Iconic
Translation Machines

"From the Lab to the Market" Commercialising MT Research

John Tinsley
Director / Co-Founder

What is Iconic Translation Machines?

We provide Machine Translation solutions with Subject Matter Expertise

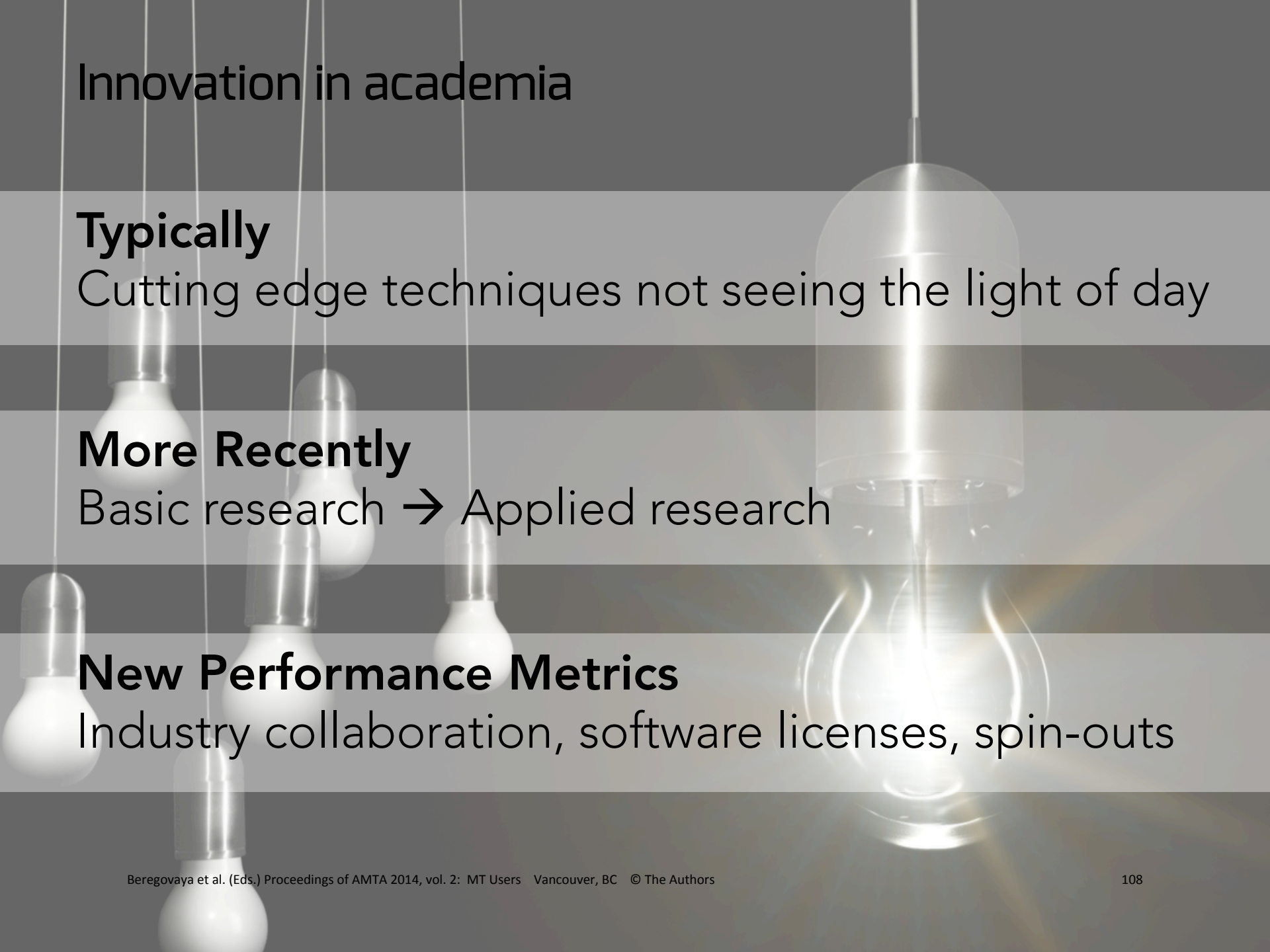
HQ: Dublin, Ireland

Founded: 2012

For the next 20 minutes…

Part 1: The journey from the lab to the market

Part 2: The technology that took the journey



Innovation in academia

Typically
Cutting edge techniques not seeing the light of day

More Recently
Basic research → Applied research

New Performance Metrics
Industry collaboration, software licenses, spin-outs

Lab to Market: The Starting Point… The Lab

The Lab
Dublin City University

The Funding
European Union (FP7 PSP)

The Goal
Adapt existing technology for patent machine translation

So what now?

PLuTo
Patent Language Translations Online

Lab to Market: Technical Development

The Process
Build with working group: release early, release often

Engagement
Identify broader user base (patent professionals, translators) and field test

The Outcome
Well-developed, well-performing prototype with a user base

Is there a business in this?

IPTranslator.com

Lab to Market: Commercial Viability

Feasibility Study Grant

Commercialisation Fund

iHPSU

Skillset

Is this worth exploring?

- Develop MVP
- Business model
- Product/Market fit?

- License IP
- Spin out

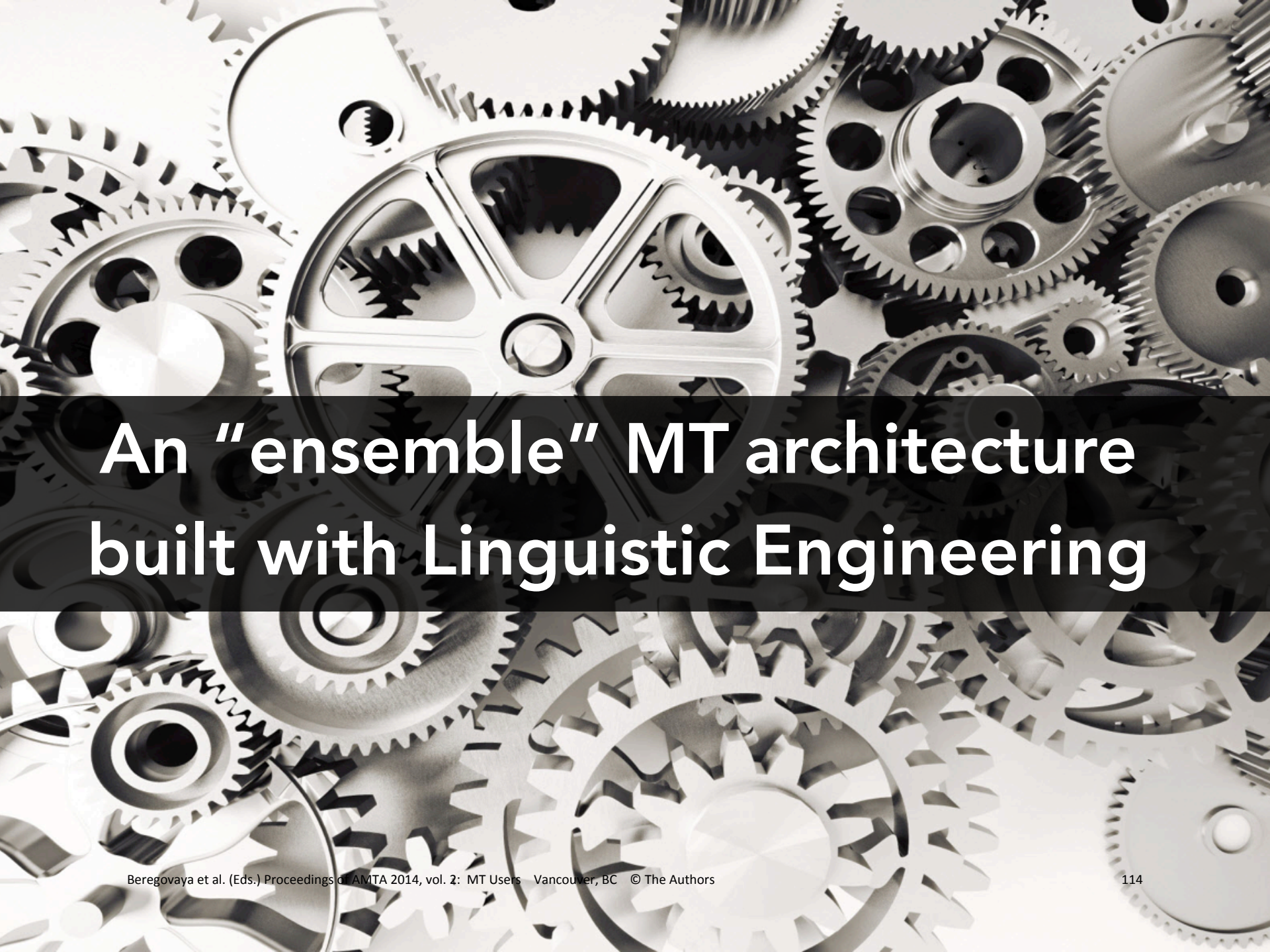
- Support
- Exporting
- Investment

Lab to Market: ... The Market!

Part 1: The journey from the lab to the market

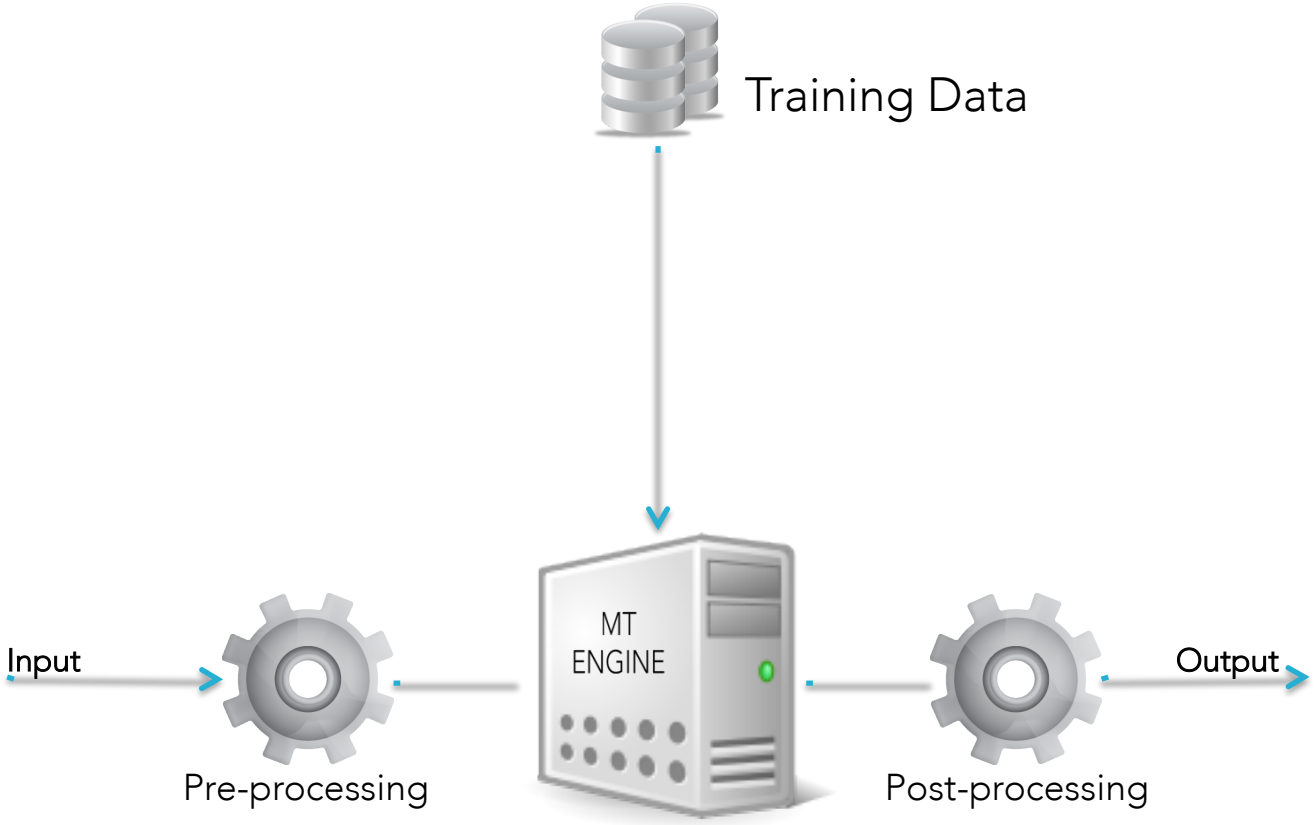
Part 2: The technology that took the journey

We provide Machine Translation solutions with Subject Matter Expertise


An “ensemble” MT architecture
built with Linguistic Engineering

What is Linguistic Engineering?

Data Engineering



The Challenge of Patents

Long Sentences

US 7,752,326 B2

41

network connections and preferably but not necessarily pre-emptive multitasking and multithreading. The interface of the present invention as it appears to the outside world (i.e. programmers and developers who provide access to end users and programmers who provide Content Access Modules to Content Engines used by content publishers) is independent of both the operating systems and the programming languages used. Adapters can be built allowing the tiers of the system to cooperate even if they use a different operating system or a different programming language. The protocol of the present invention can be implemented on top of networking standards such as TCP/I P. It can also take advantage of inter-object communication standards such as CORBA and DCOM. The object model of the present invention can be

42

a requesting client of the plurality of clients, that provides an input field and that transmits to the server within the user session a plurality of queries to retrieve content from the server, wherein each of the plurality of queries are consecutive and form a lengthening query string for retrieving content from the server; and wherein each subsequent one of the plurality of queries extends the lengthening query string by one or more additional characters; and

a server, which is configured to access one or more content sources that store content and that can be accessed by the server to respond to the queries from the clients, wherein the server further includes a unified query and result cache common to the plurality of clients and that stores

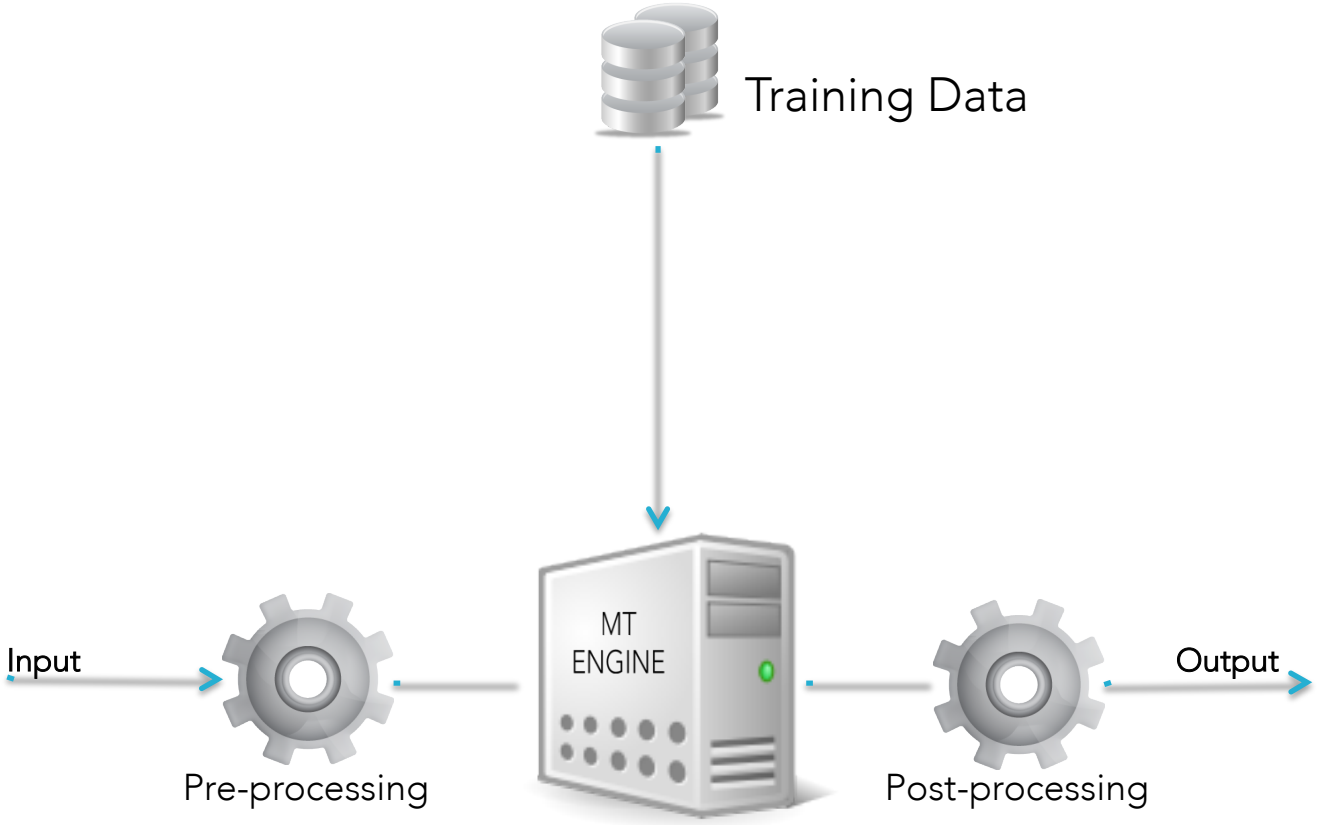
Longest Sentence: 1,417 words

Largest single document: 249,322 words

maximum stress of 1.2 to 3.5 N/mm<2> and a maximum elongation of 700 to 1,300% at 0[deg.] C.

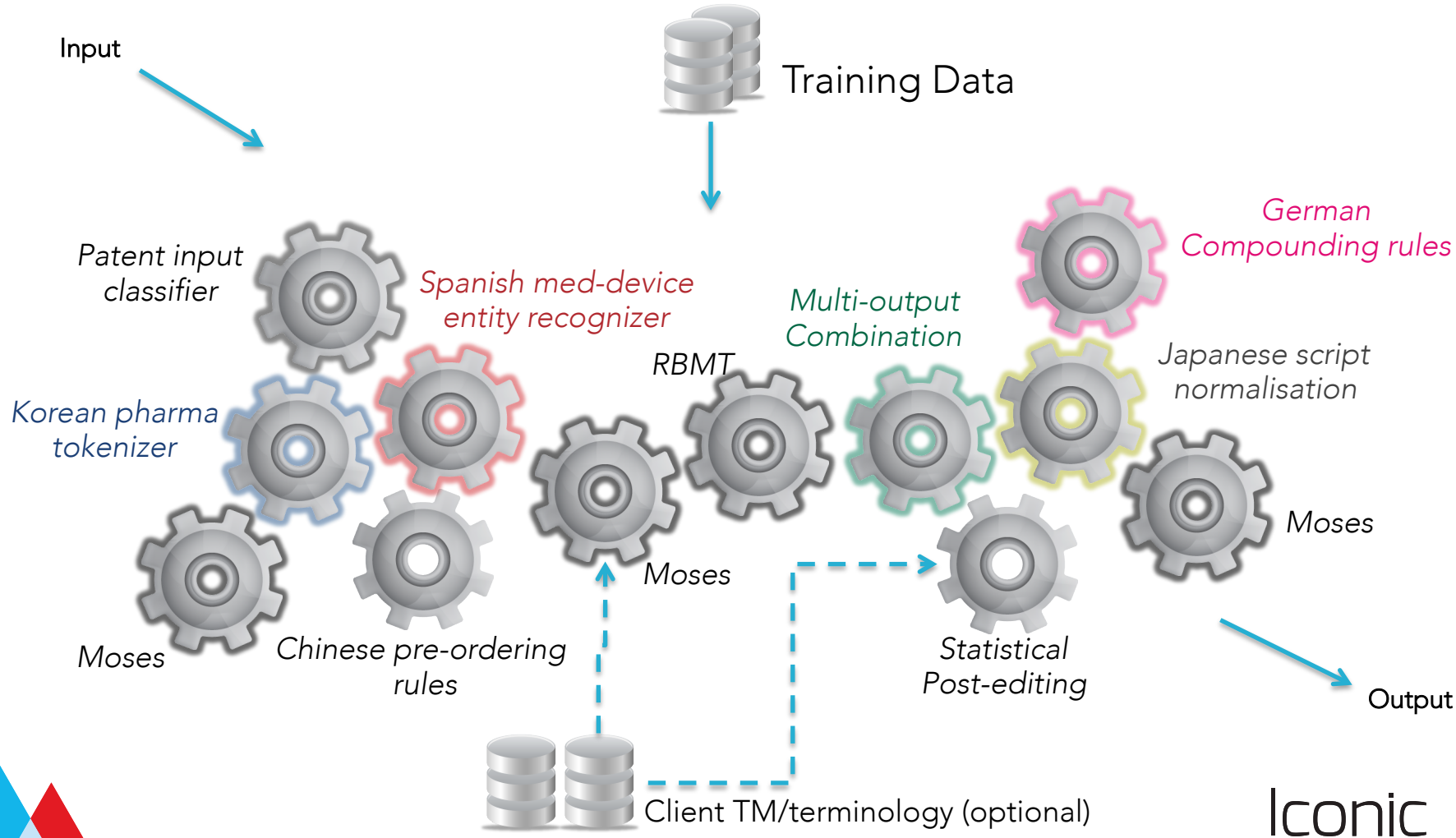
What is Linguistic Engineering?

Data Engineering



An “ensemble” architecture

Data Engineering + Linguistic Engineering



Iconic
Translation Machines

What is the value for users?

De-risking the machine translation proposition

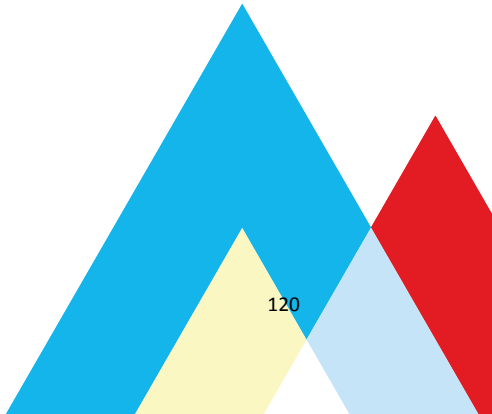
Typical Prerequisites	Our Prerequisites
+ Data	+ No data needed
+ Time	+ Systems are ready to go
+ €€€	+ No upfront cost
= ???	= Evaluate immediately

Customisation. Refinement.

» Incorporation of user feedback
» Incremental training with post-edits
» Tuning for specific input types

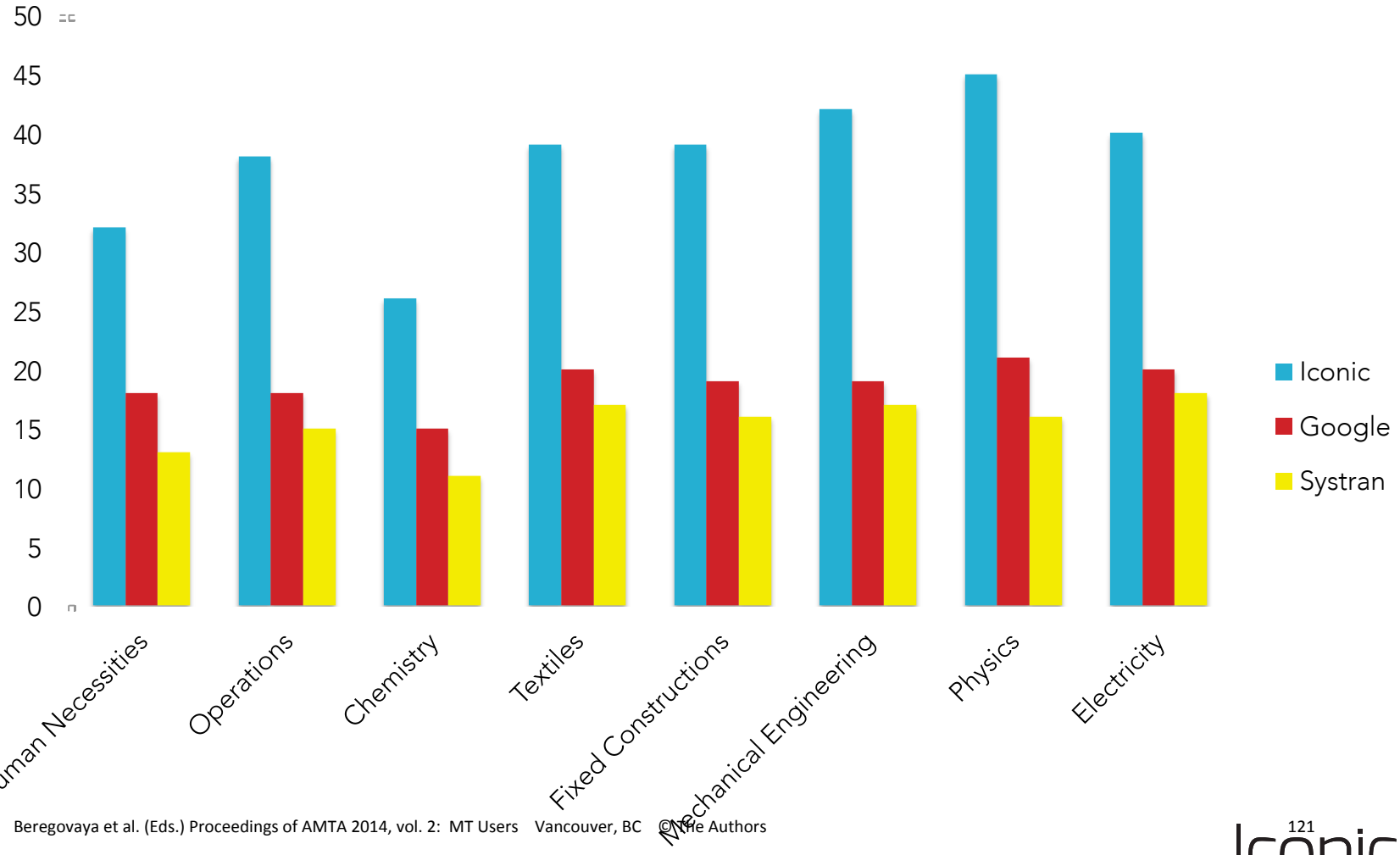
Case Studies

1. What this approach means straight up in terms of quality…
2. Productivity gains from using these systems…
3. As a foundation for client customization…



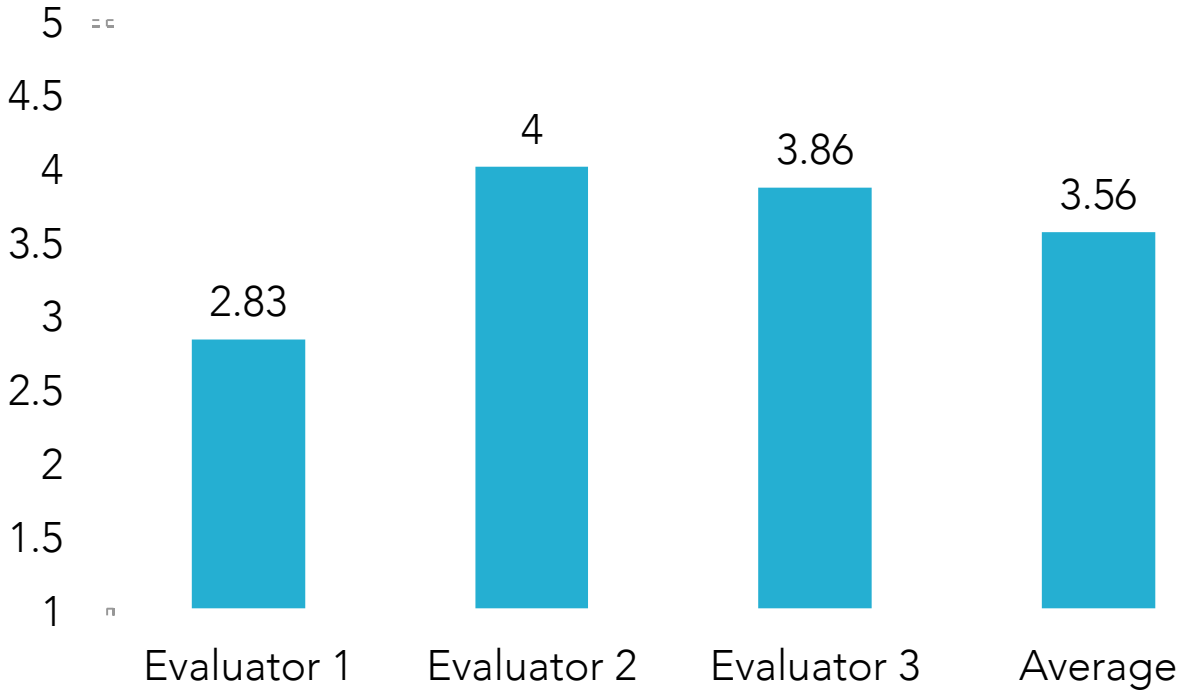
Case 1: Quality

Portuguese to English



Case 1: Quality

German to English



Case 2: Productivity

Business Need

Machine Translation technology for the legal industry

Iconic had a domain-specific MT solution for that industry

welocalize

Case 2: Productivity

Process

Translation samples required for initial evaluation

Delivered immediately and initial results were positive

welocalize

Case 2: Productivity

Performance

"MT delivered measurable productivity gains from the outset"

>20% productivity increase for translator post-editing Iconic output

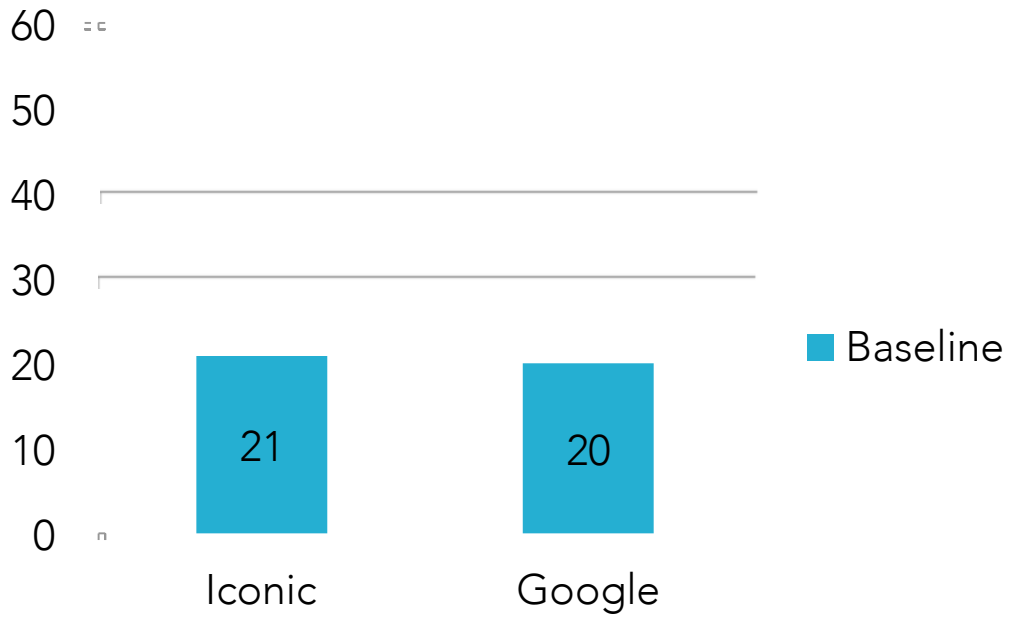
welocalize

Case 3: Customization

Chinese to English

- Modify our patent machine translation engines for "Written Opinions" on patents

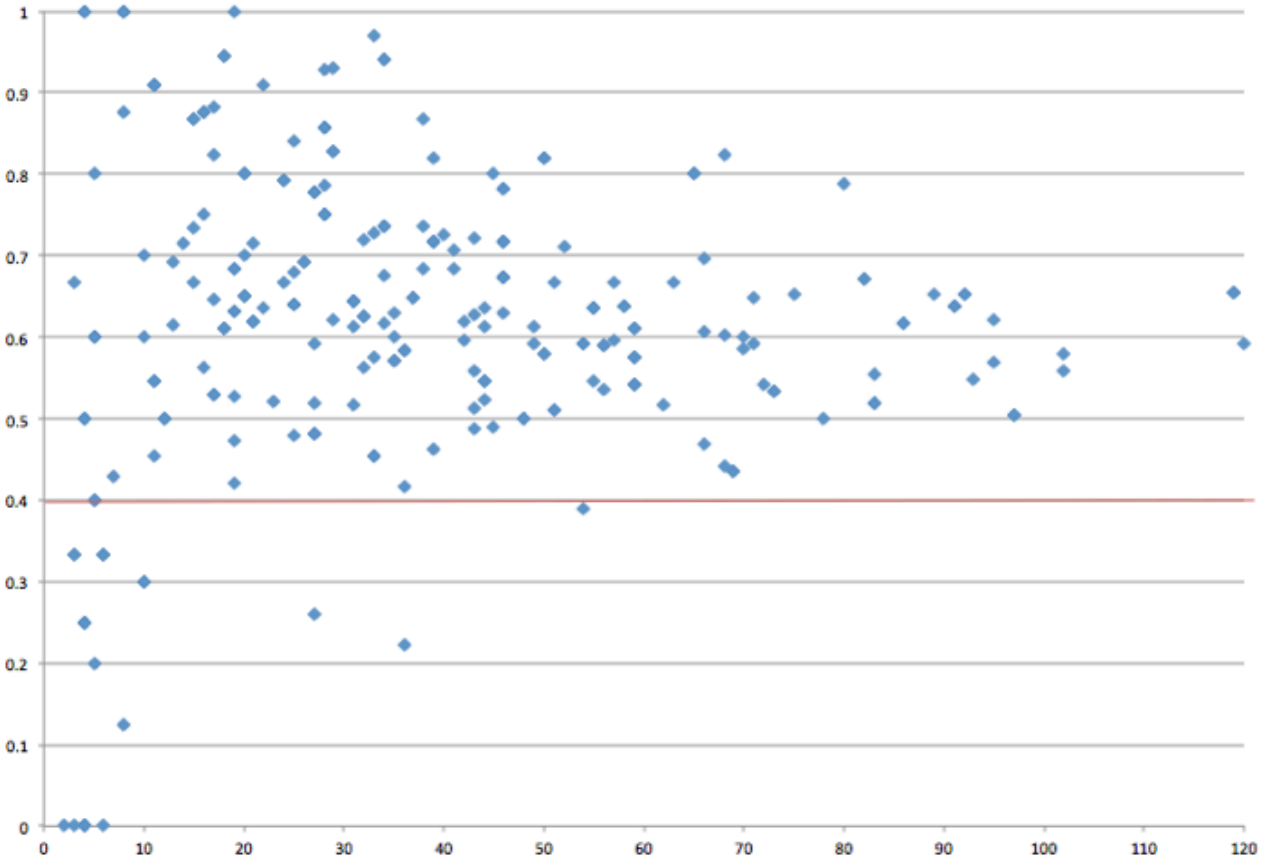
- 0.25% new data, 2 new ensemble processes



Case 3: Customization

Essentially out of domain – not viable for post-editing

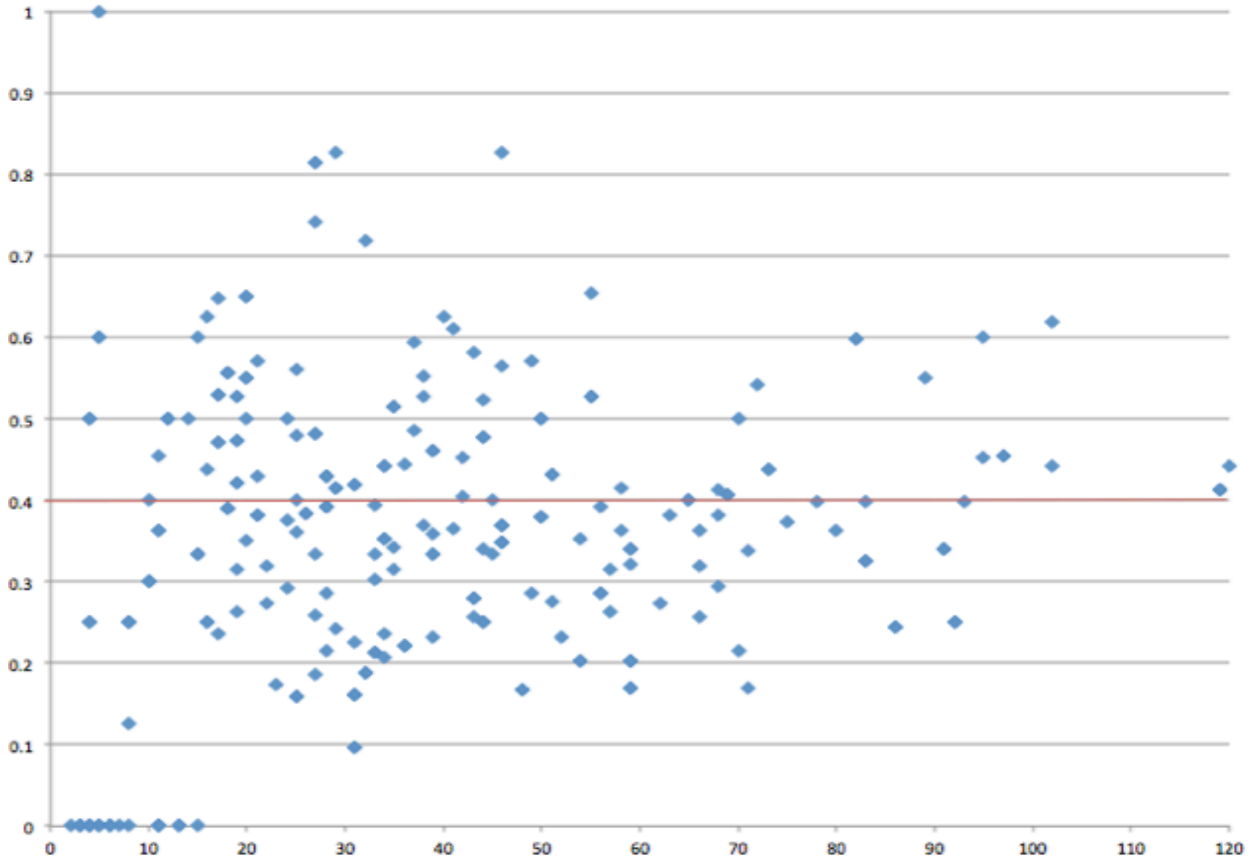
Productivity threshold



Case 3: Customization

After customization – 25% gain in productivity

Productivity threshold



Take home messages…

- Do you have technology that can solve a problem?
  - *Validate this!*

- Is there a market for this technology?
  - *Find product/market fit!*

- *Go for it!*

- This is what we did in developing domain-adapted MT solutions with subject matter expertise for LSPs and data providers.

- *Enjoy the ride!*

Thank You!

john@iptranslator.com

@IconicTrans

Iconic
Translation Machines