Rule-based MT and UTX Glossary Management – Honda's Case Dealing with Thousands of Technical Terms

MT Summit 2017 (Nagoya, Japan)

Saemi Hirayama
CAT tool leader, Honda R&D Americas, Inc.
Yuji Yamamoto
Founder/representative, CosmosHouse

Contents

Speakers

Saemi Hirayama

An in-house translator/CAT tool leader at the Ohio Center of Honda R&D Americas, Inc.

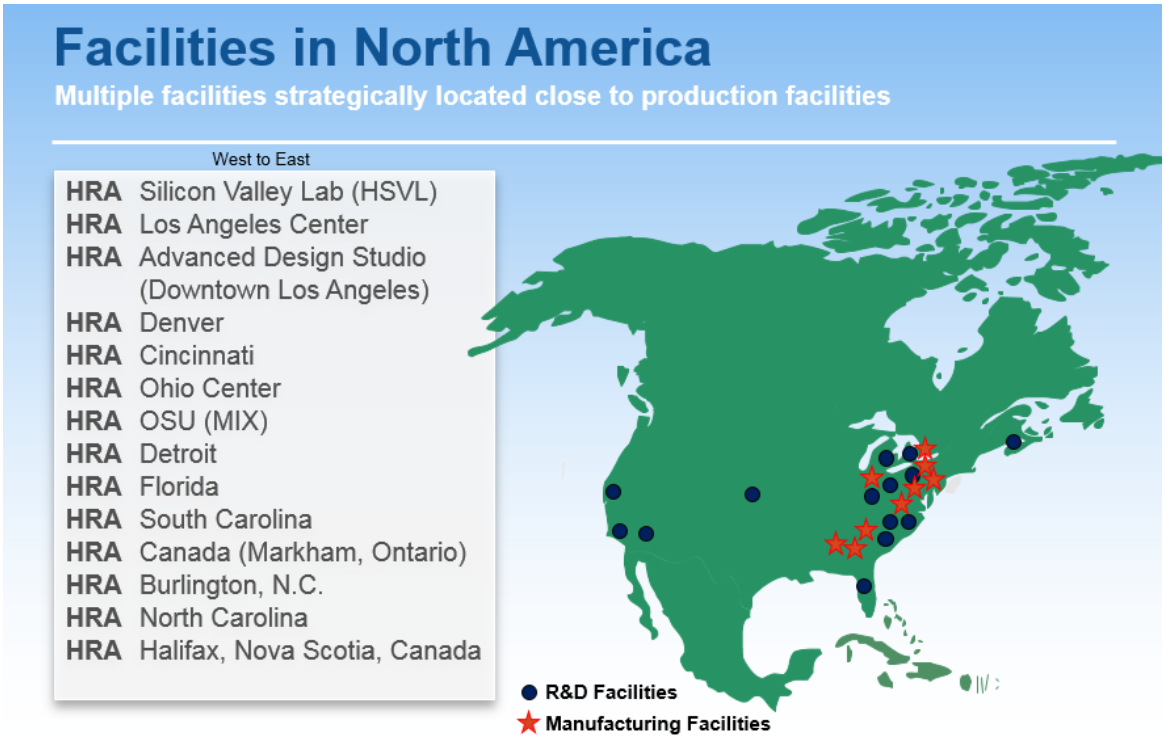
■Yuji Yamamoto

Founder/representative, CosmosHouse <http://cosmoshouse.com>

UTX team leader at AAMT (Asia-Pacific Association for Machine Translation)

Honda R&D Americas Inc.

(hereafter referred to as **Honda R&D**)



- Product Research & Development
- Product Styling Design
- Environmental Technology Development
- Safety Technology Research and Development

Honda R&D's needs

1. JA to EN, EN to JA
2. Technical documents written by engineers
3. Used for translation needs by global associates in daily business operations
4. Term-level accuracy and consistency are important
5. Speed is crucial

Honda R&D MT overview

Over a decade ago, Honda R&D adopted an RBMT (Rule-based Machine Translation) system

The current MT is a RBMT subsystem

Engineers use it to translate documents and emails

In-house translators also use it to process translation requests from engineers

Honda R&D MT overview

- Honda Jargon dictionaries categorized and added to the MT
- A feedback function added to the web-based MT for mistranslations/ unregistered terms to keep the dictionaries up-to-date

Honda R&D MT achievement

- In-house translations reduced and outsourcing cost cut by half
- Significant translation speed increase
- Better communication with accurate technical terms

Why was RBMT chosen at Honda R&D?

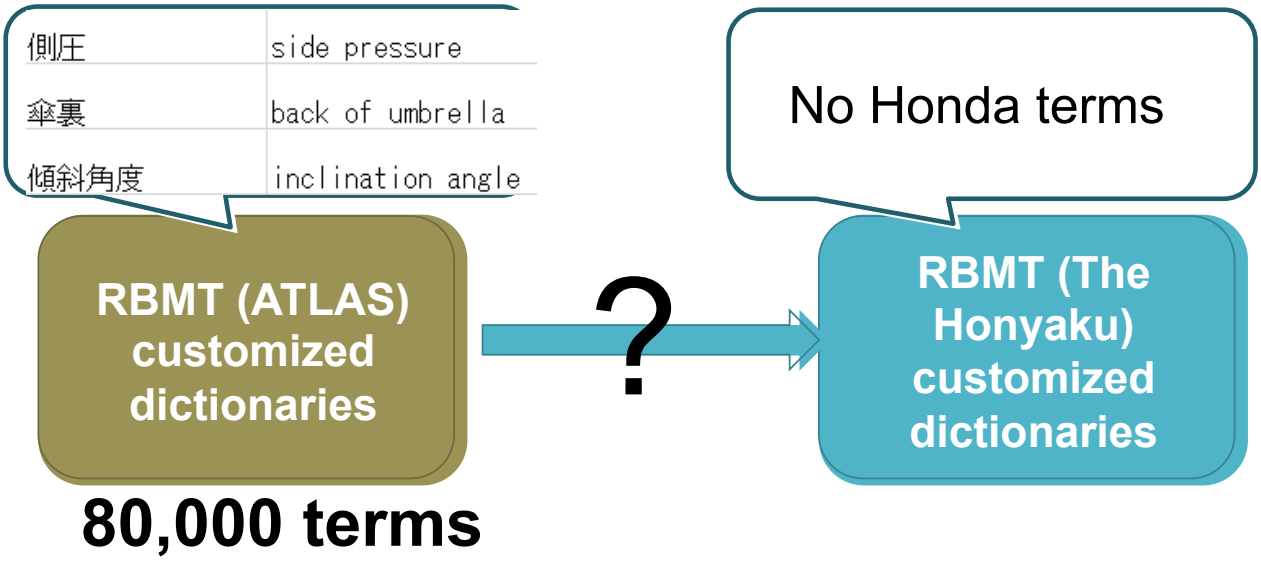
1. 80,000 terms, including many Honda-only terms
2. Many incomplete fragmental phrases/very few fixed phrases
3. File formats: complicated slides
4. No two documents are alike

Why is neural/statistical MT not used at Honda R&D?

1. Term-level accuracy and consistency are poor in NMT
2. Human-translated corpus is too small
   Because the majority of translations are lightly post-edited machine translations
3. Protection of intellectual property and secrecy
4. Higher cost
5. Most documents do not repeat

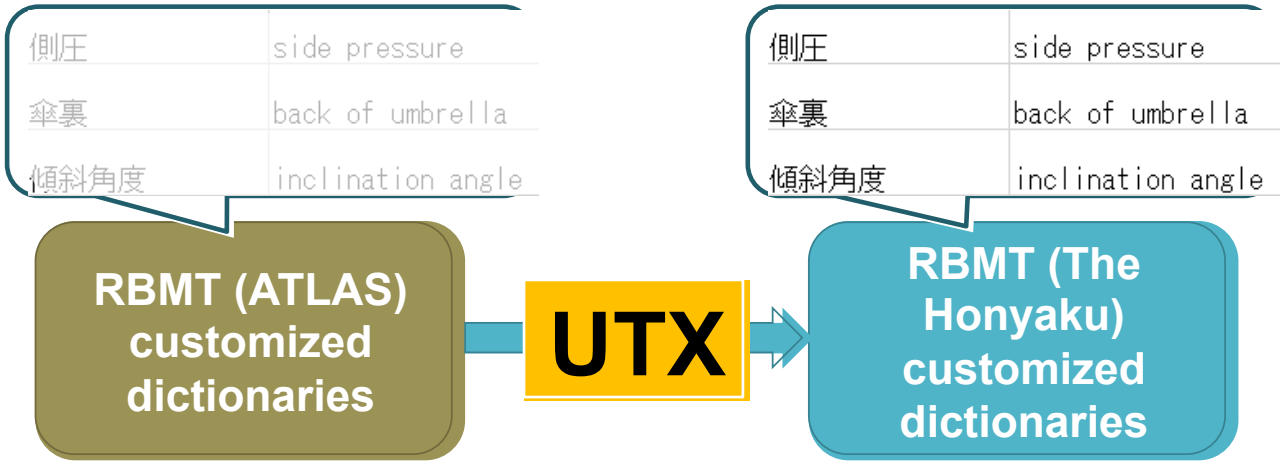
Issue 1: MT migration (RBMT to RBMT)

80,000 Honda terms in Fujitsu ATLAS were needed to be imported into a new MT system, Toshiba's The Honyaku.



Solution 1: MT migration (RBMT to RBMT)

Solution: Conversion through UTX format. The customized dictionaries were transferred to the new MT.



80,000 terms

UTX – glossary standard for sharing and reuse

UTX (Universal Terminology eXchange)

- **Developed by AAMT, initially as RBMT dictionary data exchange format**
- **Later restructured as a structured glossary format**
- **Used by companies and organizations such as Japan Patent Office**

http://www.aamt.info/english/utx/

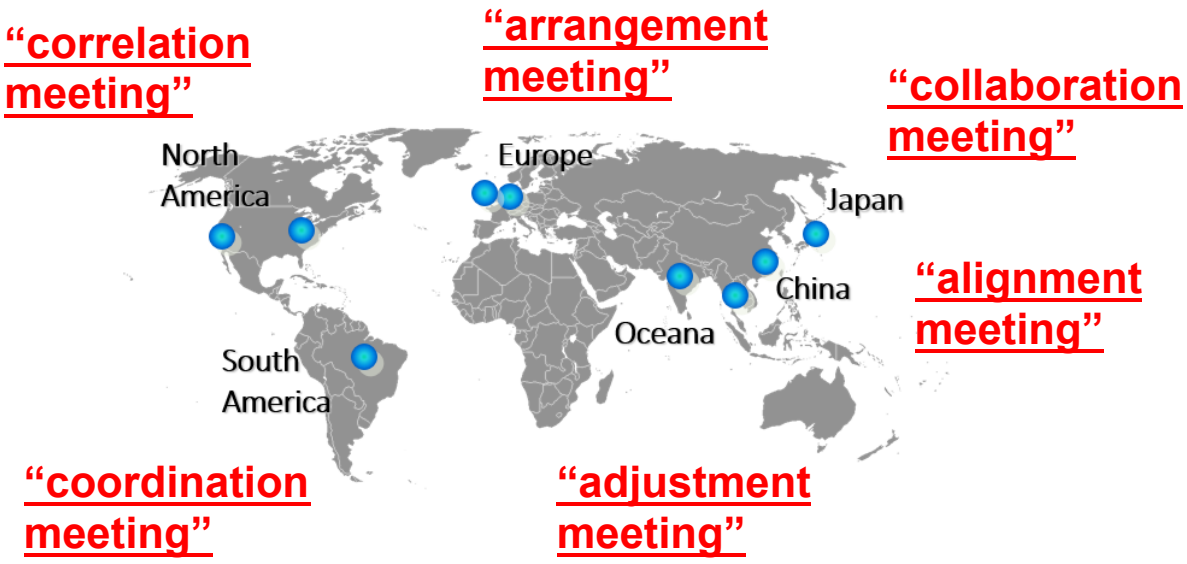
Issue 2: term inconsistency

Identical Honda jargon was being translated inconsistently at various company sites around the world.

整合会

Correlation meeting?
Coordination meeting?
.....?

“整合会”



Honda Terms were being translated inconsistently in 6 Regions Worldwide

Solution 2: term inconsistency

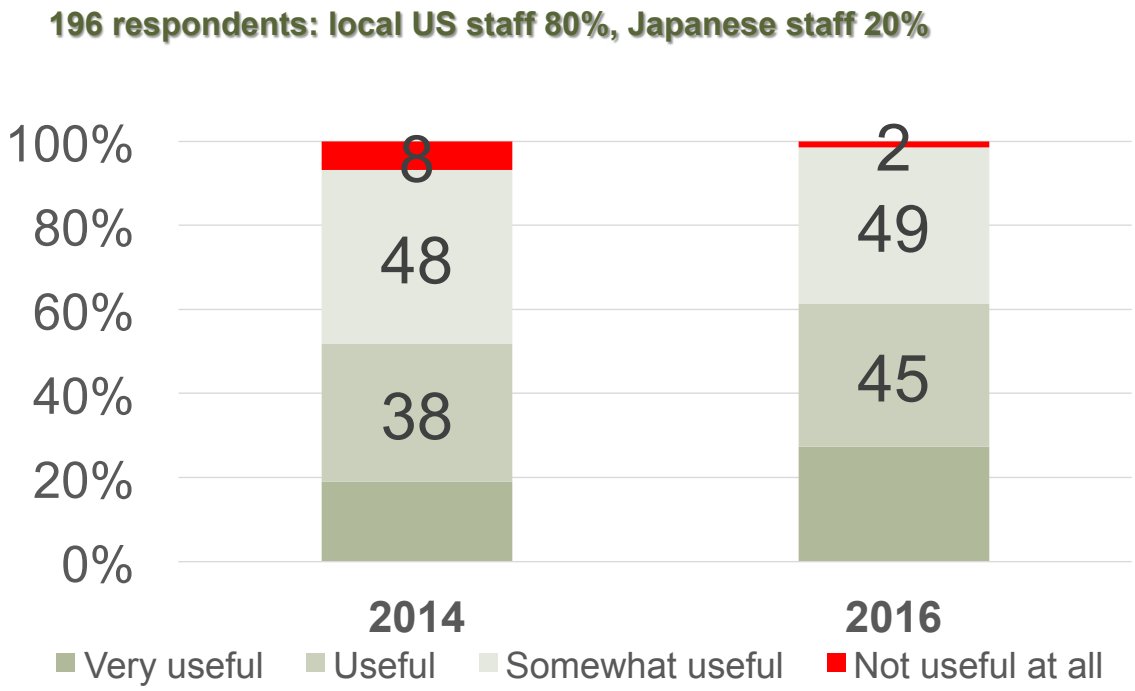
- Term statuses (approved, non-standard, forbidden etc.) were added.
- 1:n, n:1, n:n source/target term pair relationships are clearly defined.
- J to E glossary now also works as E to J.

#term:ja	term:en	term status:ja	term status:en
整合会	correlation meeting	approved	approved
整合会	coordination meeting		forbidden
整合会	collaboration meeting		non-standard

Terminology management at Honda R&D

- 2013: UTX was introduced, transferring 80,000 terms from the old MT to a new one.
- 2014: a terminology committee was established to review existing/new terms to update the MT dictionaries monthly.

Reported useful by 98% of users



Terminology management continues

1. Review terms and term statuses
2. Add new terms
3. Delete unnecessary/obsolete terms
4. Categorize terms

…to improve translation accuracy and efficiency

For fellow MT user companies

- Glossaries control your company vocabulary

  Quality of human/machine translation can be improved with terminology management

- Proper terminology management pays off!

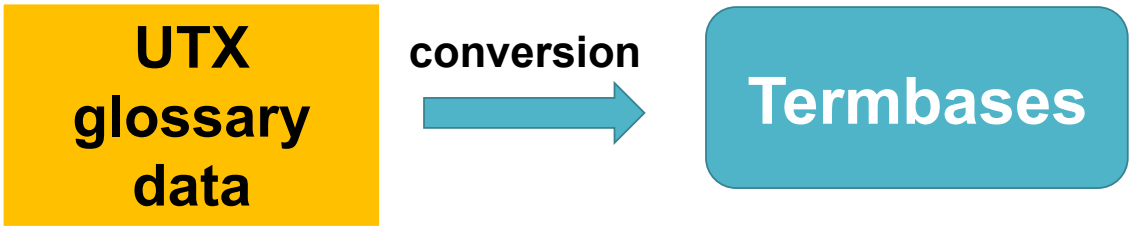
Solution 2: term inconsistency

- Term statuses (approved, non-standard, forbidden etc.) were added.
- 1:n, n:1, n:n source/target term pair relationships are clearly defined.
- J to E glossary now also works as E to J.

#term:ja	term:en	term status:ja	term status:en
整合会	correlation meeting	approved	approved
整合会	coordination meeting		forbidden
整合会	collaboration meeting		non-standard

Future actions at Honda R&D

1. Further tuning of UTX glossary
2. UTX for terminology check
   - Can be used for post editing neural/statistical MT if necessary
   - Terminology tool training for translators



Take away

1. Glossaries are necessary for both humans and MT
2. UTX glossary management has been effective at Honda R&D
3. Neural MT may not be the only future – users are satisfied with RBMT