The Emerging World of Arabic Localization & Machine Translation

15 April 2021

Nour Hussein
Chief Growth Officer

tarjama
WORDSMITHS

About Nour Hussein

Organisational Growth & Transformation
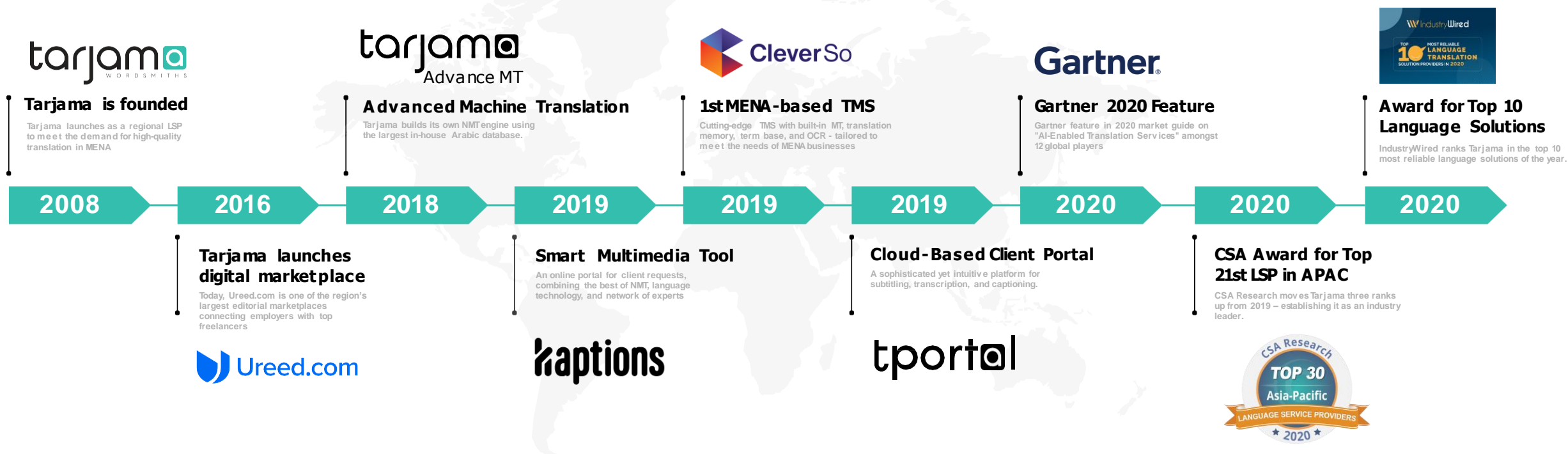
Commercial Product Development

Digital Solutions Design

Europe to MENA Connection

Egypt, UK, India, Netherlands, UAE

Poll Question 1

Which industry are you working in?

Poll Question 2

Do you have Arabic localization in your company's growth strategy?

Agenda

1. Importance of Arabic for businesses

2. The growing need for Arabic localization

3. Challenges of Arabic localization

4. Where has Arabic MT technology come so far?

5. How to overcome challenges with MT

6. The future of Arabic MT

tarjama
WORDSMITHS

Importance of Arabic for businesses & opportunities in the Arabic spoken countries

20
Countries

420
Million People

Arabic is spoken in more than 20 countries in the Middle East and North Africa by 420 million people.

5th
Place

Based on factors such as speaker numbers, economic output and importance in world diplomacy, the Power Language Index (PLI) has Arabic in 5th place.

2744.1 $
Billion USD

Arab world has a combined GDP Value of 2744.1 Billion USD

Source: Joint Arab Economic Report 2020

The growing need for Arabic localization

ENGLISH

English native speakers

4.9%

of the world's population

English content

60%

English dominates the majority of content on the web

ARABIC

Arabic native speakers

5.4%

of the world's population

Arabic content

3%

but estimates suggest that Arabic content makes 3% percent online.

The growing need for Arabic localization

tarjama
WORDSMITHS

0.9%

Arabic content on Wikipedia

WIKIPEDIA
The Free Encyclopedia

Top 12 languages of global websites

Rank	Language	Percentage
1	English	60.6 %
2	Russian	8.3%
3	Turkish	3.9%
4	Spanish	3.8%
5	Persian	3.3%
6	French	2.7%
7	German	2.3%
8	Japanese	2.1%
9	Vietnamese	1.7%
10	Chinese	1.4%
11	Arabic	1.1%
12	Portugese	0.9%

Challenges of Arabic localization

Linguistic challenges

There are many different forms of the Arabic language

1. Qur'anic or classical Arabic

2. Modern Standard Arabic

3. Colloquial or dialectal Arabic – over 30 different dialects across the MENA region

Challenges of Arabic localization

Linguistic challenges

Complex morphology (structure of language, words, and sentences)

<div dir="rtl" align="center">

فَأَسْقَيْنَاكُمُوهُ

</div>

<div align="center">

faˈasqaynākumūhu

</div>

And given you drink from it.

Poll Question 3

How do you use Arabic?

Challenges of Arabic localization

Linguistic challenges

Diacritics make Arabic words more complex & layered

"أَلَمُّ أَلَمَّ أَلَمُ أَلِمَّ أُلُّمَّ بِدَائِهِ – إِنْ آنَ آنُ آنَ آنُ أُوَانِهِ"

Challenges of Arabic localization

Cultural challenges

Heavy use of Arabic metaphors that varies based on region + wouldn't make sense in another language .

(Kol Sana Wa Enta Salem) كل سنة ونت سالم

Literal Translation: Every year you are healthy

Transcreation: Happy birthday

(Ala Rasi) راسي على

Literal Translation: On my head

Transcreation: I'd do anything for you

Poll Question 4

How often you use MT?

Challenges of Arabic localization

Cultural challenges

Arabic words that have no equivalent in English.

(Ya'aburnee) يقبرني

A declaration of one's hope that they'll die before another person because of how difficult it would be to live without them

يقين *(Yaqin)*

Certitude and freedom from doubt

Where has Arabic MT technology come so far?

Problems with generic MT models when it comes to Arabic:

- They are **crowdsourced**

- They are **non-domain-based**

- They don't consider **dialects**

- They pose **security** threats

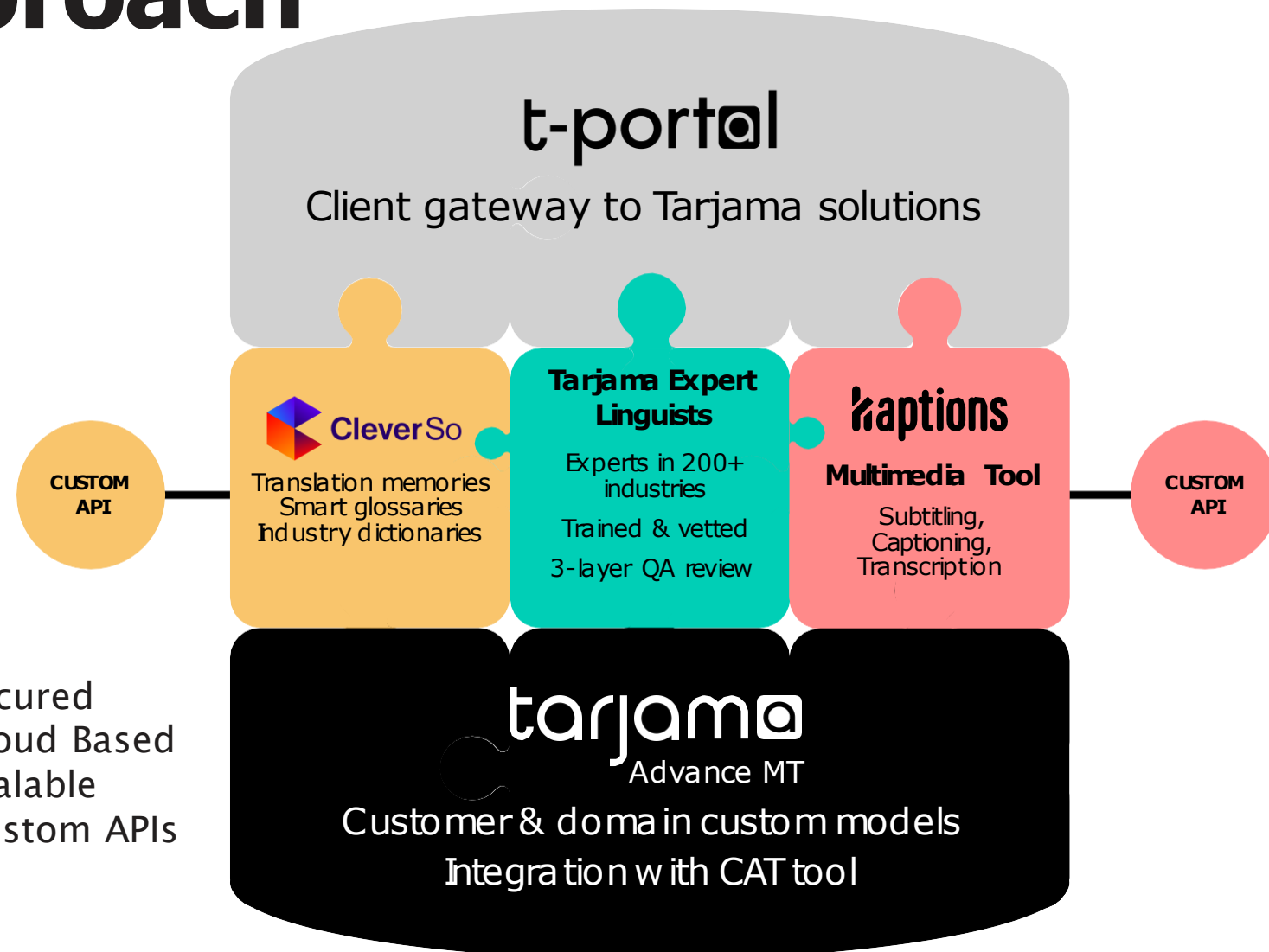# Overcoming challenges with MT

Language is subjective.

There is a difference between a **correct translation** and a **good translation**/**transcreation**.

The Augmented Intelligence Approach

At Tarjama, we advise customers to opt for an Augmented Intelligence Solution where the Machine does the heavy lifting to save time and the human does the creative and qualitative adjustments.

✓ Secured
✓ Cloud Based
✓ Scalable
✓ Custom APIs

t-portal
Client gateway to Tarjama solutions

CleverSo
Translation memories
Smart glossaries
Industry dictionaries

Tarjama Expert Linguists
Experts in 200+ industries
Trained & vetted
3-layer QA review

kaptions
Multimedia Tool
Subtitling, Captioning, Transcription

CUSTOM API

CUSTOM API

tarjama
Advance MT
Customer & domain custom models
Integration with CAT tool

Tarjama MT

Tarjama MT meets the need for high-quality Arabic machine translations that are trained by professional pre-approved data that is domain-specific and accounts for dialects.

Pre-approved Data   **Domain Specific**   **Dialect**

We tested the accuracy of our MT solution from English to Arabic

Results showed that Tarjama's MT engine had significantly higher scores in the legal and finance industry sectors and the second highest in the health and business industry sectors

	tarjama WORDSMITHS	Google	aws	SYSTRAN beyond language	Microsoft Bing
Business	63.16%	63.16%	65.78%	55.26%	52.63%
Legal	75%	35%	50%	20%	35%
Finance	82.5%	67.5%	75%	70%	70%
Health	40%	60%	40%	30%	40%

The future of Arabic MT

Arabic not as one language – but as several distinct dialects

Domain Specific MT as a Norm – organisations and government training MT engines

Data Annotation & Labelling – Evolution of a linguist role?
Customer feedback tagging/classification, Sentiment analysis

Development of NLP tools geared for Arabic
Text extraction, text clustering, sentiment analysis, chatbots, speech-to-text
Email filters, Virtual Assistants, chatbots, Automatic Summarisatiom, Predictive text

Integration with other CAT tools & elements
Integrating with TMS for example helps extract more metadata to develop the MT engine

Questions?

Abu Dhabi
twofour54
Abu Dhabi, UAE
T +971 2 401 2915

Dubai
Dubai Design District
Dubai, UAE
T +971 4 514 6421

Riyadh
Uthman Ibn Affan St.
Riyadh, KSA
T +966 1145 44509

Amman
Al Madina Al Monawara
St. Amman, Jordan
T +9627 9867 8182

✉

nour.hussein@tarjama.com

www.tarjama.com

info@tarjama.com

tarjama