“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?
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

Is this worth exploring?

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

Skillset

- License IP
- Spin out

- Support
- Exporting
- Investment

Feasibility Study Grant

Commercialisation Fund

iHPSU
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

- Input
- Pre-processing
- MT ENGINE
- Post-processing
- Output

Training Data
The Challenge of Patents

Long Sentences

Longest Sentence: 1,417 words

Largest single document: 249,322 words

maximum stress of 1.2 to 3.5 N/mm² and a maximum elongation of 700 to 1,300% at 0[deg.] C.
What is Linguistic Engineering?

Data Engineering

Training Data

Input

Pre-processing

MT ENGINE

Post-processing

Output
An “ensemble” architecture

Data Engineering + Linguistic Engineering

Input

Training Data

Patent input classifier

Spanish med-device entity recognizer

Multi-output Combination

German Compounding rules

Japanese script normalisation

Moses

Output

Korean pharma tokenizer

Chinese pre-ordering rules

RBMT

Moses

Moses

Statistical Post-editing

Client TM/terminology (optional)
What is the value for users?

De-risking the machine translation proposition

Typical Prerequisites

+ Data
+ Time
+ €€€
= ???

Our Prerequisites

+ No data needed
+ 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

- Iconic
- Google
- Systran

Categories:
- Human Necessities
- Operations
- Chemistry
- Textiles
- Fixed Constructions
- Mechanical Engineering
- Physics
- Electricity
Case 1: Quality

German to English

Evaluator 1: 2.83
Evaluator 2: 4.0
Evaluator 3: 3.86
Average: 3.56
Case 2: Productivity

Business Need

Machine Translation technology for the legal industry

Iconic had a domain-specific MT solution for that industry
Case 2: Productivity

Process

Translation samples required for initial evaluation

Delivered immediately and initial results were positive
Case 2: Productivity

Performance

"MT delivered measurable productivity gains from the outset"

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

![Bar chart showing Iconic and Google performance with Baseline comparison]

- Iconic: 21
- Google: 20

Beregovaya et al. (Eds.) Proceedings of AMTA 2014, vol. 2: MT Users Vancouver, BC © The Authors
Case 3: Customization

Essentially out of domain – not viable for post-editing

Productivity threshold
Case 3: Customization

After customization – 25% gain in productivity
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