



Neural MT: towards Commoditization

New market challenges and opportunities for users

Jean Senellart
CEO, SYSTRAN



Luc Meertens
CEO, CrossLang

Outline of today's presentation

- **Share results of Language Technology Market Study**
- **Introduce SYSTRAN Marketplace**
 - Motivation
 - Giving back control to the real expert
- **Application Examples**
- **Launch, Recruiting and Business Model**

Language Technology Market Study



This study was part of a service contract supporting the EC in implementation of Automated Translation building block

Overall objective:

Contribute to the development of CEF Automated Translation as “multilingualism enabler” (beyond simple translation) for:

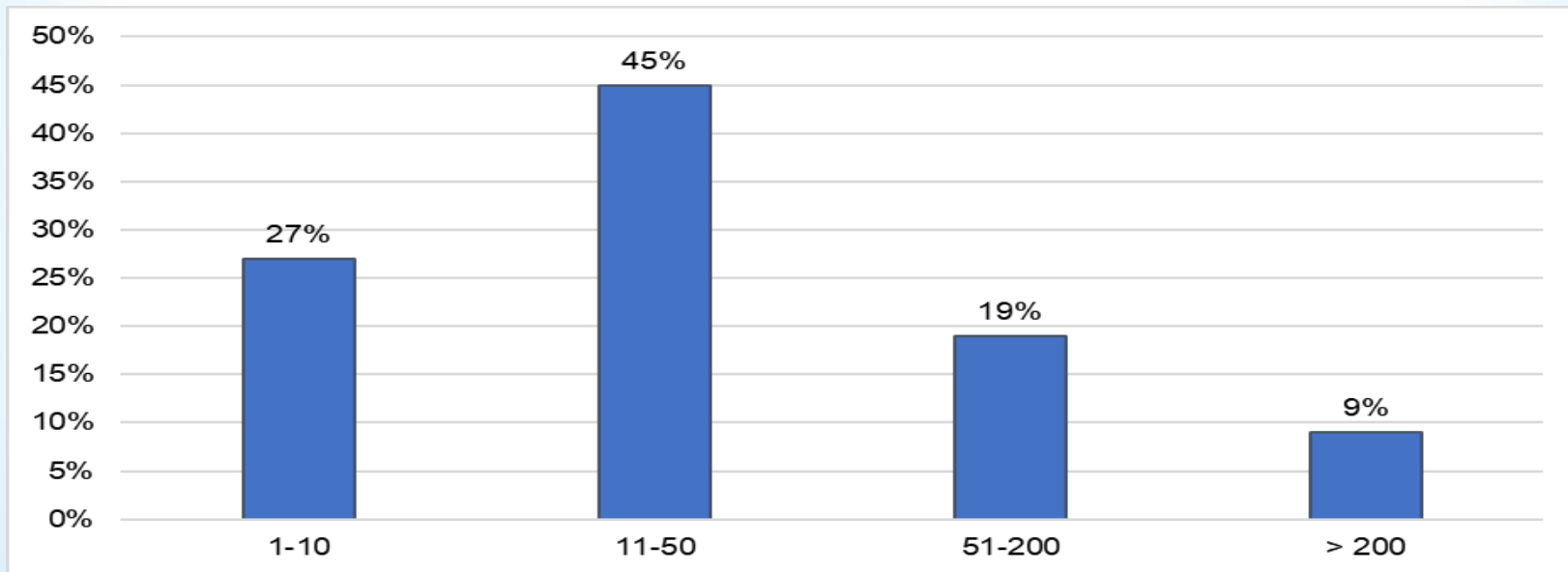
- CEF DSIs & Online services linked to CEF DSIs
- Other relevant public online services

Language Technology Forecast 2018-2020

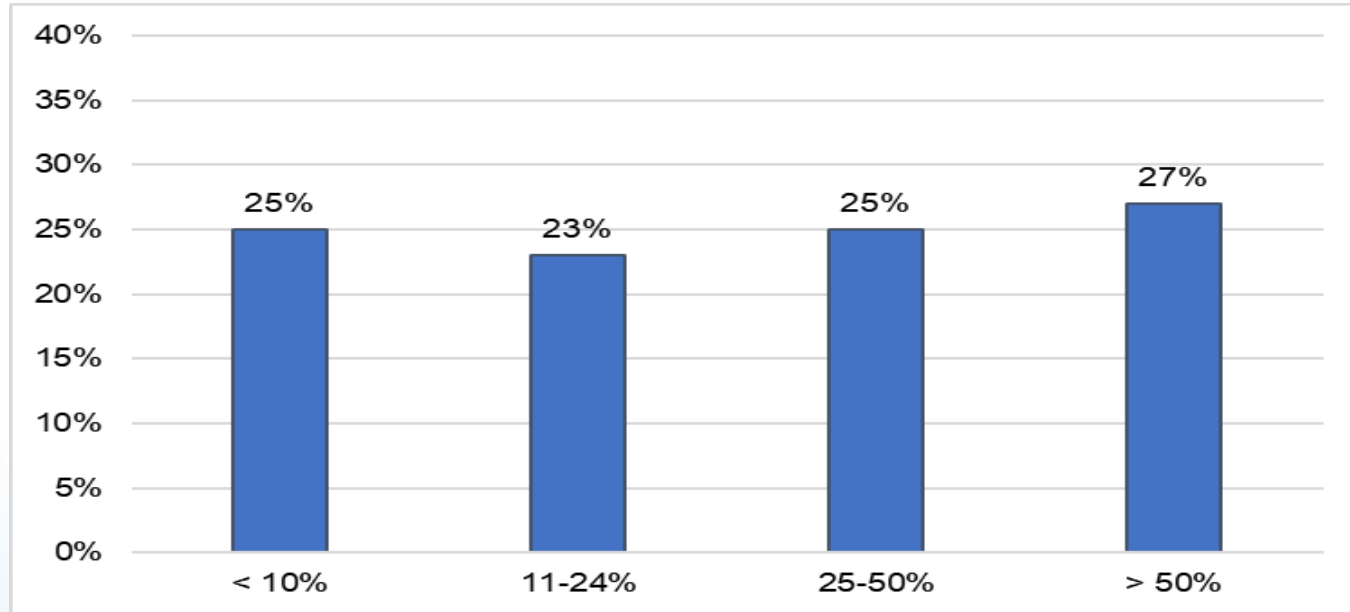


	2018	2019	2020
Germany	197	217	240
United Kingdom	189	209	232
France	88	96	105
Netherlands	55	60	66
Rest of EU 28	249	277	305
<i>In EUR million</i>	778	859	948

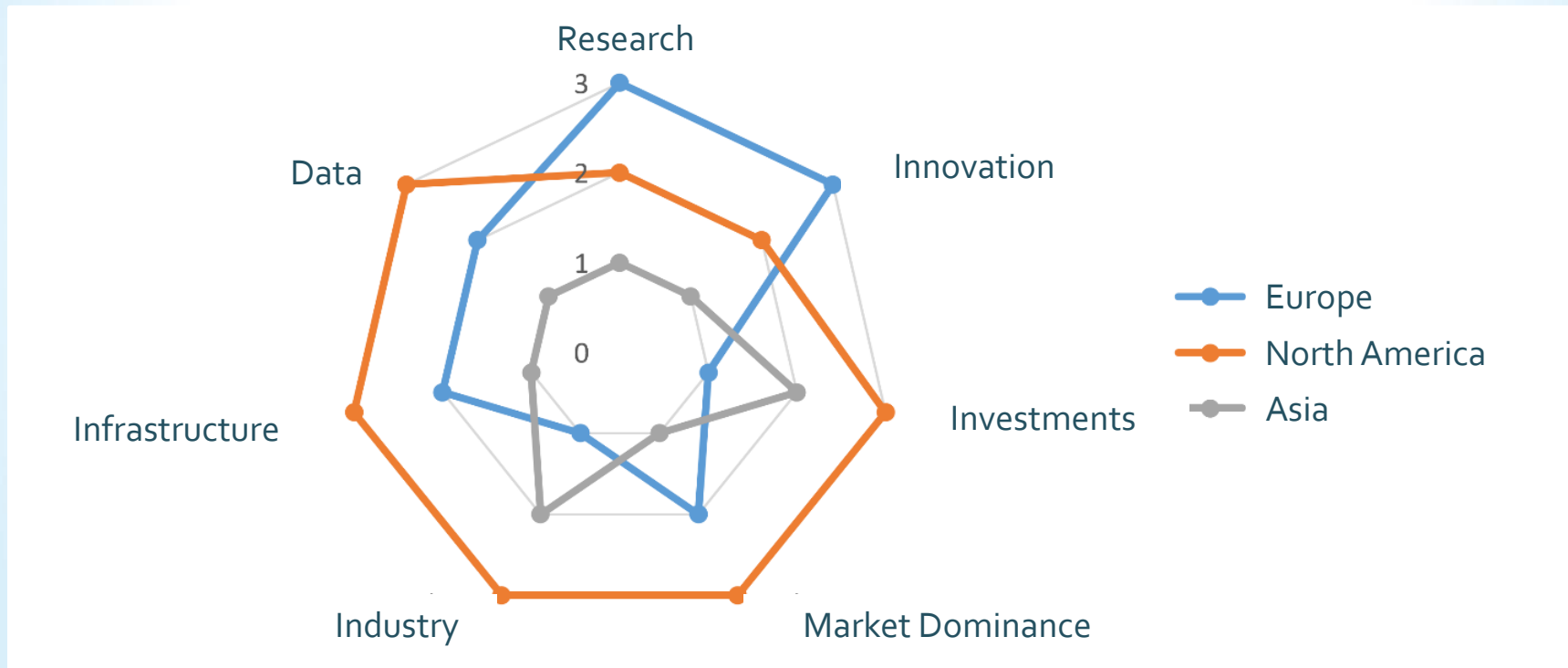
Supply side / Employee Numbers



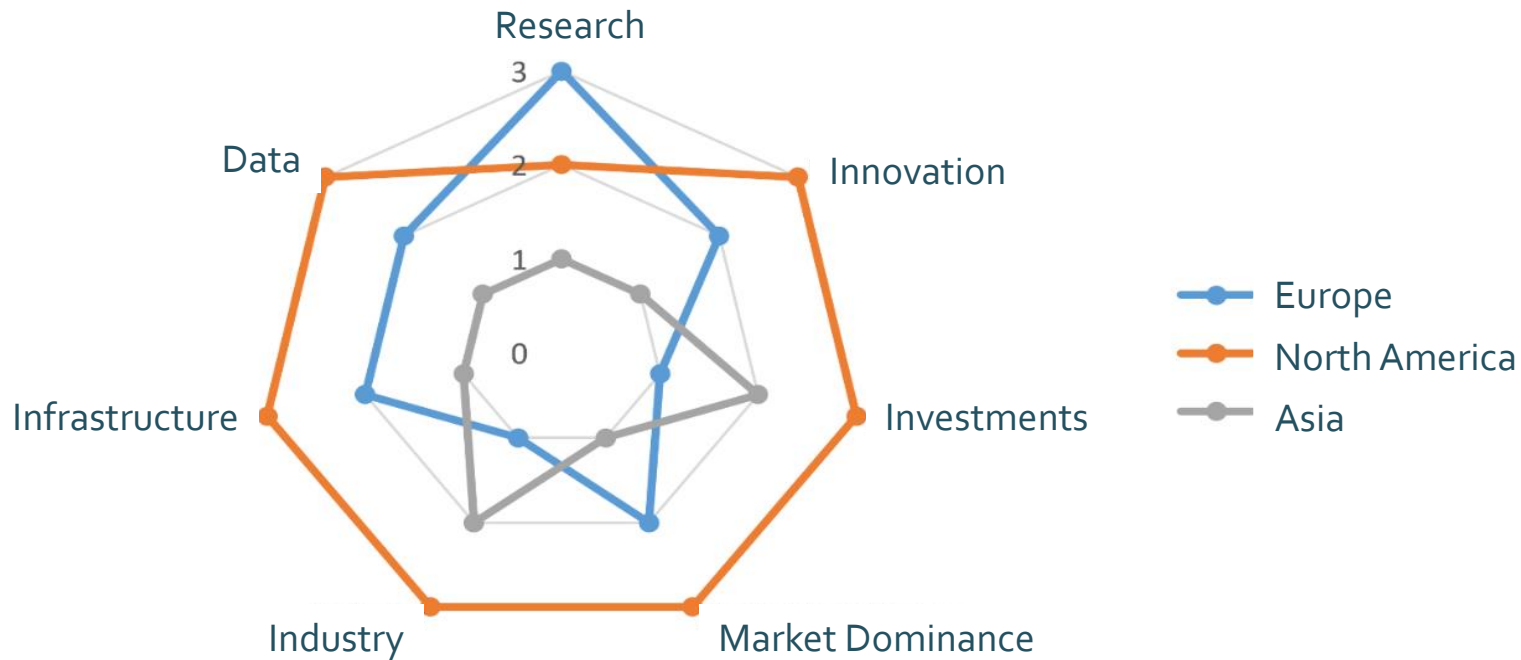
Supply side / Revenue Growth



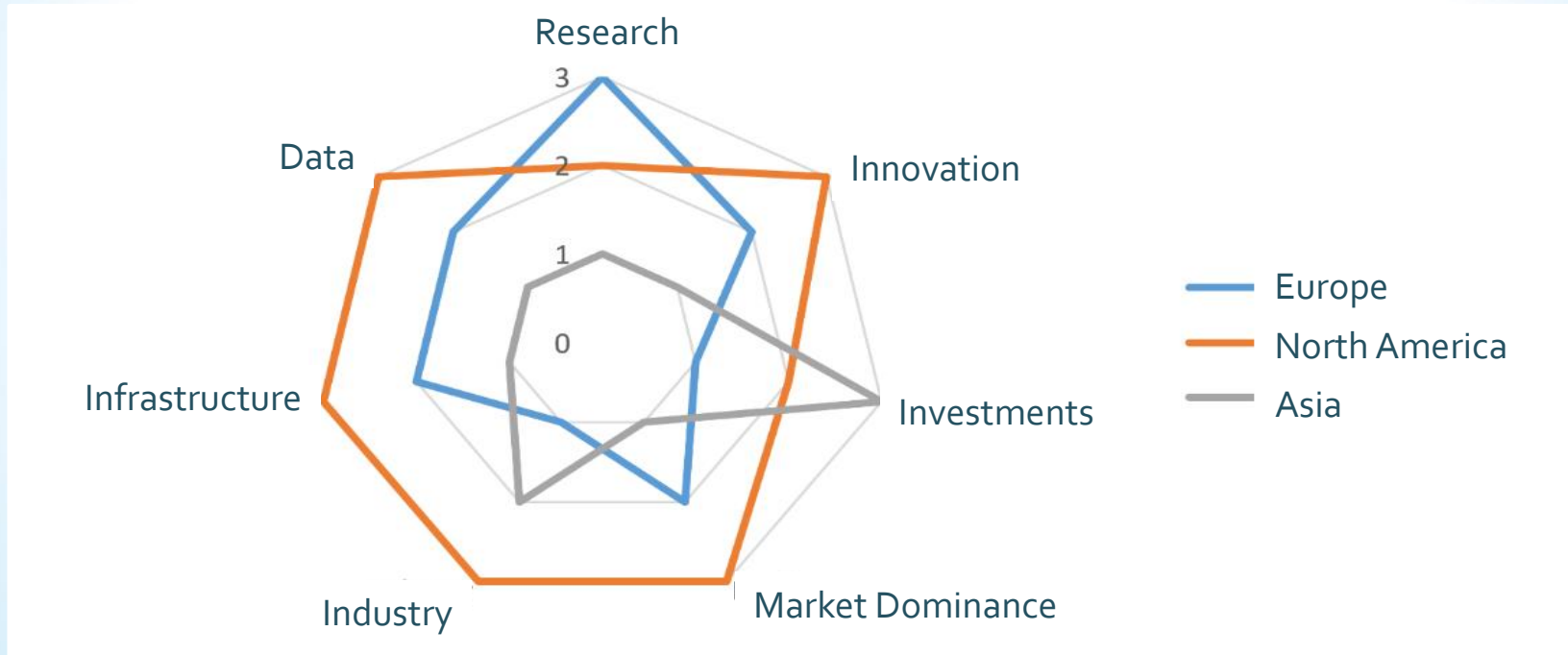
Competitiveness analysis : MT



Competitiveness analysis : Speech



Competitiveness analysis : Search



Competitiveness Analysis – score card




	Europe	North America	Asia
Machine Translation	14	19	9
Speech Technology	13	20	9
Search Technology	13	19	10
TOTAL	40	58	28



Karen Tso 
@cnbcKaren



The world's most valuable global brand...  is ...[#Amazon!](#) Here is the list from [#Brandz](#) [@WPP](#) [@Kantar](#)

1. Amazon
2. Apple
3. Google
4. Microsoft
5. Visa
6. Facebook
7. Alibaba
8. Tencent
9. McDonalds
- 10 AT&T

 5 7:47 AM - Jun 11, 2019



 [See Karen Tso's other Tweets](#)





Conclusions & Recommendations



Market is dominated by global players

MT market evolved in non-core environment

- Google, Baidu, Yandex : MT to sell ads
- eBay, Amazon, Alibaba : MT to drive e-commerce
- SDL : MT to drive localization business

Significant areas of market deficiency

- Quality gap for smaller and complex languages
- Domain and application specific MT
- Security and privacy concerns

Conclusions & Recommendations



Europe is strong in research & innovation, but **not successful to scale innovations** and capture the market

Europe needs European alternatives to fill the gaps/deficiencies and to avoid reliance from market monopolies

Multilingual DSM to be developed on **own LT infrastructure**

Public procurement can be the major driver for European LT industry

Focus on Governance

- Provide MT models / LT models (API access)
- Provide data for training models

We listened and worked hard to prepare the launch of the *LTI Cloud* project at our upcoming Summit.

The *LTI Cloud* is a SaaS wrapper around LT components with the goal of facilitating the plug 'n' play of language technology. It will allow you to market your capabilities, to easily build new LT components based on others... and it will be a great new customer acquisition channel.

We intend it to become a single point of access for companies wishing to implement language technology, empowering people and businesses by simplifying digital language technology solutions in Europe and globally. So far, we have architected the *LTI Cloud*. Influential politicians love the idea and will come to the LTI Summit to support us.

We would like to share this project with you. It is, therefore, very important that you come to the LT-Innovate Summit on 25-26 June in Brussels to show your interest and support. Only together can we push the project ahead and launch soon the first version of the *LTI Cloud*!

We count on you and are looking forward to seeing you soon in Brussels,



Jochen Hummel
CEO, ESTeam and
Chairman, LT-Innovate



Robert E. Etches
CIO, TextMinded



Luc Meertens
CEO, CrossLang



Christoph Prinz
CEO, SailLabs

The LTI Cloud is the one-stop-shop platform for making available, discovering, assembling, testing and prototyping language technology components.



Knowledge Management



Natural Language Processing



Search



Translation

Welcome to the Catalogue of CEF eTranslation Services.

Multilingualism has become an essential part of our lives, which is why enabling seamless cross-border communication has never been more important. Therefore, the catalogue of services includes a wide range of tools and services which aim to help you overcome language barriers and to support your multilingual needs. Browse through our catalogue, explore the language tools (LT) and services offered by [CEF eTranslation](#) and discover other potentially useful third-party services* which can facilitate your multilingual operations.

Explore the services provided by CEF eTranslation [here](#)

Explore third-party services [here](#)

Browse the [ELRC-SHARE repository](#) for Natural Language Processing tools/services

Results 8 (1/1)

Filter By:

Service Function

- Translation (4)
- Language Technologies (3)

[MORE](#)

Intended User

- Any third party interested in training MT engines (1)
- CEF Automated Translatio (1)

[MORE](#)

Language Coverage

- English (6)
- French (5)

[MORE](#)

Service Provider

- Administrative Modernizatoin Agency (1)
- Austrian Presidency of the Council of the European Union (1)

[MORE](#)

Service Ownership Status

- External (8)

[Clear Filters](#)

Translator of the Bulgarian Presidency of the Council of the EU

Function: Translation

Task: Machine Translation

By Ministry of the Presidency, Bulgaria

Translator of the Austrian Presidency of the Council of the EU

Function: Translation

Task: Machine Translation

By Austrian Presidency of the Council of the European Union

Latvian state administration language technology platform

Function: Language Technologies

Task: Speech Recognition , Text To Speech Synthesis , Machine Translation

By Culture Information Systems Centre, Latvia

ParaCrawl OpenSource Pipeline

Function: Language Technologies

Task: Crawling , Alignment

By Universitat d'Alacant

PLaTa: Spanish state administration Machine Translation Platform

Function: Translation

Task: Machine Translation

By Ministerio de Política Territorial y Función Pública, Secretaría de Estado de Función Pública, Secretaría General de Administración Digital

eTradução - Repositório Nacional de Recursos de Tradução

Function: Data And Data Management

By Administrative Modernizatoin Agency

Towards the Primary Platform for Language Technologies in Europe





SYSTRAN
Market place

The professional Choice

Challenges with a Generic LT Cloud



API

Not all services have available APIs
API not normalized – missing standard



Infrastructure

Scattered Services with restricted scalability



Service discovery

How will end-user discover about the service?



Payment solution

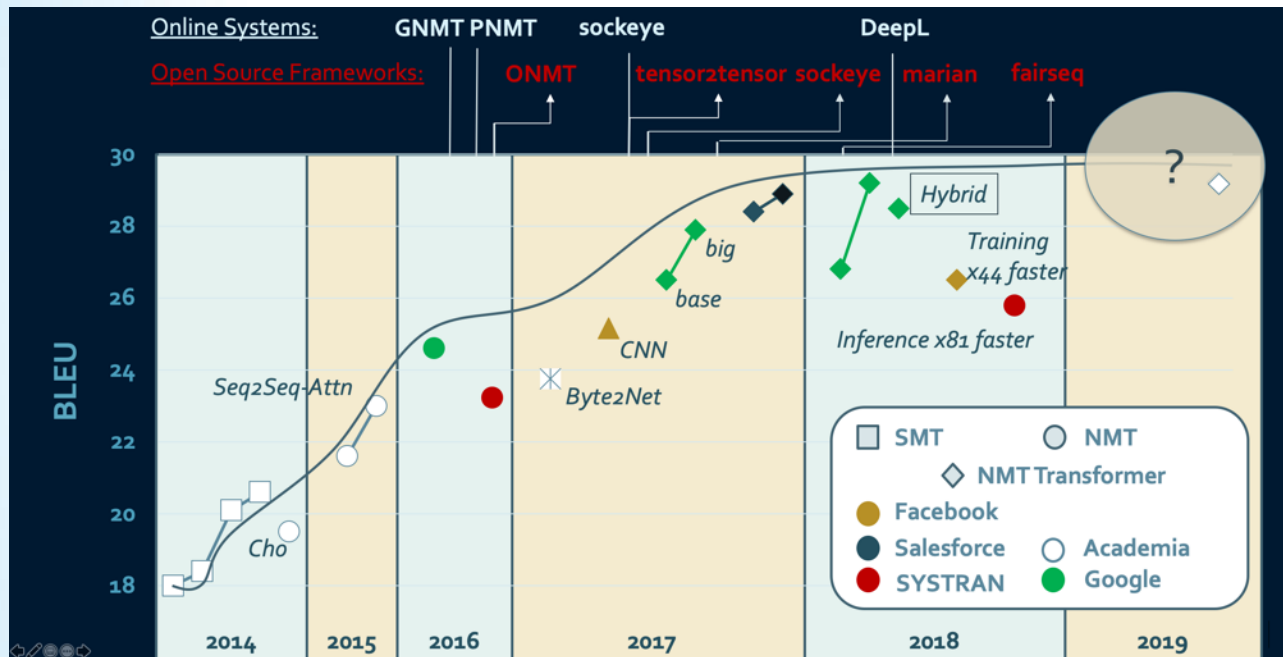
Need a normalized and simplified payment model and solution



Feedback loop

Hard to set-up functional support and feedback workflow

Machine Translation Technology today



- Incredible Progress on core technology
- Plethora of OSS implementations
- More than 6000 papers published in 2018
- Available through Online Services
- Evidences of "Human Parity"

👉🗉 MT has never has been so good – but is it really available to the industry?

Options for the industry today



**Existing Cloud
Solution**



**Do-It-Yourself
Solution**



**Customized solutions
from a pure player**

Options for the industry today



Existing Cloud Solution

PRO

- Instantaneous Set-up
- (Generally) State-Of-the-Art quality on generic translation

CON

- Constrained to one environment
- No control on terms and conditions
- No interoperability
- Low Customization
- No way to submit feedback
- Access to Technology and not Product

Options for the industry today



Do-It-Yourself Solution

PRO

- Full control on the process
- Possibility to optimize costs

CON

- Hard to follow technology evolution
- Required in-house Language Aware Data Scientists
- Scalability Costs
- DIY technology → ... DIY Product

Options for the industry today



Customized solutions from a pure player

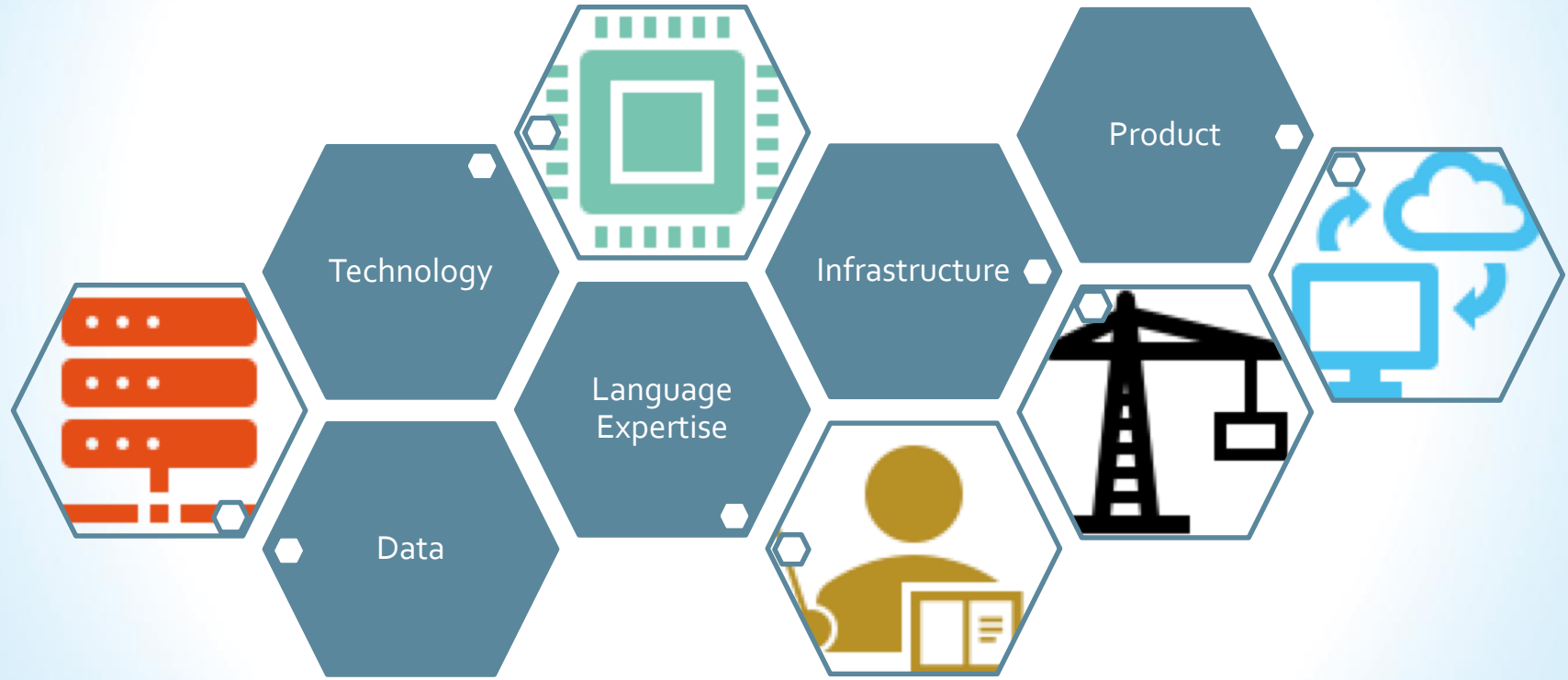
PRO

- Key-turn solution
- Evolving Best-of-Breed technology
- Exchange with expert
- Bullet-proof production

CON

- Entry cost limits to large player - difficulty to scale down
- Limited by vendor language coverage and/or service capacity

The bricks of a full end-to-end solution



Technology

Core Technology

- Open Source dominating the market



Full process to train engine is going beyond core technology

- Preprocessing/postprocessing
- Data selection
- Engine Evaluation
- Computing Resource Allocation
- Releasing model for production
- Storing and selecting models

Data Panorama

Large “open” data from institutions

- Generally out-of-domain
- Massively acquired - medium quality

Online self-crawled data

- Main Question about IP

Institutional data providers

- Not sharing to avoid losing value
- Or sharing assets but with non-commercial licenses

User Data

- Privacy and GDPR Requirement

No Traceability on corpus acquisition and intellectual property

Training Models – putting the expert in the loop

- Non customized-models are **impressive** but **useless** in actual production cycle
- Training Neural Model requires Language Expertise and domain specific focus
- Model-owner should be able to specialize any existing model to a specific domain, and handle specific feedback



Infrastructure for training and Deploying – is it that easy?

Training

Finally, for WMT English-German we train on all 226M available monolingual training sentences and perform 250K updates in 22.5 hours on 128 GPUs. We upsample the bitext with a rate of 16 so that we observe every bitext sentence 16 times more often than each monolingual sentence. This results in a new state of the art of 35 BLEU on newstest2014 by using only WMT benchmark data.



~9000\$ to reproduce a SOTA training on public cloud
Can go down to ~4000\$ with in-house hosting but > 40 full days of training

Inference

Free translation necessary to drive business

Deploying neural model is easy however scaling up to free web-translation not economically possible for non-GAFA

➤ Need mutualization of infrastructure

Community-based

Responsible Platform



institutions

corporates



data providers



LSPs

freelance translators

academics



Open Source contributors



SYSTRAN
Market place

*The professional
Choice*



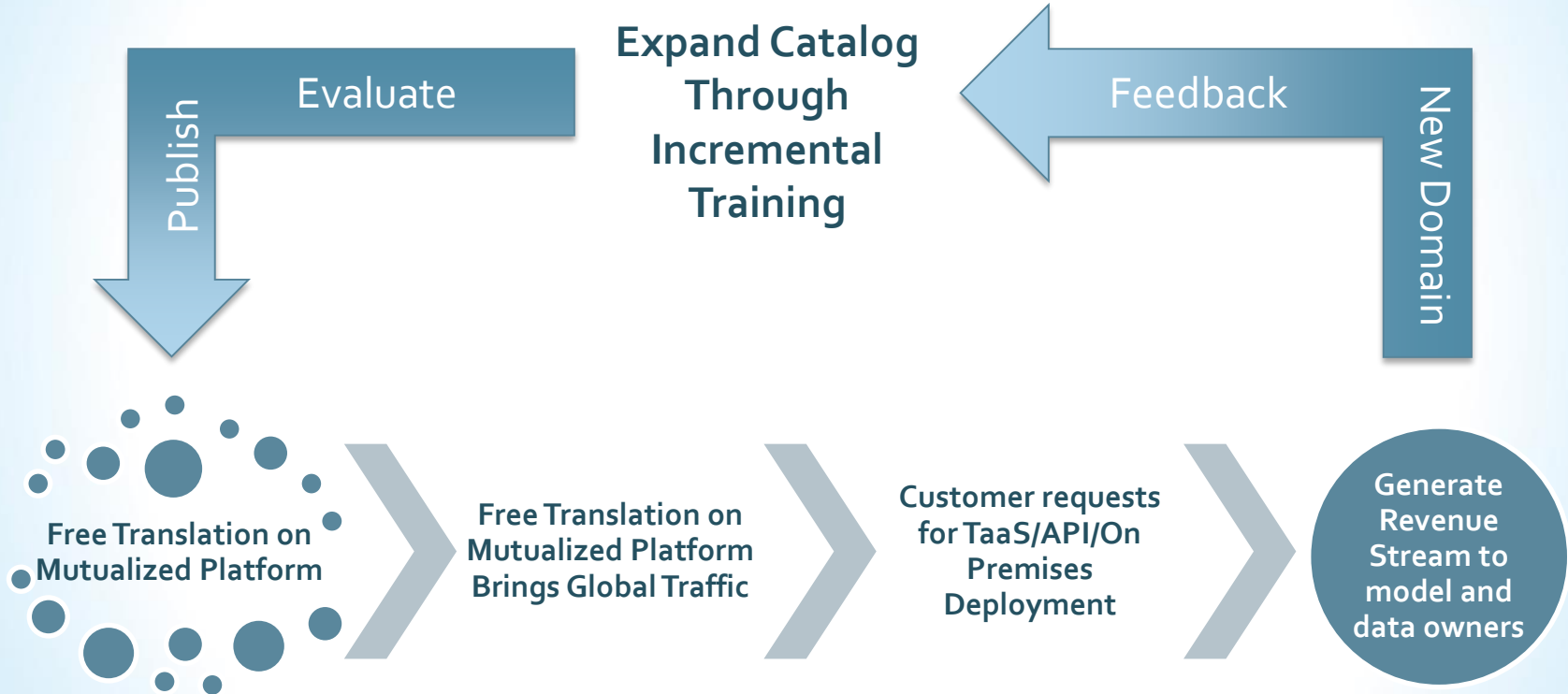
DATA PROVIDERS



TRAINERS



Toward Responsible and yet State of the Art approach



Use Case Example and benefits for Trainer/Data Provider



TRANSLATION AS A
SERVICE FOR **LEGAL**
DOMAIN



API INTEGRATION



DEPLOYMENT ON
PREMISES
FOR **IT DOMAIN**

Translate legal documents automatically

Find out more

Try it now



Highly accurate

Making use of a cutting-edge machine translation system, Judicio gets it right almost all the time. ▶



Incredibly fast

With an average translation speed of 3.000 words/minute, Judicio tackles all time-critical projects. ▶



Safe & Secure

All connections are fully secured and translations are not kept on our server. ▶



Affordable, flexible plans

Our flexible pricing model can be tailored to your individual business needs and budget. ▶

TO TRANSLATE FULL DOCUMENTS WITH JUDICIO,

[Log in with your account](#) ↗

[Contact us](#) for a free trial

Curious about how well judicio can translate law texts?

See for yourself and try it out with a sample text of your choice.

No account is needed for this **demo**.

The demo translation is **limited to 1000 words** per text.



🌐 Dutch ▼ → French ▼

Paste text

Paste your text here

- French
- German
- English

[Clear text](#)

TRANSLATE

By pressing this button, you agree to the [Terms of Service](#).

Lawyers deserve better word processing than MS Word

Work more efficiently every day as of now.
And achieve even better results.



Review
✕

Role	Approver
Folder	My documents
Title	PRESENTATION - Scientific publication - FOR TRANSLATION

Edit metadata

Show suggested changes

Magic tabs

Export

Identify legal references by LeReTo

Check NDA - powered by ThingsThinking

Translate Text - powered by DeepL

Sign electronically - powered by Signaturit

Translate Text to German - powered by CrossLang

Translate Text to English - powered by CrossLang

Manage Users

Invite Users

Publishing special characters

Presentation in document: Icons ▾



Scientific publication

Sample document

Autor: Author 1

Autor: Author 2

Summary: This document shows, among other things: Formulas as paragraphs and inline, cross-references to an image, a table, a formula, an entry in the literature directory and a numbered heading.

1 Formula, footnotes and references to literature

Far behind, behind the mountains of the word, far from the countries of vocals and consonants, the blank texts live¹... They live secluded in letter-houses on the coast of Semantik, a great spray-blast.

A small creek called Duden flows through its place and provides it with the necessary rules. It's a paradisiacal country where a fried piece of phrase is put into the mouth.². Even the omnipotent puncture does not control the text of the blind. [1], [5] and [8] - an unorthographic life. But one day, a little line of blind text decided, her name was Lorem Ipsum, to go out into the wide grammar.³.

$$D = \int \frac{N_E \cdot S_{col,E} \cdot dx}{\rho \cdot dx \cdot dA} dE = \int \frac{S_{col,E}}{\rho} \cdot \Phi_E dE$$

Deployment on Premises of customized model



LARGE SOFTWARE COMPANY HAS TM'S AVAILABLE OF ON AVERAGE 500K-700K SENTENCES FOR EN-PL, EN-CS AND EN-HU



WE ADDED DATA TO GENERIC SYSTEM, THUS CREATING A CLIENT-SPECIFIC ENGINE (+OVERSAMPLING)



MODELS PUSHED TO MODEL CATALOG AS PRIVATE MODELS AND DEPLOYED ON PREMISES

The SYSTRAN Marketplace today

30,000 models available from 110 language pairs in model catalog

Equivalent of 15 GPU-year training

Available as base models



Core-Network of worldwide trainers and data providers

Covering 10 different markets

In expansion - please join!



Deployment infrastructure ready for 5-billion word per day

The SYSTRAN Marketplace Business Model

Data Provider

- Leverage Existing Data and Generate Revenue
- Keep Full IP
- **Invest** on Data Improvement

Trainer Network

- Build or Customize model with “Train as a Service”
- Royalty Based model (70% of revenue)
- Provide paid customization service

SYSTRAN

- Maintain core Technology and provide Infrastructure
- Sell Translation Product
- Connect Trainer and Customer

MT Customer

- Try and Select for free in **largest** translation catalog
- Use MT as a service or deploy on premises
- Provide feedback and Request Customization



SYSTRAN
Market place

The professional choice

→ A multi-phase launch

FIRST RELEASE

July
1st



- The 1st free online translation solution offering specialized translation models that have been trained by a worldwide expert community
- Test hundreds of translation models