

# **Current and potential applications of machine translation**

**John Hutchins**

Presentation on 29 November 2012  
at Translating and the Computer

# MT for dissemination: traditional use

Production of 'publishable' quality texts

Editing of 'raw' output

Post-processing, post-editing

- Statistical post-editing, text prediction

Control of input (pre-editing, 'controlled language')

Domain restriction

Combined with translation memories

Numerous applications (e.g. software documentation)

Demand and need will continue to grow

- Further applications and aids

# MT for assimilation

- fast/immediate
- readable (intelligible), for information use, publication quality not necessary
  - intelligence services (e.g. NAIC)
  - occasional translation (home use)
- use as draft for translation
- aid for writing in foreign language (e.g. by EC administrators)
- emails, Web pages
- **any** system type can be used (including those originally for mainframes and PCs)
  - online MT has all types of rule-based and SMT systems

# MT for interaction

Correspondence, emails

In future: special-purpose systems for emails and for business correspondence (with interactive authoring in controlled language)

Interchange in military situations (Diplomat, Phraselator)

Interchange in medical situations (MedSLT, Converser)

Interchange in tourism situations

Interchange by telephone (ATR)

Interchange in business oral communication (Verbmobil)

# Online MT

- No data on users: ages, background knowledge, types of texts, etc.
- For many users only experience of MT – no knowledge of PC systems
- 'tested' on idioms, ambiguities, back-translation
- Most online services use statistical MT
- Used by translators as rough drafts?
- Average length 20 words; 50% of submitted 'texts' just one or two words
- Very few webpages (unexpected!)
- Overall usage continues to grow exponentially; has raised MT awareness, but at a cost...
- Wider range of access: PC, laptop, mobile, netbook, notebook, tablet, etc.
- Little use of specialist dictionaries or of value-added' services (e.g. post-editing)
- The less the language knowledge of users , the more useful
- Quality improvements will be slow or negligible

# Webpages

- Colloquial, culture-dependent language
- Texts in graphic format cannot be translated (very common in Japanese webpages)
- But website developers often recommend users to online MT services – do they know the dangers to their reputations?
- Website localisation for companies, etc. (IBM Websphere)

# MT for minorities

- May be widespread globally, but minor in particular country (e.g. Hindi in UK)
- Not commercially significant
- Poor resources (dictionaries, grammars)
- Not even word-processing (alphabets)
- Lack of bilingual corpora
  - even SMT rapid development not an option
- Instead of MT often 'low-level' aids more suitable/effective

# Embedding MT

- Information retrieval
  - Multilingual access to information
- Information extraction
  - data mining, text mining
- Transliteration
- Summarization
- Question answering (cross-language)
- Authoring software



# Subject-specific MT systems

Sublanguage systems (few successors of Meteo)

e.g. police, drug enforcement, news

- Easier with SMT systems; fast production (cheaper?)

Commercial PC systems for medicine/patents (Japanese)

Availability of special glossaries (ranked for preference)

e.g. medicine, law, Bible, business

Wide range available (but how many sold/used?)

# Other applications - current

- [mostly still mainly research]
- subtitles, broadcast transcripts, syndicated feeds
- distance education, language teaching
- doctor-patient communication
- emergency services
- MT for the deaf

## Other applications - future

- chatrooms, social networking (Facebook, etc.)
  - Problems comparable to spoken language translation
- emergency services
- Speech translation [general, telephone]
- Photocopier-MT; Scanner-MT ('pen' scanner)
- Camera-MT (menus, road signs?)
- Spectacles-MT
- MT for robots (spoken commands?)
- decipherment (back to MT origins!)
- Universal Translator

# Future expectations (1)

- merging of MT and TM for enterprise dissemination systems
- Internet as major (chief) resource - not only SMT
- rapid development of systems (SMT),
- more language pairs (SMT)
- reuse of MT components (for closely related languages)
- open source: wider range of tools (both RBMT and SMT)

# Future expectations (2)

improvements in quality (hybrid, multi-engine systems)

minor (and minority) languages

i.e. languages not of major commercial or military interest

special-purpose systems (domain and function) - also online  
spoken language MT:

domain-specific only [not general-purpose]

embedding of MT in other LT systems

MT seen in future as aids for bilingual (multilingual) communication  
as much as for translation

# Quality of MT: general considerations

- ‘inherent’ limitations of MT:
  - complexity (linguistic and computational); sentence based
  - constant creation of neologisms and new meanings for words and word sequences; constant creation of new idioms
  - lack cultural background, wider context
  - need encyclopedic knowledge; lack common sense
  - therefore no Universal Translator
- Perfectionism is not necessary or essential
  - publishable quality will always require human editing/revision
  - assimilation/interchange can always tolerate imperfect communication

# Resources online

Machine Translation Archive ([www.mt-archive.info](http://www.mt-archive.info))

My website for *history of MT* ([www.hutchinsweb.me.uk](http://www.hutchinsweb.me.uk))

Statistical Machine Translation ([www.statmt.org](http://www.statmt.org))

Associations:

EAMT ([www.eamt.org](http://www.eamt.org))

AMTA ([www.amtaweb.org](http://www.amtaweb.org))

TAUS [Translation Automation User Society]  
([www.translationautomation.com](http://www.translationautomation.com))