History of Machine Translation in the United States

AMTA 2022

Dr. Jennifer DeCamp

MITRE

Background

- 2012 contacted by Routledge Publishing to write an article with Jost Zetzsche on "The History of Translation Technology in the United States" for the *Routledge Encyclopedia of Translation Technology,* published 2014
    - 2019 contacted to write an update, to be published in 2022
    - Discussed with AMTA and ATA leadership that this is a topic to cover as a community

- Like most people in this audience, I have:
    - Taught classes and workshops that included MT and NLP history
    - Provided conference presentations on the history
    - Been around for much of MT development
    - Always had a passion for MT anthropology

- In this presentation, I would like to describe:
    - The history of the history of MT
    - Reality Check: Xerox
    - Gaps
    - Recommendations

The History of the History of MT: W. John Hutchins

1939 – 2021

|	
1960	Graduated with a bachelor's degree in French and German
1962	Obtained a diploma in librarianship
1962-1998	Worked as a librarian, publishing in translation and information retrieval
1978	Authored "Machine Translation and Machine-Aided Translation" in the *Journal of Documentation*
1986	Authored *Machine Translation: Past, Present, and Future*
1992:	Co-authored with Harold Somers: *An Introduction to Machine Translation*
2000:	Authored: *Early Years in Machine Translation* (author/editor)
2015:	Authored: "History of Research and Applications" in *The Routledge Encyclopedia of Translation Technology*
	Developed *MT Compendium of Translation Software*
	Ending in 2014? No prototypes or short-lived products

Donated his extensive library to the MT community (John W. Hutchins Machine Translation Archive)

Somers: "What perhaps many did not realize is that John's work on MT was entirely a labor of love, a kind of hobby, all completed in his own spare time: his job as a librarian did not include working on the MT Archive, nor I think did his employers properly realize and reward his fantastic contribution to the field. We were extremely fortunate to benefit from his skills: from a scientific viewpoint he was an informed observer free of any of the prejudices of the developer or researcher with his own theories and approaches to push."

Summary: Dedicated, detailed, objective librarianship
But ending around 2014; software compendium not covering prototypes and short-lived products; British focus
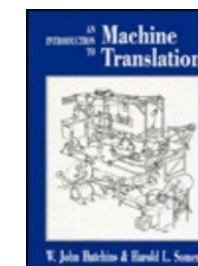
A Small Sample of Other Historians

- Harold Somers
  - 1915 - 2001
  - Hutchins, W. J., & Somer, H. L. (1992). An Introduction to Machine Translation
  - 1978 Retired

- Andy Way

- Chris Wendt

- Steve Richardson

- Mike Dillinger

- Jay Marciano

- Kathleen Egan (retired)

- DARPA and CAMT Program Managers

- Others in U.S. Government, but constrained in what they can say

Summary: More U.S. involvement

                      Many teaching courses and/or providing tutorials on Intro to MT

                      Little in U.S. Government operations

The Tenth Biennial Conference of the
Association for Machine Translation in the Americas

Introduction to MT

Mike Dillinger
Ebay
Jay Marciano
Lionbridge
Technologies

SAN DIEGO, CA
OCTOBER 28 - NOVEMBER 1, 2012

A Selection of Other Histories

- Timelines
  - Wikipedia
  - TAUS

- Short histories
  - Wikipedia
  - IBM
  - Systran
  - AMTA
  - Many others

Summary: Documentation by companies (some—like IBM and SYSTRAN—focusing only on their own contributions), professional organizations, blogs, and Wikipedia

A Selection of Other Resources

- Publications (e.g., Routledge)
- Conference tutorials and presentations
- The W. John Hutchins MT Archive
- The EAMT Software Compendium
- AMTA Resources
- ACL Archives

Summary: Massive information but little curation, except in history-focused publications, tutorials and presentations
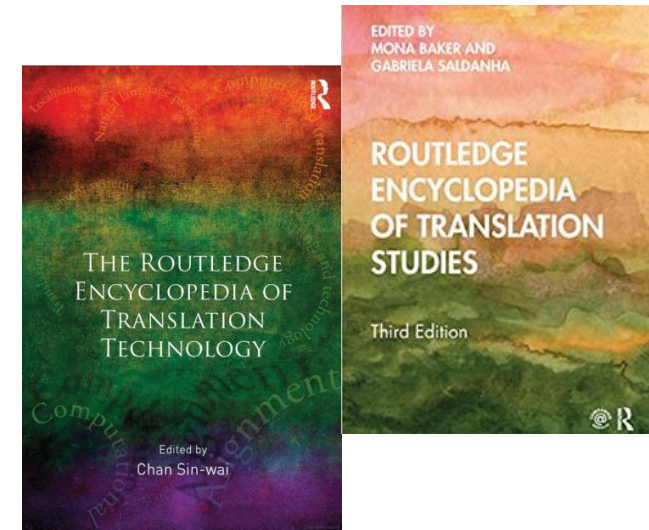
Reality Check: Xerox

- From: *An Introduction to Machine Translation* (Hutchins and Somers 1992)
  - Also checked Hutchins' *Machine Translation Past, Present, and Future (1986) and Early Years in Machine Translation (2000)*

- "Xerox installed Systran in 1982 for technical manuals," using "Multinational Customized English (MCE)," which had about 3000 words and "rules for unambiguous English"

- "At Xerox, texts for translation by Systran are composed in a controlled English vocabulary and syntax; and a major feature of the SMART systems is the pre-translation editor of English input."
  - **Not mentioned:**
    - Areas in source document would be highlighted for editing
    - Corresponding output would be highlighted to alert post-editors

- "The texts that their writers produce are clearer and more understandable"
  - **Not mentioned:**
    - The MT output had more consistent terminology
    - Some technical writers refused to use the pre-editor

- "Output from the system needs little or no post-editing"
  - **Not mentioned:**
    - The system reduced the highly valuable time at the end of the production cycle, when companies would start waiting on purchases until the new version came out
    - The system also reduced the time to produce last-minute revisions and post-shipment revisions.
- No discussion of Xerox DocuTrans
  - In 1989, Xerox provided MT from multiple engines with pre-editing and post editing, including confidence measures
    - Combination of SMART, SYSTRANn, and METAL (Mechanical Translation and Analysis of Languages, started by the Air Force)

Summary: No mention of key applications, confidence measures, post-editing tools, or multi-engine configurations
No mention of refusal by some technical writers to use pre-editor

Gaps

- **Time**, particularly before 1980 and after 2014
  - Due to lack of digitized resources and loss of key librarians
- **Efforts by the U.S. Government**, except for DARPA, IARPA, Wright Patterson Air Force Base, and occasionally a few general papers
  - Due to constraints on what could be publicly released
- **Efforts by LDS Church and other religious organizations**
  - Due to constraints on the quantity of data that could be handled
- **Lack of detail** (e.g., Xerox example)
  - Due to constraints on the quantity of data that could be handled
- **Lack of larger context** (e.g., histories of translation theory and practice, innovation, computer technology, popular culture, etc.)
  - Due to the constraints on the quantity of data that could be handled
- **Lack of information on how practices and decisions turned out** (e.g., Xerox pre-editing interface led to some groups not using the system)
  - Due to lack of time and/or focus
  - Maybe due to Hutchins waiting to see if the system had longevity

Why Try to Fill These Gaps and Provide Analysis?

- Identify best practices (e.g., responding to user feedback)
- Identify requirements and motivations that may have been forgotten (e.g., user interfaces)
- Analyze trends and identify areas of high potential
- Provide long term evaluation of processes and products (e.g., pre-translation editing)
- Improve planning through understanding the accuracy of past projections and forecasts
- Recognize outstanding work
- Protect and celebrate our remarkable history of MT, that helps to build our sense of community

Recommendations

- Address underrepresented areas (e.g., through AMTA panels)
    - Before 1980
    - After 2014
    - Efforts by the U.S. Government
    - Efforts by the LDS Church and other religious groups
    - Long-term results
- Plan AMTA panel on U.S. Government work in MT
    - Obtain more detail on work
    - Obtain official government disclosure and perhaps push the bar on what can be disclosed
- Plan cross-government panel at IAMT
    - Obtain more detail on work
    - Obtain more insights and ideas
- Encourage historical analysis as a field of research in MT/NLP
- Review and encourage expansion histories, timelines, and databases
    - AMTA site
    - EAMT site
    - Wikipedia
    - Publications
    - Company sites

References

Association for Machine Translation in the Americas (AMTA) (2022). "Machine Translation." AMTA Website. Available at: https://amtaweb.org/machine-translation/

AMTA (2022). "Resources." AMTA Website. Available: https://amtaweb.org/resources/#1506630470744-1df80718-e8a9

Baker, M. and Saldanha, G. (2021). *Routledge Encyclopedia of Translation Studies*, Third Edition. Routledge Publishing, London and New York.

Chan, S. W. (2022). *The Routledge Encyclopedia of Translation Technology.* Routledge Publishing Company, London and New York.

Chan, S. W. (2015). *The Routledge Encyclopedia of Translation Technology.* Routledge Publishing Company, London and New York.

DeCamp, J. (2022). "The History of Translation Technology in the United States." In *The Routledge Encyclopedia of Translation Technology*, Ed. Chan S. W. Routledge Publishing Company, London and New York.

DeCamp, J. (2015). "The History of Translation Technology in the United States." In *The Routledge Encyclopedia of Translation Technology*, Ed. Chan S. W. Routledge Publishing Company, London and New York.

Egan, K., Kubala, F., and Sears, A.(24 October 2008). "User-Centered Development and Implementation." Presentation at AMTA Government Users Presentation, AMTA Conference Waikiki. Available: https://aclanthology.org/2008.amta-govandcom.9.pdf

Hutchins, W.J. (2022). The W. John. Hutchins Machine Translation Archive. European Association for Machine Translation (EAMT) website. Available at https://mt-archive.net/

Hutchins, W.J. (2022). "Compendium of Translation Software." EAMT website. Available at: https://compendium.eamt.org/compendium/index.php

Hutchins (2015). "History of Research and Applications." In *The Routledge Encyclopedia of Translation Technology*, Ed. Chan S. W. Routledge Publishing Company, London and New York.

Hutchins (Ed.) (2000). *Early Years in Machine Translation*. John Benjamins Publishing Company, Amsterdam and Philadelphia.

Hutchins, W. J. (June 1978). "Machine Translation and Machine-Aided Translation" in the *Journal of Documentation,* Vol. 24, No. 2, pp. 119-159. Available at: https://mt-archive.net/70/JDoc-1978-Hutchins.pdf

Hutchins and H. Somers (1992). *An Introduction to Machine Translation*. Academic Press. Harcourt Brace Jovanovich Publishers, London.

IBM (2022). "701 Translator: IBM." IBM Archives website. Available: https://www.ibm.com/ibm/history/exhibits/701/701_translator.html

SYSTRAN (2022). "SYSTRAN: 50 Years of MT Innovation." SYSTRAN website. Available: https://www.systransoft.com/systran/translation-technology/systran-50-years-of-mt-innovation/

TAUS (2022). "Timeline." TAUS website. Available https://TAUS.net

Wikipedia (2022). "History of Machine Translation." Available at https://en.wikipedia.org/wiki/History_of_machine_translation