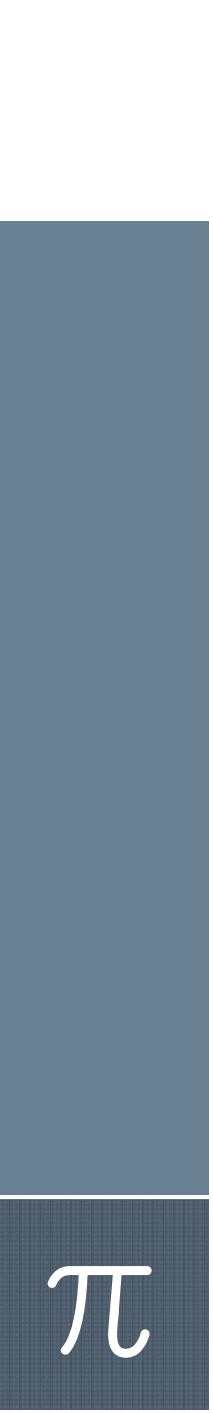
Associating Structured Translation Specifications with Points on a Spectrum from MT to HT

T&I LEARN workshop,
May 16-18, 2016,
Loyola Columbia Graduate Center



The new age of machine translation (MT)

- Why are we in a ***new*** age of machine translation? It has been around for over 50 years. It is approaching old age.

- What is new is that more people use machine translation in various ways, and there is an increased optimism about artificial intelligence in general and its application to machine translation in particular.

- On what basis should someone decide whether to use machine translation, human translation, or some mix?

On the basis of the answers to 3 questions:

(1) Have we reached the Singularity?
(2) What are the Structured Translation Specifications for the project at hand?
(3) Which point on the Spectrum is most appropriate for this project?

Note: we have not yet reached the Singularity. When we do, everything changes for everyone, so I will delay further discussion of it until later on in this session.

A VISION OF THE FUTURE

Human Translation (trees) and Machine Translation (pagoda)

Living in harmony

image: Xianhu Botanical Garden Shenzhen, China

Presenter: Alan K. Melby

- Emeritus Professor of Linguistics, Brigham Young University
- Member of the governing council of FIT (International Federation of Translators)
- President of LTAC Global (a small non-profit promoting the use of translation-related standards)
- Certified French-to-English translator
- First university degree: Mathematics (hence Pi symbol) but later switched to Linguistics

Preview of the rest of this presentation

- **Specifications** (how to develop them for each project)
- **Spectrum** (what are the five points on it?)
- **Singularity** (why some translation projects will be best assigned to human translators until the arrival of the Singularity)

Specifications

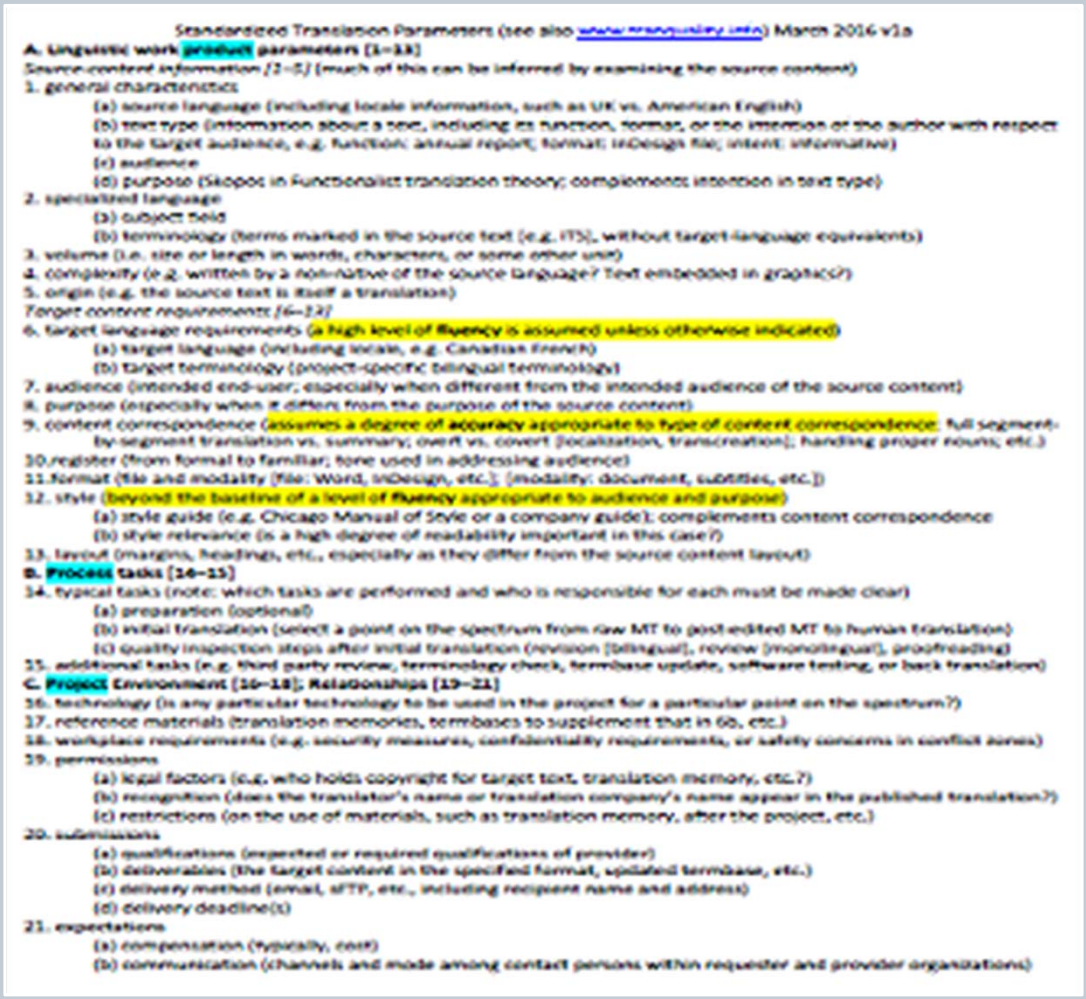
- There are three main types of **stakeholders** in a translation project:
  - Requesters
  - Providers
  - End Users
- Translators, who are **providers**, and translation project managers need to learn how to develop **structured translation specifications** collaboratively with requesters, so they are appropriate to end users, based on initial requirements, which are typically incomplete

Specifications are based on parameters

- Translation parameters are questions that need answers before a translation project begins
- There are 21 standardized parameters in ASTM F2575-14, structured hierarchically under the following high-level categories:
  - **Product**
    - Source-content characteristics
    - Target-content requirements
  - **Process** (tasks to be performed during the project)
  - **Project** Environment and Relationships among stakeholders

STANDARDIZED TRANSLATION PARAMETERS

- See handout
- Notice high-level organization into Product, Process, and Product
- These parameters are from ASTM International F2575, 2014 edition, Section 8 and are compatible with Annex B of ISO 17100 (May 2015)



Standardized Translation Parameters (see also www.multilingual.com) March 2016 v1a

A. Linguistic work product parameters [1–13]

Source content information [1–5] (much of this can be inferred by examining the source content)

1. *general characteristics*
   - (a) source language (including locale information, such as UK vs. American English)
   - (b) text type (information about a text, including its function, format, or the intention of the author with respect to the target audience, e.g. function: annual report; format: InDesign file; intent: informative)
   - (c) audience
   - (d) purpose (Skopos in Functionalist translation theory; complements intention in text type)
2. specialized language
   - (a) subject field
   - (b) terminology (terms marked in the source text [e.g. ITS], without target-language equivalents)
3. volume (i.e. size or length in words, characters, or some other unit)
4. complexity (e.g. written by a non-native of the source language? text embedded in graphics?)
5. origin (e.g. the source text is itself a translation)

Target content requirements [6–13]

6. target language requirements (a high level of **fluency** is assumed unless otherwise indicated)
   - (a) target language (including locale, e.g. Canadian French)
   - (b) target terminology (project-specific bilingual terminology)
7. audience (intended end-user, especially when different from the intended audience of the source content)
8. purpose (especially when it differs from the purpose of the source content)
9. content correspondence (assumes a degree of **accuracy** appropriate to *type of content correspondence*: full segment-by-segment translation vs. summary; overt vs. covert [localization, transcreation]; handling proper nouns; etc.)
10. register (from formal to familiar; tone used in addressing audience)
11. format (file and modality [file: Word, InDesign, etc.]; [modality: document, subtitles, etc.])
12. style (beyond the baseline of a level of **fluency** appropriate to audience and purpose)
    - (a) style guide (e.g. Chicago Manual of Style or a company guide); complements content correspondence
    - (b) style relevance (is a high degree of readability important in this case?)
13. layout (margins, headings, etc., especially as they differ from the source content layout)

B. Process tasks [14–15]

14. typical tasks (note: which tasks are performed and who is responsible for each must be made clear)
    - (a) preparation (optional)
    - (b) initial translation (select a point on the spectrum from raw MT to post-edited MT to human translation)
    - (c) quality inspection steps after initial translation (revision [bilingual], review [monolingual], proofreading)
15. additional tasks (e.g. third party review, terminology check, termbase update, software testing, or back translation)

C. Project Environment [16–18]; Relationships [19–21]

16. technology (is any particular technology to be used in the project for a particular point on the spectrum?)
17. reference materials (translation memories, termbases to supplement that in 6b, etc.)
18. workplace requirements (e.g. security measures, confidentiality requirements, or safety concerns in conflict zones)
19. permissions
    - (a) legal factors (e.g. who holds copyright for target text, translation memory, etc.?)
    - (b) recognition (does the translator's name or translation company's name appear in the published translation?)
    - (c) restrictions (on the use of materials, such as translation memory, after the project, etc.)
20. submissions
    - (a) qualifications (expected or required qualifications of provider)
    - (b) deliverables (the target content in the specified format, updated termbase, etc.)
    - (c) delivery method (email, sFTP, etc., including recipient name and address)
    - (d) delivery deadline(s)
21. expectations
    - (a) compensation (typically, cost)
    - (b) communication (channels and mode among contact persons within requester and provider organizations)

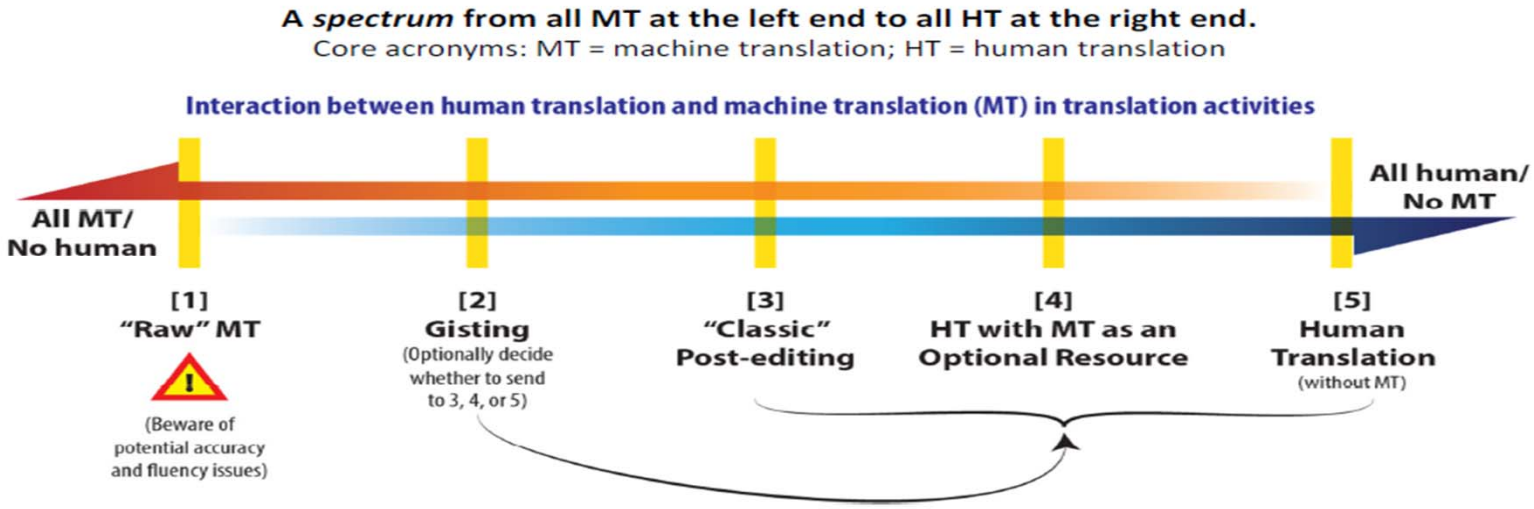
Please turn to your handout consisting of a one-page summary of the 21 parameters

The parameters are tied to the spectrum

› Please see parameter 14b and consult the two-page "spectrum" handout

see parameter 14b (Initial Translation)

› "Select a point on the spectrum" (see handout)



Characteristics of the 5 points on the spectrum

Point on Spectrum	Human Post Processing
1: “Raw” MT	none
2: Gisting	monlingual gisting/triage
3: “Classic” post-editing	bilingual editing
4: HT with MT as option	bilingual translation
5: HT without MT	bilingual translation

- Point 1: Based on specifications, a translator could even recommend raw MT.
- Point 2: Gisting is monolingual; but gisting-based triage might bring in a bilingual.
- In points 4 and 5: the human translator is often using both TM and automatic terminology lookup

Discussion time

- Some translation scenarios from your world
- For each:
  - Preliminary requirements from requester
  - Structured translation specifications, after negotiation
  - Appropriate point on the Spectrum
- Note: in commercial world, Microsoft uses point 1 (raw MT) for tech support, with optional end-user post-editing; in the EU, staff members use point 2 every day to decide whether to send a document to the human translation service; much translation for dissemtination uses point 3, 4, or 5. Prediction: Point 4 will grow the fastest.

As time permits:

› Back to the question of the Singularity

What is the Singularity?

- The word Singularity has several meanings.
- In this lecture, it is a potential future event, predicted by Ray Kurzweil, when computers become more intelligent than humans and then evolve quickly on their own.
- We should remember that if the Singularity arrives, computers will be able to replace translators but will also be able to replace all other professionals, so there is no need for translators to worry about it. *And it is counter-productive to say it won't happen*.

The Singularity and the Spectrum

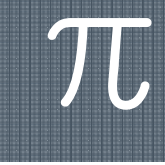
- Wrong question to ask:
  - When will translators be replaced by computers?
- Right question to ask:
  - When will **all** specifications result in a recommendation to use raw machine translation?
- AKM answer:
  - Only when the Singularity has arrived! But then all professional, not just translators, will be replaced by computers.
  - Note: this answer is controversial. Some think all human translators will be replaced before the Singularity.

A very optimistic view of how quickly machine translation will advance

- “Machine translation … still falls short in accuracy, functionality and delivery. That won't be the case for long. A decade from now, I would predict, everyone reading this article will be able to converse in dozens of foreign languages, eliminating the very concept of a language barrier.”
- “… machines will grow exponentially more accurate … It is just a matter of more data, more computing power and better software.”
- Alec ROSS. Wall Street Journal, Eastern edition 30 Jan 2016: page C.3. (see www.alexross.com for his awards)

In contrast with this optimistic view: The reality of current MT is that computers make big mistakes they cannot see, except when tailored to a narrow domain.

See video about following two translations of a recipe for sushi cake: one by a human and the other by a computer. The results are different!



A view from President Xu (CUHK at Shenzhen) said on 2016 April 15:

- Computers are better than humans at memory and reasoning [in a well-defined framework], but
- Humans are better than computers at imagination, emotion, and creation;
- Therefore, humans will remain better than computers for some kinds of translation.
- Concerning the possibility of computers becoming autonomous and refusing to obey, he said that scientists need to think about their responsibility toward society; and
- Therefore, some kinds of development should be slowed down or stopped.

The key question

- Will all translation projects be assigned to point one (raw machine translation) before the Singularity?
- My answer: no!
- On what basis do I make this claim?
  - A career working on translation, starting with machine translation in 1970, continuing with human translation, working on translator tools, and studying philosophy of language

What types of specifications will not be met by computers before the Singularity?

- Those that involve
  - **dynamic general language**
    - dynamic vs. frozen
    - general vs. domain-specific
  - **connecting cultures**, and
  - using "**agency**" (i.e. free will) to deal with aspects of language that go beyond text (words)

Dynamic General Language (three cases)

- New expressions not in the corpus of bitexts (source texts and their translations, segmented and aligned)
- Expressions that are in the corpus with multiple translations, but it is hard to decide which is appropriate
- Multiple translations that are in the corpus, but none of them are appropriate (the “none-of-the-above” scenario)

CONNECTING CULTURES

"Translation is not a matter of words only: it is a matter of making intelligible a whole culture."

-- Anthony Burgess

More about Agency (in the sense of "free will")

- Claim 1: Interesting translation (not just literary translation) goes beyond frozen language, involves non-textual aspects of language, and requires agency.
- Claim 2: Machines do not have agency. If they did, they would have the option to refuse to do what we tell them to do and choose to do something else.
- For a discussion of free will, see Stanford Encylopedia (http://plato.stanford.edu/entries/freewill/) for a philosophical treatment of translation & agency see Melby 1994 (MT archive) and 2010 (Trans-Int.org Context App. 3).

What all translators and translation teachers need to believe in the new age of machine translation

- MT provides translators an opportunity to become language services advisors, if they are not afraid to recommend raw machine translation for some specifications, even though this cuts them out
- Technology, such as translation tools, can help translators deliver better translations while remaining in the driver's seat
- Unless the Singularity actually arrives, the best professional human translators will always have plenty of well-paid work and that work will be interesting because it will require applying agency to connect cultures using dynamic general language

Returning to the title: Associating Structured Translation Specifications with Points on a Spectrum from MT to HT

A

- *Learn* about **specifications** and the **spectrum**

B

- *Believe* in the future of human translators or abandon all hope and change jobs

C

- *Become* excellent as both **translators** and **language services advisors, if you stay**

Pause for questions and comments

Comments after this lecture:

- Alan K. Melby, Brigham Young University Translation Research Group
- Face-to-face today at the workshop or
- Email akmtrg@byu.edu
- with “LEARN spectrum” in the subject line: