

Machine translation: past imperfect, future indefinite

John Hutchins

(Email: WJHutchins@compuserve.com)

[<http://ourworld.compuserve.com/homepages/WJHutchins>]

Imperial College, London

March 2003

Early methods of MT

- word for word
 - Richens and Booth, Yngve
- lexicographic
 - Reifler, King (IBM-USAF system)
- direct
 - Georgetown, ITMVT (Moscow)
- interlingua, semantics
 - CLRU, Milan, Mel'chuk, Leningrad
- syntax
 - Bar-Hillel (categorical), Oswald+Fletcher, MIT (phrase structure) , NBS (predictive), RAND (dependency)
- transfer
 - MIT (syntactic transfer)
- statistics
 - information theory (Weaver), distribution analysis (RAND), redundancy (King)
- pre-editing, controlled language, restricted language
 - Reifler, Dodd
- post-editing
 - Bar-Hillel

The development of MT: 1950s and 1960s

- Sponsored by government bodies in USA and USSR (also CIA and KGB)
 - assumed goal was fully automatic quality output (i.e. of publishable quality) [dissemination]
 - actual need was translation for information gathering [assimilation]
- Survey by Bar-Hillel of MT research:
 - criticised assumption of FAHQT as goal
 - demonstrated ‘non-feasibility’ of FAHQT (without ‘unrealisable’ encyclopedic knowledge bases)
 - advocated “man-machine symbiosis”, i.e. HAMT and MAHT
- ALPAC 1966, set up by disillusioned funding agencies
 - compared latest systems with early unedited MT output (IBM-GU demo, 1954), criticised for still needing post-editing
 - advocated machine aids, and no further support of MT research
 - but failed to identify the actual needs of funders [assimilation]
 - therefore failed to see that output of IBM-USAF Translator and Georgetown systems were used and appreciated

Consequences of ALPAC

- MT research virtually ended in US
- identification of actual needs
 - assimilation vs. dissemination
- full automation vs. HAMT and MAHT
- recognition that ‘perfectionism’ (FAHQT) had neglected:
 - operational factors and requirements
 - expertise of translators
 - machine aids for translators
- henceforth three strands of MT:
 - translation tools
 - operational systems (post-editing, controlled languages, domain-specific systems)
 - research (new approaches, new methods)

From 1967 to 1979

- Continuation of research in US (Texas, Wayne State), Soviet Union, UK, Canada, France
- rule-based approaches: interlingua and transfer
- 1970: Systran installed at USAF (Foreign Technology Division)
- 1970: TITUS installed (restricted language: textile industry abstracts)
- 1975: Météo ‘sublanguage’ English-French system (weather broadcasts)
- 1975: CULT Chinese-English (restricted language: mathematics)
- 1976: European Commission acquires Systran
- 1979: Pan American Health Organization system (SPANAM)
- 1979: Eurotra project begins

MT research in 1970s and 1980s

- Rule-based systems:
 - involving long-term efforts compiling grammar rules (interlocking) and creating dictionaries
- Interlingua systems
 - DLT, Rosetta, Carnegie Mellon
- Transfer-based systems
 - GETA (Ariane), SUSY, Eurotra, Mu (Kyoto)
- Knowledge-based systems
 - Carnegie Mellon, New Mexico, Pangloss
- Speech translation
 - ATR, C-STAR, Verbmobil
- **Computer-based tools**

Changes since late 1980s

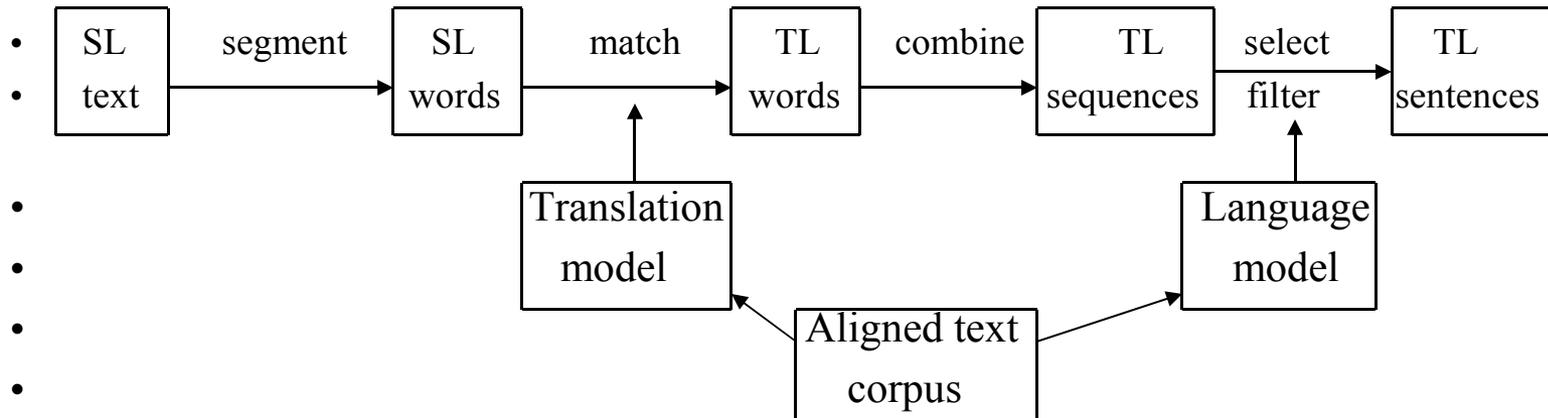
- Increasing use of MT by large enterprises
- Translation memory and translation workstations
- Localization
- Growth in PC systems
- The impact of the Internet
- Online translation
- MT and other language activities
- Research on corpus-based MT methods

System architectures and strategies

- Rule-based
 - Direct translation
 - Interlingua-based MT
 - Transfer-based MT
- Corpus-based MT
 - Statistics-based
 - Example-based
- Hybrid systems

Statistics-based MT

- Based on observations that translations observe statistical regularities
 - TL words are chosen as those most likely to correspond with the SL words in specific context
 - TL words are combined in ways most appropriate for the TL in a specific context/domain and style/register etc.



Statistics-based MT

- Bilingual corpora: original and translation
- little or no linguistic ‘knowledge’, based on word co-occurrences in SL and TL texts (of a corpus), relative positions of words within sentences, length of sentences
- Sentences aligned statistically (according to sentence length and position)
- compute probability that a TL string is the translation of a SL string (‘translation model’), based on:
 - frequency of SL/TL co-occurrence in aligned texts of corpus
 - position of SL words in SL string, and TL words in TL string
- compute probability that a TL string is a valid TL sentence (based on a ‘language model’ of allowable bigrams and trigrams)
- search for TL string that maximizes these probabilities
- first example: IBM Candide (1988) on Canadian Hansard (English and French)

Statistics-based MT: problems

- still insufficient corpora
 - but Internet may solve this
- corpus must be aligned and analysed before translation of (similar) text in same domain
 - unless large corpus for domain available
- word frequencies not sufficient: Candide intended to add morphological information, and some grammatical categories
 - some of this information may be statistically derived from large corpora
- most research aims to test how far purely statistical methods can go
 - laudable as research project, but not for developing working systems
 - in my view, some research needed on practicality of SMT for ‘real’ systems

Problems of alignment (1)

- bilingual corpora
 - suitability (i.e. appropriate domain, style, audience)
 - availability, e.g. for uncommon languages (lack of electronic resources)
- matching sentence lengths (for European languages, not for English/Japanese)
- matching words
 - cognates: first four letters and ‘same’ meaning (*mathematics* and *mathématique*)
 - - but fails for *government/gouvernement*, and *actual/actuel*
 - morphological patterns: *book/books*, *box/boxes*, *lady/ladies*, *wife/wives*, etc.
- using bilingual dictionaries (as seed for alignment: simple word pairs)

Problems of alignment (2)

- Work best for word-to-word alignment

– well, I think if we can make it at eight on both days
– ja, ich denke wenn wir das hinkriegen an beiden Tagen acht Uhr

- Difficulties when a SL word group (phrase) corresponds to TL word group

– yes, then I would say , let us leave it at that.
– Ja, dann würde ich sagen , verbleiben wir so.

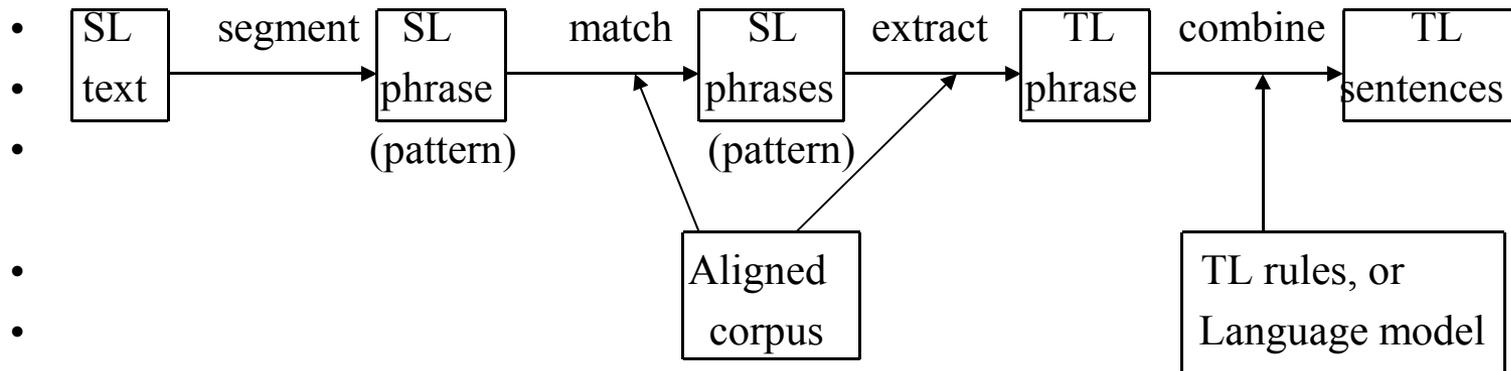
- Problems with inadequate training corpus

Translation memories: weaknesses

- Expensive to build (in time and money)
- sentence-based comparisons restrict potential use (no phrase matching); whole sentence repetition is rare (except with revised texts)
- loss of context beyond sentence
- any TM likely to contain redundant, ambiguous versions
- any TM likely to contain conflicting translations (with little or no guidance)
- sentences are edited by translators outside TM environment and therefore not included in the database
- TM systems do not ‘learn’ decisions/choices made by users (e.g. which potential translations are preferred, which rejected)
- fuzzy matching often fails (hidden tags) and too complex, and translators opt not to use the facility; prefer translating from scratch
- combining extracted translation segments left entirely to user/translator
- developments needed:
 - **finding phrases (retrieval, fuzzy matching)**
 - **combining phrases; searching for words in combination**
 - **phrase repetition (‘example-based’ approach?)**

Example-based MT

- Based on observation that translators try to find similar SL phrases and sentences and their TL equivalents in previously translated texts
 - seek sets of analogies and examples from bilingual corpora



Example-based MT: some problems and issues (1)

- bilingual aligned corpora
 - size: adding examples may improve performance or may degrade performance
 - repetition of same or similar examples may reinforce selection or may be unnecessary clutter
 - suitability of examples: automatically compiled or manually compiled
 - need: phrases/clauses aligned (not sentences), length is open issue
 - stored: as word strings or as annotated trees(e.g. dependency or case grammar trees)
- analysis of corpus at run-time or in advance
- use of grammatical categories (patterns)
 - templates (e.g. <1st name><family name> flew to <city> on <date>)
 - X [pron] eats Y [noun/NP] ↔ X [pron] ga Y [noun/NP] o taberu
 - X o onegai shimasu → may I speak to the X (if X=jimukyoku ‘office’, ... etc.); or: please give me the X (if X=bangō ‘number’, ... etc.)

Example-based MT: some problems (2)

- matching by characters:
 - This is shown as A in the diagram ↔ This is shown as B in the diagram
 - The large paper tray holds up to 400 sheets <≠> The small paper tray holds up to 300 sheets
 - (because system does not know that *large* and *small* are similar/substitutable)
- matching by words via thesaurus (close in meaning)
 - English *eat* → Japanese *taberu* or *okasu*
 - *A man eats vegetables* ↔ *Hito wa yasai o taberu*
 - *Acid eats metal* ↔ *San wa kinzoku o okasu*
- problem of ‘boundary friction’
 - that old man has died ↔ ce vieil homme est mort
 - that old woman has died ↔ (not simple substitution: ce viel femme est mort), **but**: cette vieille femme est morte

Bilingual lexical differences

- bilingual lexical ambiguity (more than one equivalent, whether ambiguous in SL or not):
 - river: fleuve/rivière
 - Taube: dove/pigeon
 - Schraube: screw/bolt/propellor
 - corner: coin or angle; Ecke or Winkel
 - light: léger, clair, facile, allumer, lumière, lampe, feu
 - look: regarder, chercher, sembler
- lexical gaps
 - dacha, cottage, marmelade, vodka, etc.
 - snub: infliger un affront; verächtlich behandeln, or: derb zurückweisen
 - het Turks kennen: to know Turkish
 - kenner van het Turks: *knower of Turkish, someone who knows Turkish
- **Solved (?) by contextual rules (RBMT), or examples (EBMT), or word-word frequencies and ‘language models’ (SMT)**

Structural ambiguity

- (1) Peter mentioned the book I sent to Mary
 - Peter mentioned the book which I sent to Mary
 - Peter mentioned to Mary the book which I sent [to Peter/David]
- (2a) We will meet the man you told us about yesterday
 - ... the man you told us about yesterday
- (2b) We will meet the man you told us about tomorrow
 - we will meet tomorrow the man...
- (3a) pregnant women and children
 - des femmes et des enfants enceintes
- (4a) Smog and pollution control are important factors
- (4b) Smog and pollution control is under consideration
- (4c) The authorities encouraged smog and pollution control
- (5a) old men and women may usually mean ‘old men and old women’
- (5b) [but perhaps not in] Tickets were refunded for children, old men and women
- **Problems (1), (2), (3), and (5a) may be ‘solved’ by SMT ‘language model’ and by EBMT databases. But problems (4c) and (5b) require ‘knowledge’ (i.e. rule-based KBMT)**

Bilingual structural differences

- (1) Young people like this music
 - Cette musique plaît aux jeunes gens
- (2) The boy likes to play tennis
 - Der Junge spielt gern Tennis
- (3) He happened to arrive in time
 - Er ist zufällig zur rechten Zeit angekommen
- (4) Le moment arrivé je serais prêt
 - When the time comes, I shall be ready
- **Difficult to specify rules to cover all circumstances and contexts; example-based (EBMT) and statistics-based (SMT) yet to prove any better; possibly examples like no.4 are inherently unsolvable**

Anaphora

- Die Europäische Gemeinschaft und ihre Mitglieder
 - The European Community and its members
- The monkey ate the banana because it was hungry
 - Der Affe ass die Banane weil er Hunger hat
- The monkey ate the banana because it was ripe
 - Der Affe ass die Banane weil sie reif war
- The monkey ate the banana because it was lunch-time
 - Der Affe ass die Banane weil es Mittagessen war
- Particular problem when translating from Japanese when it is good style to omit the subjects of verbs and to avoid repetition.
- **Sentence-orientation of all systems makes most anaphora problematic (unresolvable); possibly only a discourse-oriented ‘language model’ is the only chance**

Non-linguistic problems of ‘reality’

- The soldiers shot at the women and some of them fell
- The soldiers shot at the women and some of them missed
 - must know what ‘them’ refers to e.g. if translating into French (ils or elles)
- **No solutions with linguistic rule-based approaches**
- **No solutions with corpus-based approaches**
- **Perhaps only solution using Artificial Intelligence approaches
(Knowledge-based machine translation, e.g. Carnegie-Mellon University)**
- However, perhaps this aspect is sometimes exaggerated: no need to understand what AIDS and HIV are in order to translate:
 - The AIDS epidemic is sweeping rapidly through Southern Africa. It is estimated that more than half the population is now HIV positive.

Problems of stylistic difference

- The possibility of rectification of the fault by the insertion of a valve was discussed by the engineers [nominalization style]
- The engineers discussed whether it was possible to rectify the fault by inserting a valve [preference for verb forms]
- [English] Advances in technology created new opportunities
- [Japanese] Because technology has advanced, opportunities have been created
- [or Japanese] Technology has advanced. There are new opportunities.
- **All methods of MT tend to retain SL structural features; however, theoretically SMT ‘language model’ approach should be more TL-oriented.**

Hybrid systems

- **clearly, none of the current MT ‘models’ are capable of solving all problems**
- **hence search for hybrid architectures**
- **in theory, it would seem that (on average):**
 - **RBMT better for SL analysis**
 - **EBMT better for transfer**
 - **SMT best for TL generation**
- **Problem is that different approaches not easily compatible:**
 - **there are however research prototypes combining:**
 - **EBMT with statistical methods**
 - **EBMT using rules similar to those in RBMT systems**
 - **perhaps a version of EBMT will be the answer**
- **Currently ‘hybrid’ systems are parallel systems with a selection mechanism, as in:**

Speech translation: problems

- speech recognition, speech synthesis
- highly context dependent, use of ‘knowledge databases’
- discourse semantics, ‘ill-formed’ utterances
- ellipsis, use of stress, intonation, modality markers
- restricted domain (e.g. hotel booking by telephone)
- colloquial usage not yet investigated sufficiently (even in linguistics)

- half-way ‘solutions’ available with voice input/output
 - Word processing add-ons (Dragon Naturally Speaking, IBM ViaVoice)
 - PC translation systems with voice input/output (Al-Wafi, CITAC, ESI, Korya Eiwa, Personal Translator PT, Reverso Voice, TranSphere, Vocal PeTra, ViaVoice Translator)
 - Online translation with voice output (Translation Wave)

The translation demand

- Dissemination: production of ‘publishable quality’ texts
 - but, since raw output inadequate:
 - post-editing
 - control of input (pre-editing, controlled language)
 - domain restriction (reducing ambiguities)
- assimilation: for extracting essential information
 - use of raw output, with or without light editing
- interchange: for cross-language communication (correspondence, emial, etc.)
 - if important: with post-editing; otherwise: without editing
- information access to databases and document collections
 - limited use before 1990

MT for corporations: issues

- General-purpose system or specialised system
- Controlled language
- Lexical resources
- Management implications
- Control of terminology
- Consistency
- Standards; exchange formats
- Compatibility (hardware, software)
- Integration: technical authoring, publishing

Large-scale translation and MT

- accurate, good quality, publishable (dissemination)
- publicity, marketing, reports, operational manuals, localization
- technical documentation; large volumes
- repetitive, frequent updates; saving costs (and staffing?)
- multilingual output (e.g. English to French, German, Japanese, Portuguese, Spanish)
- available in-house terminological database; user (company) dictionaries
- backup resources (translated texts, personnel for dictionaries, etc.)
- human assistance for quality (controlled language input, post-editing)
- integrate with technical writing and publishing
- availability of in-house printing/publishing
- technical expertise (computers, printers, etc.)

Controlled language

- Controlled authoring of the source text in standard manner, suitable for unambiguous translation
- Typical rules:
 - use only approved terminology, e.g. *windscreen* rather than *windshield*
 - use only approved sense: *follow* only as ‘come after, not ‘obey’
 - avoid ambiguous words: *replace*, either (a) remove and put back, or (b) remove and put something else in place; not *appear* but: come into view, be possible, show, think
 - only one ‘topic’ per sentence, e.g. one instruction, command
 - do not omit articles; use relative pronouns (which, in order that); avoid post office-nominally gerundive form (*wires connecting...→ wires that connect...*)
 - do not use pronouns instead of nouns if possible
 - do not use phrasal verbs, such as *pour out*
 - do not omit implied nouns
 - use short sentences, e.g. maximum 20 words
 - avoid co-ordination of phrases and clauses
- **advantage of controlled language is improvement of original SL text; sometimes translation no longer necessary; later revision can be faster**

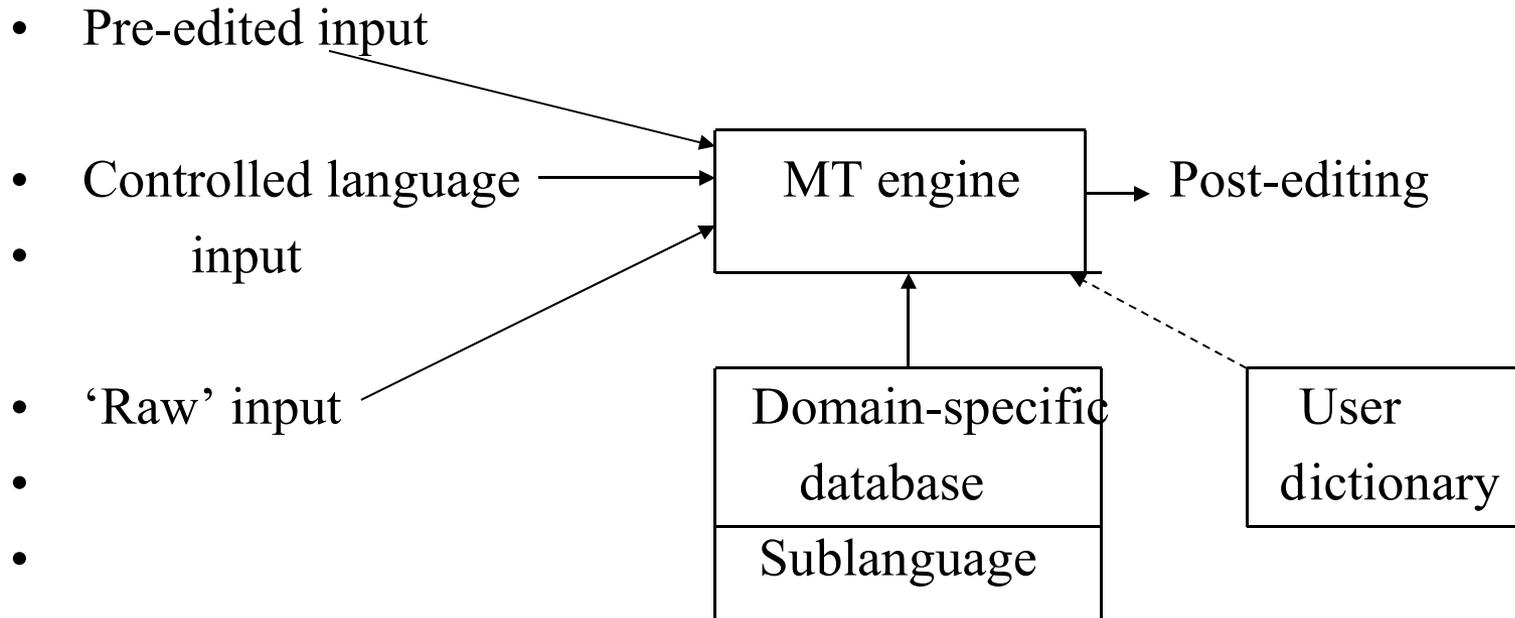
Controlled language and special-purpose systems: requirements and issues

- system developed by external agency (e.g. Smart, LANT) or in-house?
- special dictionaries (domain, company): existing, or to develop?
- terminology databases
- new or adapted from existing controlled languages
 - despite previous models, SAP developed own language (SKATE)
- grammar and style analysis (usual grammar checkers inadequate)
- lexicon
 - internal (company) and external (standard terminology)
- grammar
 - to be recommendations or to be obligations

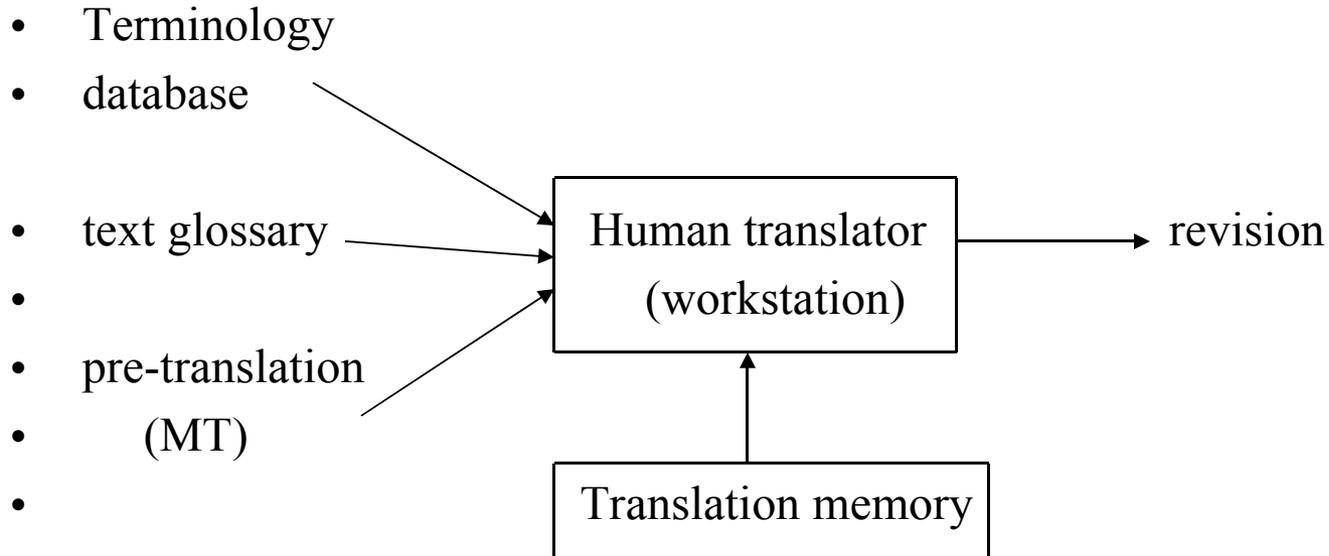
Lexical acquisition

- dictionary building
 - hand-crafted (pre-1990) was expensive in time and effort
 - required information: morphological variants, grammatical categories, syntactic contexts, lexical co-occurrences, semantic conditions/constraints, translation options
 - generally more detailed than terminology information for human translation (and includes **all** words)
- major problem for all current (commercial and custom-built) systems
- providers: vendor vs. customer
 - basic dictionary, special dictionaries, user dictionary (customer-specific)
 - terminologists, database managers
- resources for creating dictionaries
 - size (what is adequate? definition of domain)
 - use of lexical resources (printed dictionaries, Internet dictionaries)
 - extraction from electronic texts (monolingual/bilingual, internal, Internet, Web pages): word alignment
 - validating, standardization, checking, updating, sharing
 - conversion into required formats for particular MT system
 - software (MultiTerm, TMX, etc.)
- **corpus-based methods do not require detailed dictionaries** (future prospect)

Human-assisted MT



Machine-aided human translation



Computer-aided translation and translation tools

- recognition that fully automatic translation not appropriate for professional translators
- PCs and multilingual word processing, desk top publishing
- Translator ‘in control’
- dictionaries (monolingual, bilingual): on-line access
- grammar aids, spelling checkers
- user glossary, terminology management, ‘authorised’ terms, standards, specialist glossaries
- input, output, transmission (OCR, pre-editing, controlled language)
- translation memory, alignment
- management support tools (project control, budgeting, workflow)
- previous antagonism of translators to MT diminished

Convergence of HAMT and MAHT

- increasingly, systems straddle different categories
- workstations (TM systems) include MT components (e.g. Trados, Atril)
- MT systems include TM components (e.g. globalwords)
- localization systems embracing, or as components of, either TM or MT systems
- common facilities:
 - terminology management; integration with authoring and publishing systems; project management; quality control; Internet access and downloading; Lexical acquisition; Web translation
- common aim: production of quality translations for **dissemination**; utilization of translator skills
- at present: both approaches in parallel rather than integrated
- in research: EBMT investigates merging of rule-based and database methods
- future: full integration (no distinctions)

MT for translators (office systems): issues

- translation database -- ownership, copyright
- terminology management -- acquisition
- integration with other IT equipment
- translation workstations still too expensive for individual translators
- insufficient functionality in downsizing systems for large organizations onto stand-alone (PC) systems
- suitable project management tools (currently most for large agencies and companies)
- **currently downsized versions of 'enterprise' systems, or upgraded version of 'home' systems - not yet well defined category**

MT for assimilation (home use, online)

- must be fast, immediate, real-time
- must be readable, but accept poor quality
- more languages
- webpage compatible (translate graphics)
- translate electronic mail
- steady improvement

Has MT improved?

- In what respect?
 - translation quality: general-purpose vs. domain-specific
 - usability (ease of use)
 - adaptability (integration with other software)
- Since when?
 - quality perhaps not in last ten years, but since 1980 it has
- Why not?
 - inherent problems of language
 - inherent problems of ‘cultural’ differences

Systran at EC example (English to French)

- [English original]
 - Since no request concerning changed circumstances with regard to injury to the Community industry was submitted, the review was limited to the question of dumping.
- [French 1987]
 - Puisqu’aucune demande concernant les circonstances changées en ce qui concerne la blessure à l’industrie communautaire n’a été soumise, l’étude était limitée à la question de déverser.
- [French 1997]
 - Comme aucune demande concernant un changement de circonstances en ce qui concerne le préjudice causé à l’industrie communautaire n’a été présentée, le réexamen était limité à l’aspect du dumping.

Systran at EC example (French to English)

- [French original]
 - Leur objet n'était pas de formuler des recommandations politiques, mais de servir de base analytique à la réflexion politique.
- [English 1987]
 - Their object was not to formulate of the political recommendations, but to be used as a basis analytical for the political reflexion.
- [English 1997]
 - Their object was not to make political recommendations, but to serve as an analytical base to political reflection.

Online and PC translation: why so bad?

- old models (word for word, simple transformer architecture)
 - often single equivalents, no morphological analysis or target adjustment
- dictionaries too small, insufficient information, and difficult (or impossible) to update
- weak syntactic analysis/transfer
- poor disambiguation (little semantic information)
- general-purpose (not domain restricted)
- not designed for language/style of emails
- web page translations: graphics not translated, distorted, ignored; format lost
- need special functions if used as aid for writing in foreign language
- language coverage uneven; many languages of Africa and Asia are lacking
- translation from English often poorer than into English

- **conclusion: of use/value only if source language unknown or known only poorly and if essence and not full information is adequate**
- **the less the user knows the source language, the more useful becomes automatic translation**

MT in the marketplace

- retail availability
 - many only purchased direct from manufacturer
- promotion by vendors
 - confusion of terminology:
 - some ‘translation systems’ are no more than dictionaries
 - ‘computer aided translation’ either HAMT or MAHT
 - combination of MT and support tools
 - translation memories either independent or components
- low profits, slow quality improvement, few differences between rivals
- categorisation (enterprise, professional, home, workstations) unclear
- expectations of users
 - steady (faster) quality improvement
 - more languages
- suitability of system to expected use
- bench marks, consumer reports/reviews

MT for interchange: what's needed?

- correspondence, emails, etc.
- in principle, any systems can be used for written interchange
 - many PC systems have specific facilities for email translation
- in future there may be special-purpose systems for business correspondence (e.g. with interactive authoring in controlled language)
 - has been subject of research (e.g. UMIST)
- interchange in military ('field') situations
 - e.g. systems for translating standard phrases (Diplomat, Phraselator)
- interchange in tourist situations
 - so far only dictionaries of words and phrases (hand-held devices)
- interchange with deaf and hearing impaired
 - translation into sign languages [mainly research so far]
- interchange by telephone or in business oral communication
 - still at research stage (speech translation)
- interpreting ex tempore (unlikely ever to be even semi-automated) , but:
 - interpreters (at EC etc.) do use rough MT of technical speeches to aid them

MT and other LT applications

- document drafting
 - Japanese researchers, EC administrators, school essays
- information retrieval (CLIR): translation of search terms
- information filtering (intelligence):
 - for human analysis of foreign language texts
 - document detection (texts of interest); triage (ranking in order of interest)
 - deciding whether text worth translating (discard irrelevant ones)
- information extraction: retrieving specific items of information (domain-tuned, captured by key words/phrases)
 - e.g. specific events, named people or organizations
- summarization: producing summaries of foreign language texts
- multilingual generation from (structured) databases
- localization of interactive commands (computers, mobile phones)
- television subtitling
- language teaching: MT as aid for teaching translation

MT: when it works and when it doesn't

- cannot be both fully automatic (no pre- or post-editing) and general-purpose
- beyond its scope:
 - literature, philosophy, sociology, law
- large corporations, cost-effective if:
 - controlled input, standardised terminology, multilingual output, repetitive documentation, restricted domain
- occasional (information-only)
 - rough, not for publication; immediate (fast) production
- small-scale MT
 - ‘formulaic’ documents (business correspondence), restricted domain
 - interactive assistance

MT and HT in complementation

- Dissemination
 - HT: single item, context/culture-sensitive,
 - HT with TM: repetitive (e.g. localization, web localization)
 - MT only: restricted language, repetitive (e.g. Meteo); document drafting
 - MT with post-editing/controlled language: large scale, technical, localization
- Assimilation
 - MT with (rapid) post-editing: scientific, technical
 - MT only (PC or online): single item (non-literary), general purpose; information monitoring/filtering
 - MT domain-specific (online)
- Interchange
 - HT: business correspondence; interpreting
 - MT: email, personal correspondence; database searching; TV captions
 - MT domain-specific: business correspondence
 - MT (speech) domain-specific: telephone enquiries

General comments

- MT is not *translation* as usually understood, it is merely a computer-based tool
 - for translators
 - for cross-language communication
 - for access to information resources
- Perfectionism is not necessary or essential
 - publishable quality will always require human editing/revision
 - assimilation/interchange can always tolerate imperfect communication
- MT should be used only as required to save costs/effort in appropriate circumstances
- Judgement should be based
 - ***not*** on whether system produces ‘real’ translations
 - and particularly not whether it produces ‘good’ translations
 - ***but***: whether the output can be *used*
 - and: whether its use will save time or money

New directions and challenges

- Spoken language translation
 - general-purpose?
- ‘Minor’ languages
 - languages of India, Africa, Asia
 - non-national (‘official’) languages (e.g. Welsh, Basque, Catalan)
 - languages of minorities (e.g. non-indigenous languages in Britain)
- Systems for monolinguals
 - from unknown source language
 - to unknown target language
- Improvement expectations
 - particularly PC commercial and Internet systems
- Reusability of resources (particularly dictionaries and translation memories)
- Integration
 - MT as option in word processing packages, on Web pages
 - MT as option with summarization, information extraction, information retrieval, data retrieval, question-answering, Internet search tools

Sources of information

- EAMT website (www.eamt.org) with links to other IAMT sites, etc.
- LISA website (www.lisa.org)
- Conferences:
 - MT Summit, EAMT workshops, AMTA conferences, LISA Forums
- Journals (non-research):
 - *Multilingual Computing and Technology*
 - *MT News International*
- *Compendium of translation software*
- my website:
 - <http://ourworld.compuserve.com/homepages/WJHutchins>