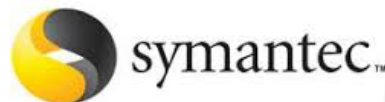
Phrase Based Machine Translation

Rule Based Machine Translation

1968



2007



2016



The previous rules of the game

Rule-Based MT era

- Large number of “small” industrial players that make regional baseline
- Asset is the code and the linguistic resources
- Subjective Quality Analysis

Statistical MT era

- Supremacy of Infrastructure
- One reference academic Open Source project – Moses
- Asset is the data
- Beginning of shared evaluation tasks

NMT era: new game, new rules

- Openness
- Very fast changing technology
- Data - a quest for quality
- Very quick transition from academia to industry

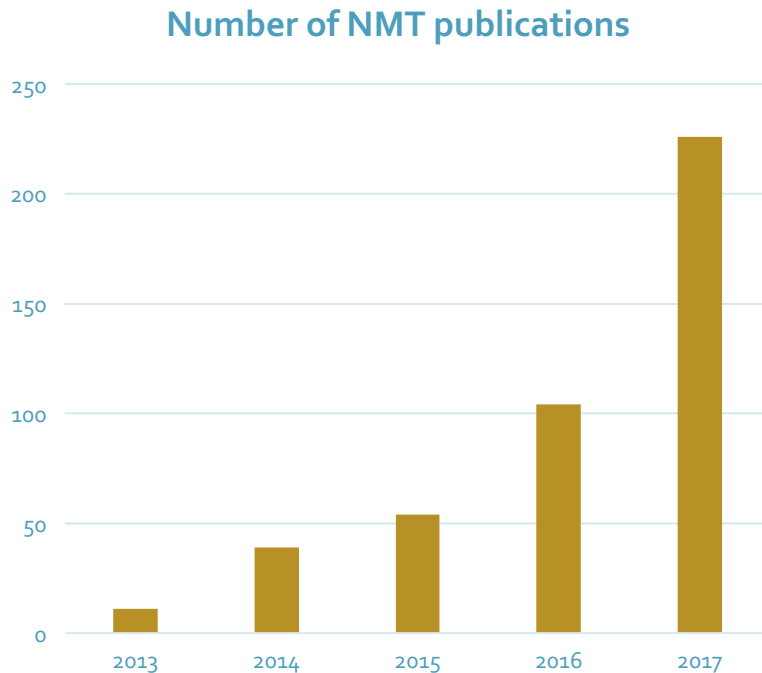
NMT Openness

- About 2 new Open Source projects per month
- Dominated by industry
- **Systematic Publications for industrial deployments**
 - GNMT, PNMT, Wipro, ...
 - Microsoft
- **Proof by number**
 - DeepL, Omniscient

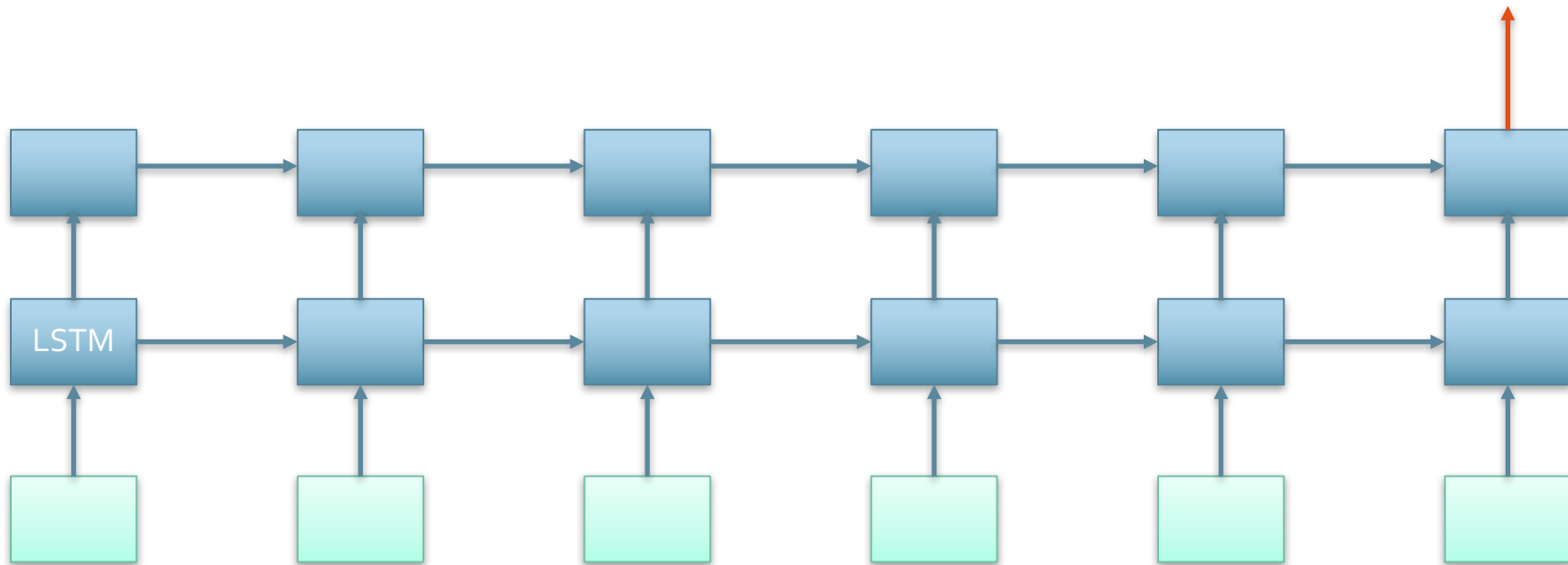


Very fast changing technology

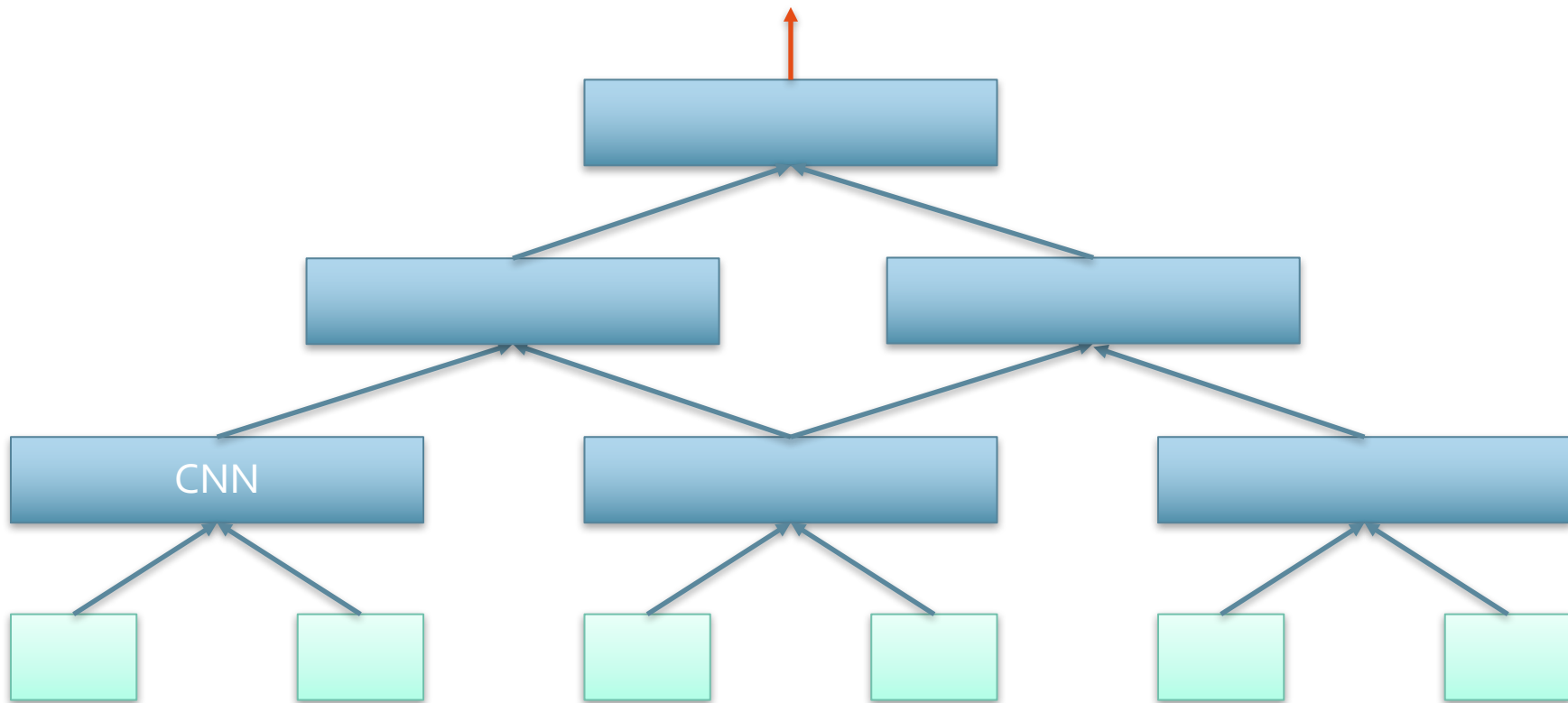
- Huge number of publications
- Major paradigm change
 - RNN
 - CNN
 - Attention-Based



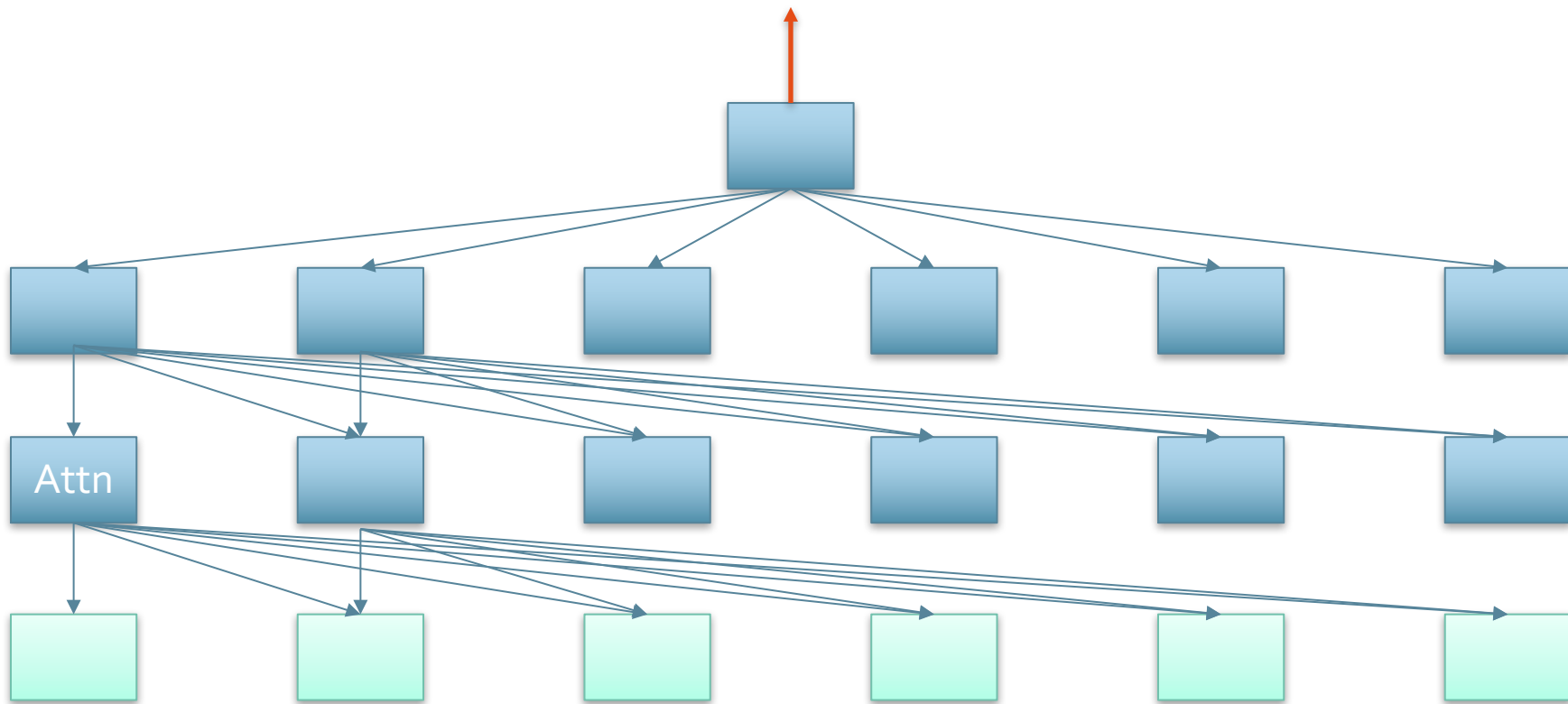
Recurrent Neural Network



Convolutional neural network



Attention-based neural network

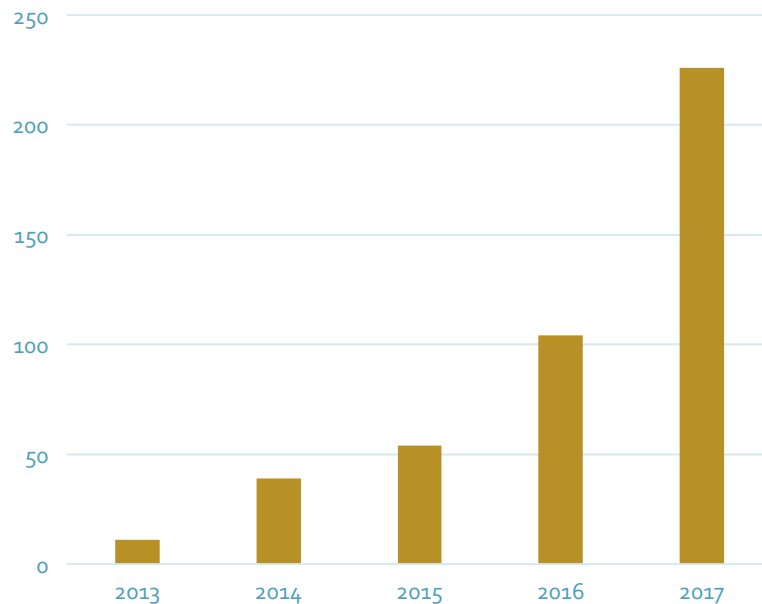


Very fast changing technology

- Huge number of publications
- Major paradigm change
 - RNN
 - CNN
 - Attention-Based
 - ...

➤ Code-base remains ... tiny

Number of NMT publications



Open Source survival rules

- **Open Source is not enough, the community expects:**
 - Daily support
 - Sharing of data, recipes and good practices
 - Sharing of negative findings
 - Integration of the latest, brightest publications
 - Modularity
 - ... Stability

- **A lot of work – why are we fighting for this?**

What is really at stake here?

- The real battle is on the AI framework and computation backend



- NMT is the gateway for all NLP technologies

➤ Like AI, NMT is becoming a commodity

What will be the winning computing framework?



ONNX



Facebook
Open Source



Microsoft

Standardization and Interoperability will enable NMT to reach a new maturity level

Are we done with MT?

- **Technology still at the beginning**
 - Promising developments on unsupervised NMT
 - NMT is still missing basic memorization abilities
 - Beyond sentence level
 - ...
- **Generic Translation is not a challenge anymore**
 - But expectations have increased along with translation quality
- **Specialized translation is a must have**
- Still a full time job to build solutions for 100 language pairs

Are we done with MT?

- **Open Source business model**
 - Give away recipes and code
 - Support ... competition
 - Constant evaluation and benchmarking
 - + **Receive community feedback**
 - + **Proven quality and expertise**
 - + **Expanding markets**

- **Toward Life long learning AI**
 - Can we train adapted models to each individual?
 - Can we train models with all available data and integrate daily new content?
 - Can we interact with NMT to increase our own language skills?

We are System, we love languages

