March 2018

Neural Won:

Now What?

Alex Yanishevsky

Senior Manager

MT and NLP Deployments

welocalize
doing things differently

Our History

1997. 2000. 2005. 2006. 2007. 2008.

Company Started

Expanded Market into Germany (Acquisition)

Established European Headquarters in Dublin (Acquisition)

Opened APAC Services in Jinan and Beijing, China (Acquisitions)

US Acquisitions Including Global Sight Technology, Leading Open Source Training Management System

Added to APAC Presence with Tokyo Office (Acquisition)

2010. 2012. 2014. 2015. 2016.

Major Expansion in Europe with UK Acquisition of Lloyd International

Added Legal Services with Market Leader Park IP Translations (Acquisition)

Acquired CD Language Solutions in Houston, TX

Acquired Agostini Associati in Milan, Italy

Significant Investment from Norwest Equity Partners

Acquired Adapt Worldwide (Traffic Optimiser) in London, United Kingdom

11th Year on Inc. 5000 Fastest Growing Private Companies

Nova Language Services Acquisition Expands Regulated Industry Solutions in Life Sciences (Acquisition)

Global Language Solutions (GLS) Acquisition Strengthens Life Sciences Market Leadership (Acquisition)

welocalize
doing things differently

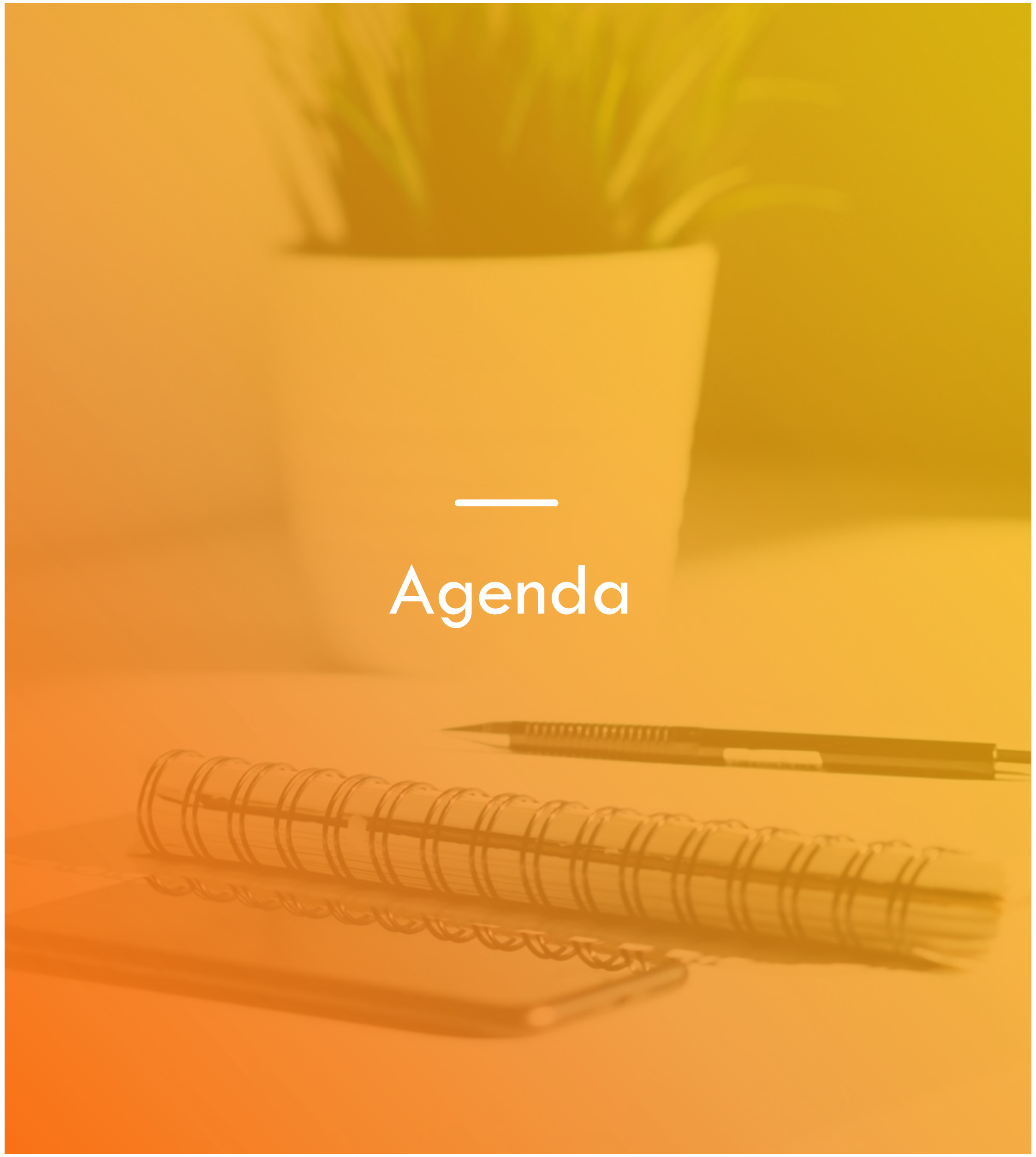
The **Facts**

- **175+**
  Languages
- **1500+**
  Employees
- **20+**
  Global Offices
- **72,859**
  Projects 2016
- **1,946**
  Global Clients
- **1.16 Billion**
  Words Translated 2016
- **4th Largest**
  Language Services Provider in the US
- **7th Largest**
  Worldwide*

Welocalize is one of the largest language service providers in the world.

*Source: Common Sense Advisory, 2015

welocalize
doing things differently



Agenda

- **Did NMT really win?**
- **Migration path**
  - Build or buy?
  - Infrastructure and Cost
  - TMS and Connectors
  - Additional Use Cases – CMS, applications using MT such as chat, KB, forums
  - Training and Maintenance
  - Supply Chain
- **Case studies**
- **What else can we do with neural technology?**

Did NMT Really Win?

Did NMT Really Win? /1

Generally, yes, and the future lies in NMT, but…

- **Locale variants such as ES-ES>ES-MX: consider transformation tables or Apertium (RBMT)**
- **Related language pairs such as ES-ES>PT-PT: consider Apertium (RBMT) or SMT**
- **Rare, long-tail language translation directions: consider SMT**
- **In some cases, well trained SMT engine in Romance languages can be preferred to NMT**
- **In some cases, SMT better at short sentences**

welocalize
doing things differently

Did NMT Really Win? /2

Locale	Evaluation	Light Marketing				Technical Documentation			
		Generic NMT1	Generic NMT2	Customized SMT	Diff Best NMT & SMT	Generic NMT1	Generic NMT2	Customized SMT	Diff Best NMT & SMT
de-DE	Ranking	√	2	3	6.02 pp	2	√	3	7.38 pp
	Accuracy			√	0.06		√		0.08
	Fluency		√		0.07		√		0.45
	Edit Distance	2	3	√	3.32 pp	√	3	2	1.12 pp
	Edit Distance (PE)	2		√	1.55 pp				
fr-FR	Ranking	√	3	2	1.97 pp	√	2	3	7.29 pp
	Edit Distance	2	3	√	2.02 pp	2	3	√	0.62 pp
ja-JP	Ranking	√	2	3	12.96 pp	√	2	3	10.51 pp
	Accuracy		√		0.32		√		0.76
	Fluency		√		0.2		√		0.49
	Edit Distance	√	3	2	8.17 pp	√	3	2	5.79 pp
	Edit Distance (PE)	√		2	21.07 pp				
pt-BR	Ranking	√	3	2	4.59 pp	√	2	3	6.65 pp
	Accuracy		√		0.09		√		0.26
	Fluency		√		0.45		√		0.28
	Edit Distance	2	3	√	1.68 pp	√	3	2	0.28 pp
	Edit Distance (PE)	2		√	3.62 pp				
zh-CN	Ranking	√	2	3	10.57 pp	√	2	3	10.40 pp
	Edit Distance	√	3	2	5.87 pp	√	3	2	3.12 pp
ru-RU	Ranking	√	2	3	5.95 pp				
	Edit Distance	2	3	√	1.58 pp				

- SMT better for DE for accuracy and edit distance
- SMT better for PTBR for edit distance
- SMT better for RU for edit distance

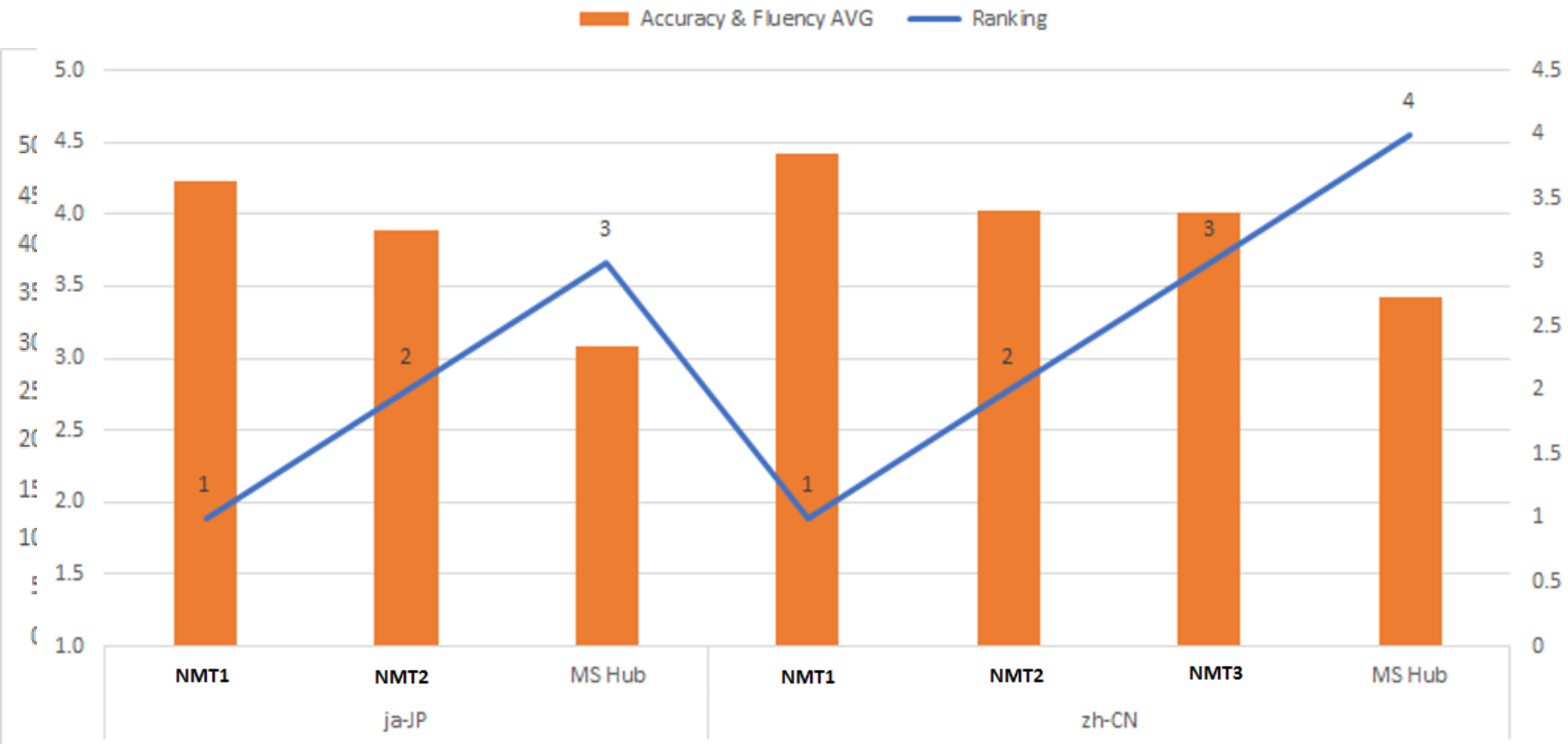
Did NMT Really Win? /3

Engine	Content	BLEU	NIST	METEOR	GTM	PE Dist	TER	Precision	Recall	Length (Hyp./Ref.)	Segs.	Words	PE Diff	Ranking 1	Ranking 2
NMT	Test set	64.10	10.33	73.72	80.11	37.05%	30.93	0.82	0.79	0.96	2460	33322	17.11%		
MS Hub	Test set	60.22	9.75	71.64	78.96	54.16%	36.61	0.79	0.79	0.99	2500	33863			
NMT	Aug-projects	63.07	8.00	73.28	77.57	50.44%	38.84	0.76	0.79	1.03	513	3852	-5.77%		
MS Hub	Aug-projects	72.38	9.34	81.98	86.89	44.67%	23.41	0.88	0.86	0.97	559	4201			
NMT	Oct-projects	54.90	7.91	66.81	72.63	59.63%	45.65	0.71	0.74	1.04	940	7265	-8.85%	43%	37%
MS Hub	Oct-projects	60.96	8.84	72.95	79.79	50.78%	34.49	0.80	0.80	1.00	1057	8395		33%	26%

- The NMT engines scores better in human ranking
- NMT engine has a lot of omissions, duplications and unusual mistranslations
- Results for auto-scoring are mixed

welocalize
doing things differently

Did NMT Really Win? /4



- Hub in last place in all rankings
- NMT1 in first place in all rankings
- NMT technology better overall

welocalize
doing things differently

Now What?

PAIN POINTS

Raw MT, PE, both

NUMBER OF ENGINES

How many domains and engines do you have and for how many languages?

STRATEGY

What is your migration path and strategy?

welocalize
doing things differently

Migration Path

welocalize
doing things differently

Migration Path

- BUILD OR BUY?
- TMS AND CONNECTORS
- ADDITIONAL USE CASES: CMS, CHAT, KB, FORUMS
- TRAINING AND MAINTENANCE
- SUPPLY CHAIN
- CASE STUDIES

welocalize
doing things differently

Build or Buy /1

BUILD

- Customized needs
- Internal expertise
- Flexibility

BUY

- Lack of time
- Lack of expertise
- Lack of customizability

BUILD

- Competitive advantage
- Build from scratch or adapt open source solutions

BUY

- Lack of influence over product roadmap
- Reliable tech support
- Reliable solutions available out of the box

welocalize doing things differently

Build or Buy /2

BUILD

- Modern MT
- Open NMT
- Tensor Flow
- Nematus
- Marian
- Moses

BUY

- Google, Amazon, Bing – not customizable
- MS Hub SMT, Globalese, Kantan, Omniscien, SDL, Systran, Iconic, etc. - customizable

BUILD

- Limited baseline
- Difficult to enforce terminology

BUY

- Robust or limited baseline based on provider
- Generally difficult to enforce terminology, but based on provider

welocalize
doing things differently

Build or Buy /3

BUILD

- More options to control (epochs, layers, baseline vs domain data)
- Quality of documentation and code samples may be more uneven

BUY

- Less options to control
- Very good documentation and code samples

BUILD

- Unlimited usage
- $3/hr for cloud for processing MT requests
- 2K per engine for training per month for 20 epochs at 4 hours an epoch

BUY

- $10-20 for 1 million characters – not customizable and MS Hub SMT
- Several hundred to several thousand per engine – customizable

welocalize
doing things differently

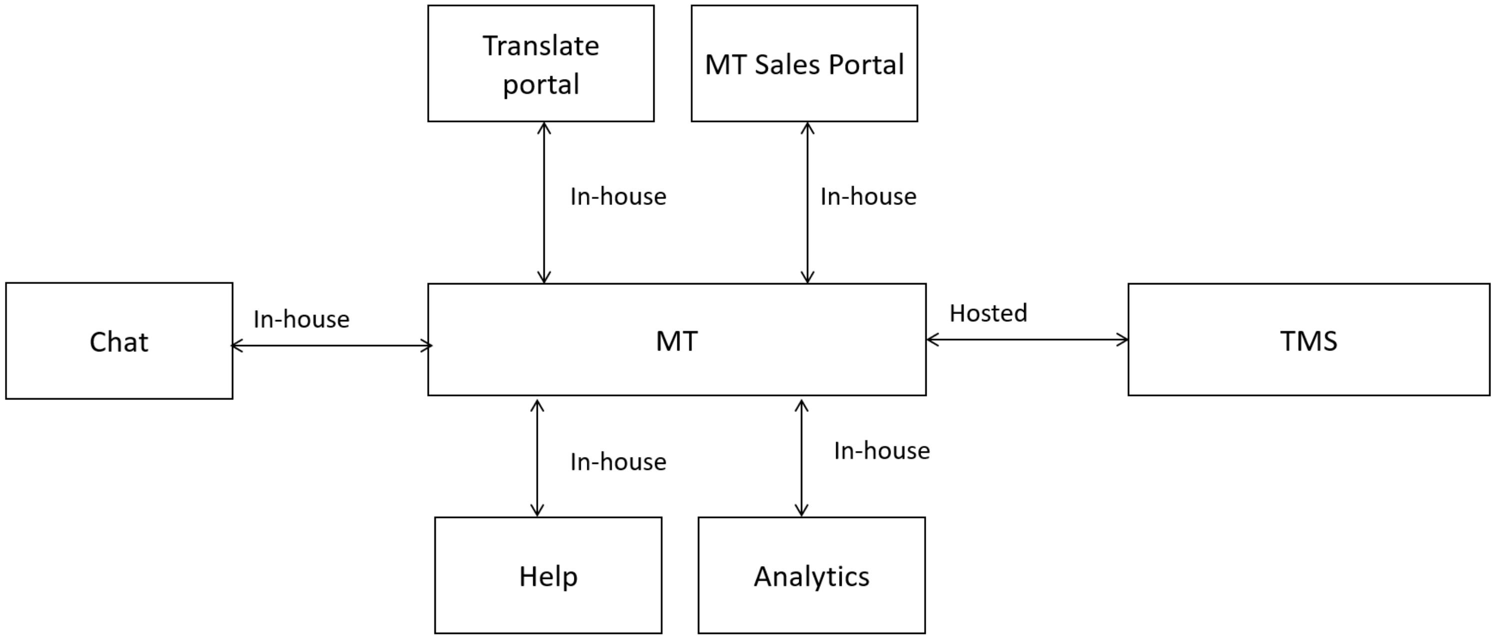
TMS and CAT Tool Considerations

- Availability and additional cost of connectors depends on TMS or CAT tool
- Tag handling
- Pre and post processing scripts
- Tags as sentence breakers
- Capabilities for providing feedback
- Interacting with Adaptive MT

Ideally, the TMS has several MT connectors so you can pick and choose and migrate when results are conclusive and/or run several MT providers in parallel.

welocalize
doing things differently

Additional Use Cases of MT



Rewrite connectors for
- KB
- Forums
- Chat
- Any other applications

6 points of MT integration!

welocalize
doing things differently

Training and Maintenance /1

Initial training
Computational costs of building NMT vs SMT are higher

Maintenance
Computational cost of enforcing specific patterns from linguistic feedback is higher; it’s not a matter of modifying phrase tables or language models as with SMT or rules/dictionaries with RBMT.

welocalize
doing things differently

Training and Maintenance /2

- Data availability
  - ✓ Some NMT systems with restricted options require a lot more training data than comparable SMT or RBMT systems
  - ❑ 5-7 million TUs (sometimes 10-11 million) overall to match the quality of an SMT engine in MS Hub with 500-600K TUs and MS Models
  - ❑ Client data ranged from 50K to 700K TUs
  - ✓ Possible to train decent engine with 1-2 million TUs in a different framework with more options available

welocalize
doing things differently

Training and Maintenance /3

Data Quality

Bad for both

- Uneven or misaligned TUs
- Wrong target language
- Poor, unreliable or inconsistent translations
- Really long segments (NMT – attention mechanism keeps track for only so long due to vanishing gradients, SMT – can't focus on long term dependencies, e.g. English with relative clauses)

Bad for NMT only

- Short segments (1-3 words)
- High ratios of DNT if you do not have method to enforce dictionary

welocalize
doing things differently

Training the Supply Chain

- NMT output is remarkably more fluent.
- However, **this fluency does not guarantee accuracy**. The cognitive load can be higher for a post-editor to review the source and suggested target.
- OOVs and DNT mistranslations

59805384	Arianna Fontana ha vinto al fotofinish la medaglia d'Oro nella finale dei 500m di short track ai Giochi olimpici invernali di PyeongChang.	59805384	Arianna Fontana won the medal's medal on the end of the short track of short track listeners at the Olympic Olympic Games of the FIFA Winter.

Source	Hypothosis	Reference
6 Div(Low)	6、 、	6 分割(低)

Source	Hypothesis	Reference
a04JwiqW9El4hce/Z3+nOHOckWJ0VSCFoqox1FVpYW4fXSeHfuQ0ktVnylyMz/vYTAWrnj493YlY	X/Z3+Delete/bbr	a04JwiqW9El4hce/Z3+nOHOckWJ0VSCFoqox1FVpYW4fXSeHfuQ0ktVnylyMz/vYTAWrnj493YlY
Examples: file:///remote/file/system/mount/point, \\\\server\\path or nfs://server:/path	示例：、或更高版本	示例: file:///remote/file/system/mount/point、\\\\server\\path 或 nfs://server:/path

Source	Hypothesis	Reference
<proxyAddress:port>	<>	<proxyAddress:port>
GuestRpc:	:	GuestRpc:

welocalize doing things differently

Case Studies

Case Study 1: Internal Dept, MTPE

- MemoQ as CAT tool
  Numerous MT connectors
- Currently on MS Hub SMT
  EN<>FR, IT, DE, ES, PTBR
- One domain, life sciences
  Any SMT or NMT solution must be customizable
- OpenNMT adaptation shows markedly improved scores

welocalize
doing things differently

Case Study 1 ROI Calculations

ITEM	COST	SAVINGS
Connector	0	
MT Usage	0	
Engine Cost ($1000 per locale pair per year)	$8000	
Vendor discounts (100K new words per year* 8 locale pairs* .01 per word)		$8000

- By how much does NMT need to win in order to move now?
- How can we put a price on this?
- How much volume? What languages?
  NOTE: How likely is additional .01 per word for each locale pair? What additional discount does it represent? For rate of .15, that's an additional 7%.

welocalize
doing things differently

Case Study 2, Tech Support, Raw MT

- All possible language combinations, i.e. over 50 languages
- UGC – prone to slang, typos, incorrect formatting
- MT embedded into chat application
- How important is lexical coverage?
- How many MT connectors does the chat application support?
- If several, can you mix and match?
- If you deploy several, what's the administrative overhead of licensing and retraining from several different MT providers?
- Is normalization taking place?

What is the minimum allowable level of quality for the lowest cost?

welocalize
doing things differently

Case Study 3, Enterprise, MTPE

- Enterprise level TMS
- Currently on MS Hub
- Each MT connector costs money and has to be vetted by TMS provider
- Many (but not all) languages do better in NMT

- What business problem are we solving – TAT, quality, cost of delivery? How will the move to NMT be a game changer?
- Split the languages amongst the connectors or only move when you can do all of them?
- As in case study 1, what's the cost of each connector relative to the expected volume, increased quality and expected discount by moving to neural?

welocalize
doing things differently

What Else Can We Do with Neural Technology?

welocalize
doing things differently

What Else Can We Do?

- NLG (Natural Language Generation) with subsequent NMT
- Sentiment analysis
- Predictive analytics for localization program management and linguist selection
- Predictive input
- Virtual assistants
- Machine learning for LQA and evaluation of source
- Document summarization

welocalize
doing things differently


Thank you
welocalize
doing things differently