Machine translation: current research problems and issues, and future prospects

John Hutchins

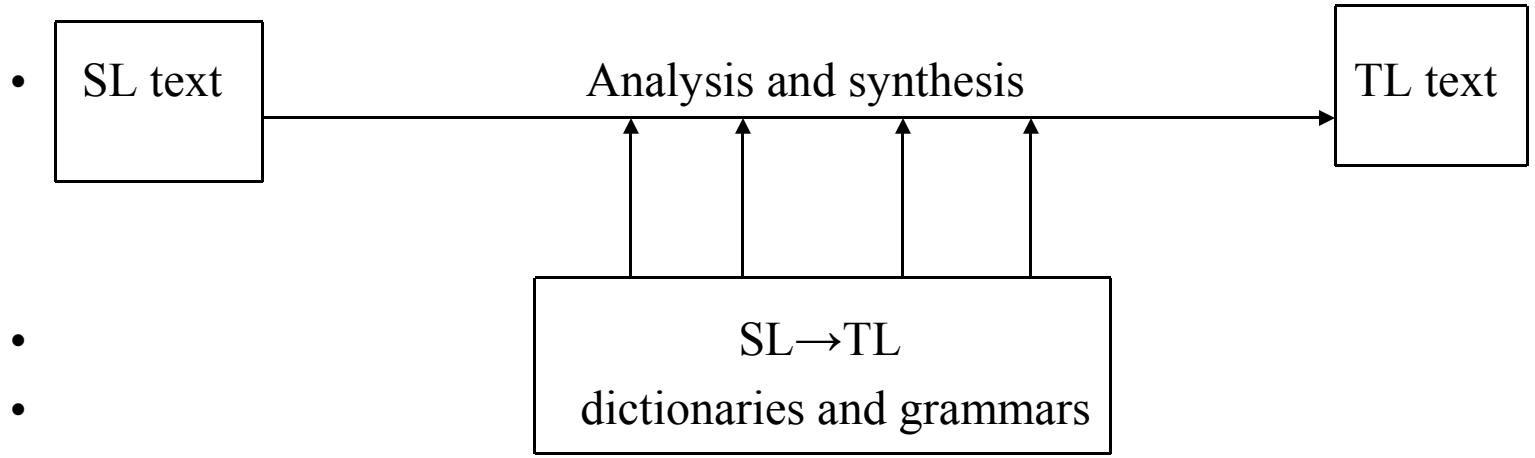
(Email: WJHutchins@compuserve.com)
[http://ourworld.compuserve.com/homepages/WJHutchins]

University of Valladolid
March 2003

System architectures and strategies

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

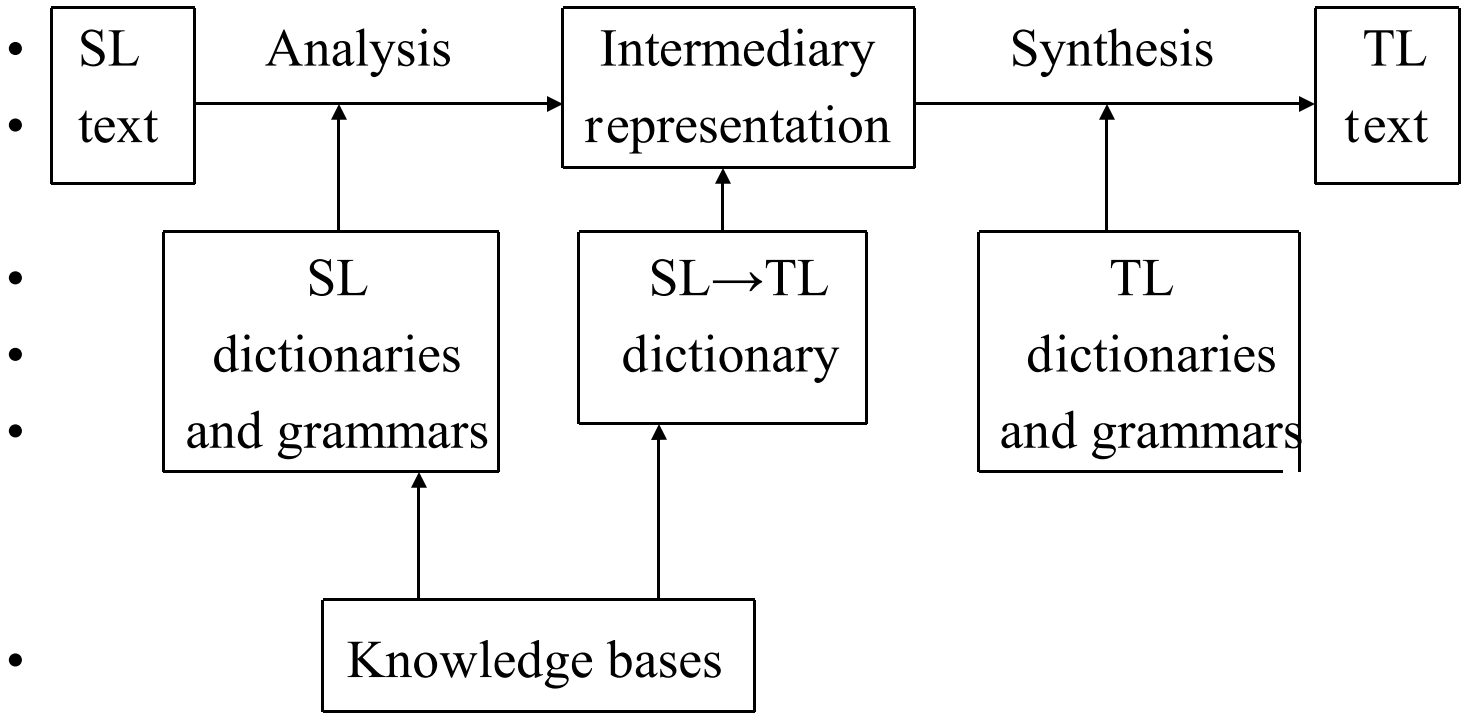
Direct translation



Direct translation

- Analysis of SL only as much as necessary for conversion into particular TL
- Dictionary lookup followed by TL word-for-word output, then TL rearrangement
- Dictionary entries include TL rearrangement rules
- Use of ‘cover’ words
- no analysis of SL syntax or semantics
- output too close to SL structure
- example (Russian to English):
  - On dopisal stranitsu i otložil ručku v storonu.
  - It wrote a page and put off a knob to the side
  - (i.e.) “He finished writing the page and laid his pen aside”
- systems:
  - Univ.Washington, IBM (US)
  - Georgetown University (US)
  - Ramo-Wooldridge (US)
  - Institute for Precision Mechanics and Computer Technology (USSR)
  - National Physical Laboratory (UK)

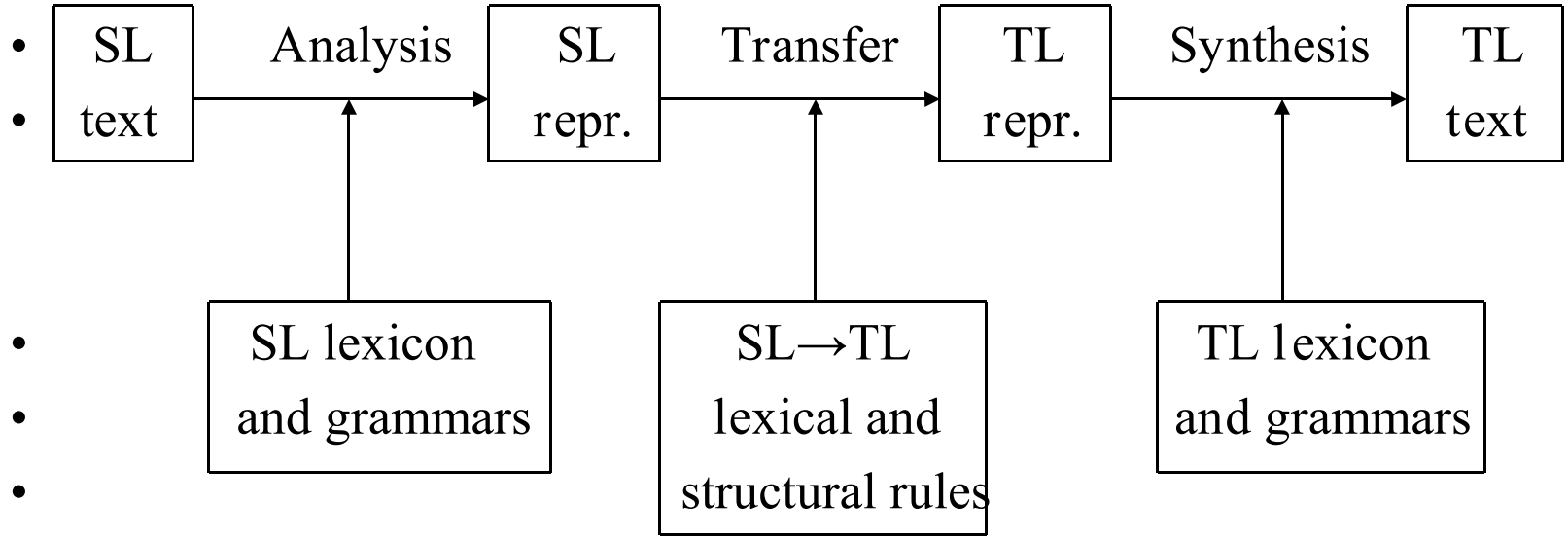
‘Interlingual’ system



Interlingua-based MT

- two independent stages: analysis, synthesis
- abstract language-neutral representation
- multistratal: morphology, syntax, semantics
- semantics-oriented (‘understanding’)
- domain-specific ‘knowledge bases’ (AI-oriented)
- projects:
  - Grenoble (CETA), Texas (METAL)
  - DLT, Rosetta, Pivot (NEC)
  - Carnegie-Mellon University (KBMT, KANT, CATALYST)
  - New Mexico State University (ULTRA, Pangloss)
  - Univ. Maryland (UNITRAN)
  - CICC Japan, China, Thailand, Malaysia
  - United Nations University (UNL)

‘Transfer’ system

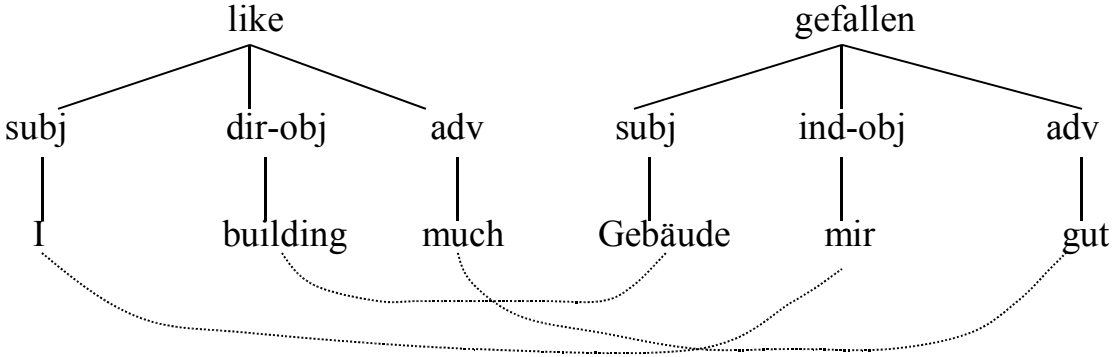


Transfer-based MT

- three stages: analysis, transfer, synthesis
- abstract semantico-syntactic interfaces/representations
- multiple level/strata: morphology, syntax, semantics
- syntax-oriented, tree-transduction
- may use finite-state models for morphology and ‘shallow’ syntax
- batch processing, post-edited
- little/no discourse information (anaphora, etc.)
- projects/systems:
  - GETA-Ariane, Eurotra, LMT, Mu, METAL
  - later/current Systran
  - most current commercial systems are direct/transfer hybrids
  - current research: e.g. interNOSTRUM, Microsoft (Example-based)

Tree transduction

- I like the new building very much ↔ Das neue Gebäude gefällt mir gut



- I like coffee ↔ ich trinke gern Kaffee
- He has just broken his leg ↔ il vient de se casser la jombe

Theories and formalisms

- Information theory
- Finite state grammar
- Transformational-generative grammar
- Dependency grammar
- Stratificational grammar
- Case grammar
- Artificial intelligence
- Lexical-functional grammar
- Generalized phrase-structure grammar
- Definite clause grammar
- Principles and parameters, Government-binding theory
- Categorial grammar
- Montague grammar
- Neural networks

Lexicalist tendency

- Problems of transfer/interlingual multistratal model:
  - complex parsing, transduction, failure = no output
  - complex dictionaries, non-universal case frames, etc.
- preference for shallow parsing
  - simple phrase relations (valencies) rather than case frames and logical (argument) structures
- syntactic and semantic (knowledge-based) constraints in lexical entries
- simple bilingual lexical transfer
- monolingual generation (with sparse structural transfer)
  - ‘shake and bake’ model
- efficient tagging
- efficient (statistics-based) SL lexical disambiguation

Unification grammar:
example (LFG)

- SL f-structure *John likes Mary*
- [PRED like
- SUBJ [PRED John]
- OBJ [PRED Mary]]
- like, V:
- (↑ PRED) = like < SUBJ, OBJ >
- (↑ PRED FR) = plaire < SUBJ, OBJ >
- (τ↑AOBJ OBJ) = τ (SUBJ)
- (τ↑ SUBJ) = τ (OBJ)
- TL f-structure *Marie plaît à Jean*
- [PRED plaire
- SUBJ [PRED Marie]
- AOBJ [OBJ [PRED Jean]]]

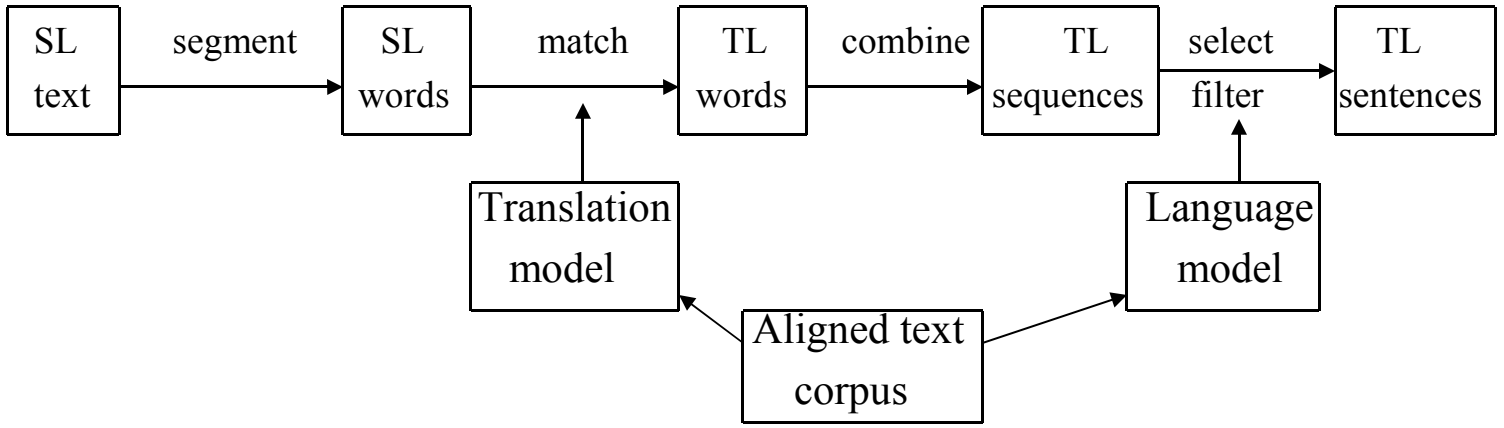
Corpus-based methods

- Not rule-based: in order to circumvent problems of complex grammar rules (analysis, transfer, synthesis), multiple strata, ‘deep’ semantic analysis; complex dictionary entries
- based on bilingual text resources

- Corpus information as source for rules (syntactic or pattern)
  - as earlier RAND project
  - e.g. lexical disambiguation rules, syntactic collocations, patterns/templates
- Extraction of phrases for re-combination [Example-based MT]
- Statistical alignment, word-word frequencies, word co-occurrences [Statistics-based MT]

Statistics-based MT (1)

- 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 (2)

- Based on:
  - 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
- Method:
  - sentences aligned statistically (according to sentence length and position)
  - compute probability that a TL string is the translation of a SL string (based on frequency of co-occurrence in aligned texts of corpus)
    - including ‘fertility’ (how many TL words correspond to SL word)
    - position of SL words in SL 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

Statistics-based MT (3)

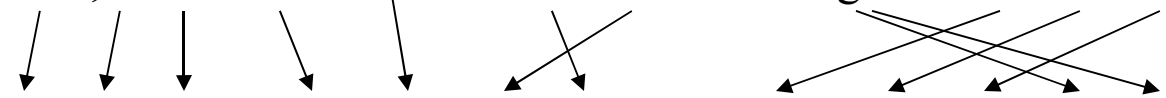
- still insufficient corpora
  - but Internet may solve this
- corpus must be aligned and analyzed before translation of (similar) text in same domain
  - unless large corpus for domain available
- word frequencies not sufficient: *Candide* intended to add morphological information, and some grammatical 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 operational systems
- but also statistics-based translation aids:
  - text prediction (TransType) for translators

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 'translationally equivalent' words
  - cognates: first four letters and 'same' meaning (*mathematics* and *mathématique*)
    - - but fails for *government/gouvernement,* and *actual/actuel*
  - can fail to recognise morphological variants: *book/books, box/boxes, lady/ladies, wife/wives*, etc.
  - by 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

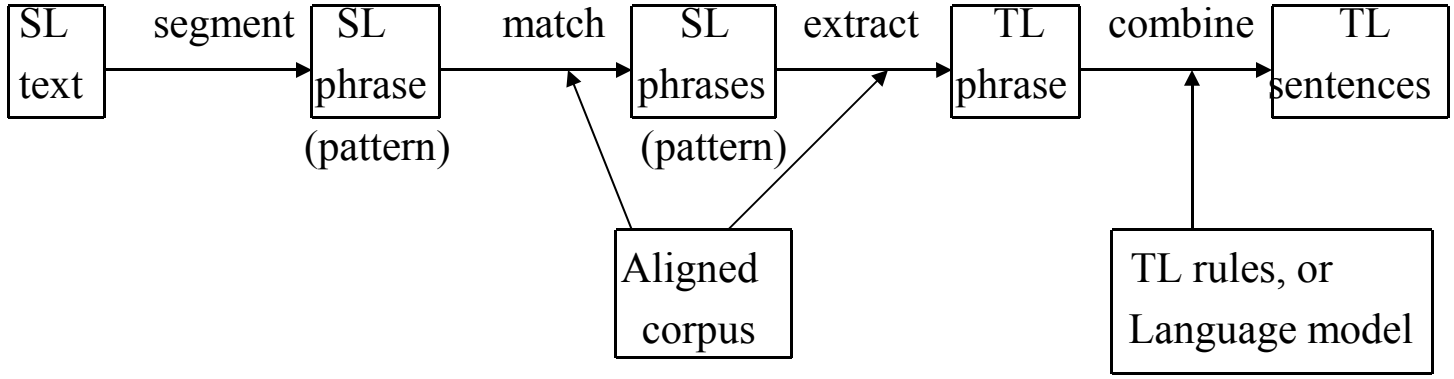


- Difficulties when a SL word group (phrase) corresponds to TL word group
  - yes, then I would say , let us leave it at that.
  - Ja, dann würde ich sagen , verbleiben wir so.

- Problems with inadequate training corpus

Example-based MT (1)

- 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 approaches (2)

- Basis: Use of already translated sentences or phrases either from actual translations (corpus) or from data supplied by user or developer
- Method:
  - Sentences/phrases aligned in database (either by rule-based parser or statistically)
  - matching algorithm (exact and close) of SL input and TL examples
  - combination algorithm (for generating a TL sentence from extracted examples); combination not trivial task

Example-based MT (3)

- bilingual aligned corpora: problems/issues --
  - suitable corpora (as for SMT may be extracted from Internet)
  - 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 (training, extracted data)
- but, nearly all use also bilingual dictionaries, tagging/parsing systems, etc.
- similarity comparisons often employ a thesaurus ('knowledge base' of wrds in semantic fields).
- may also be domain-specific (e.g. speech systems based on EBMT approach)

Example-based MT (4)

- Basic approaches
  - for SL phrase, find a similar SL phrase and its TL equivalent, and insert words
  - for SL phrase, find two (overlapping) SL phrases and equivalents, and combine them
  - for SL phrase, find its template, and insert TL words (phrases)
- use of grammatical categories (patterns): transfer-based
  - templates: <1st name><family name> flew to <city> on <date>
  - grammar patterns: X [pron] eats Y [noun/NP] ↔ X [pron] ga Y [noun/NP] o taberu
  - phrase patterns: 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.)
- problem of matching by characters when searching:
  - 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)
- problem of 'boundary friction' when combining:
  - 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

Example-based MT: boundary friction

- Examples in database:
- (1e) The obstinate man refused all help
- (1g) Der hartnäckige Mann hat alle Hilfe verweigert
- (2e) Help was rejected by the stubborn man
- (2g) Hilfe wurde von dem starrköpfigen Mann abgewiesen

- sentence to be translated:
- (3e) Help was rejected by the obstinate man
- matching fails to relate verweigern and abweisen, and hartnäckig and starrköpfig
- problem of boundary friction if example for 'obstinate man' inserted into (2g):
- (3g) * Hilfe wurde von der hartnäckige Mann abgewiesen.

- **In general, morphological variation handled more easily by rule-based systems than by corpus-based systems**

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

- **Bilingual lexical ambiguity solved (?) by contextual rules (RBMT), or examples (EBMT), or word-word frequencies and 'language models' (SMT), but perennial problem whatever the methods**
- **lexical gaps are virtually unsolvable by any method**

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 receive a state pension
- (5b) 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)**
- **In practice (for 'information' only purposes), ambiguities may be bearable as the intention can be understood from wider context**

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 (tree-transduction) 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 (*ihr* usually translated by *her*)
- 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 where 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 (no sign of one yet!)**

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]

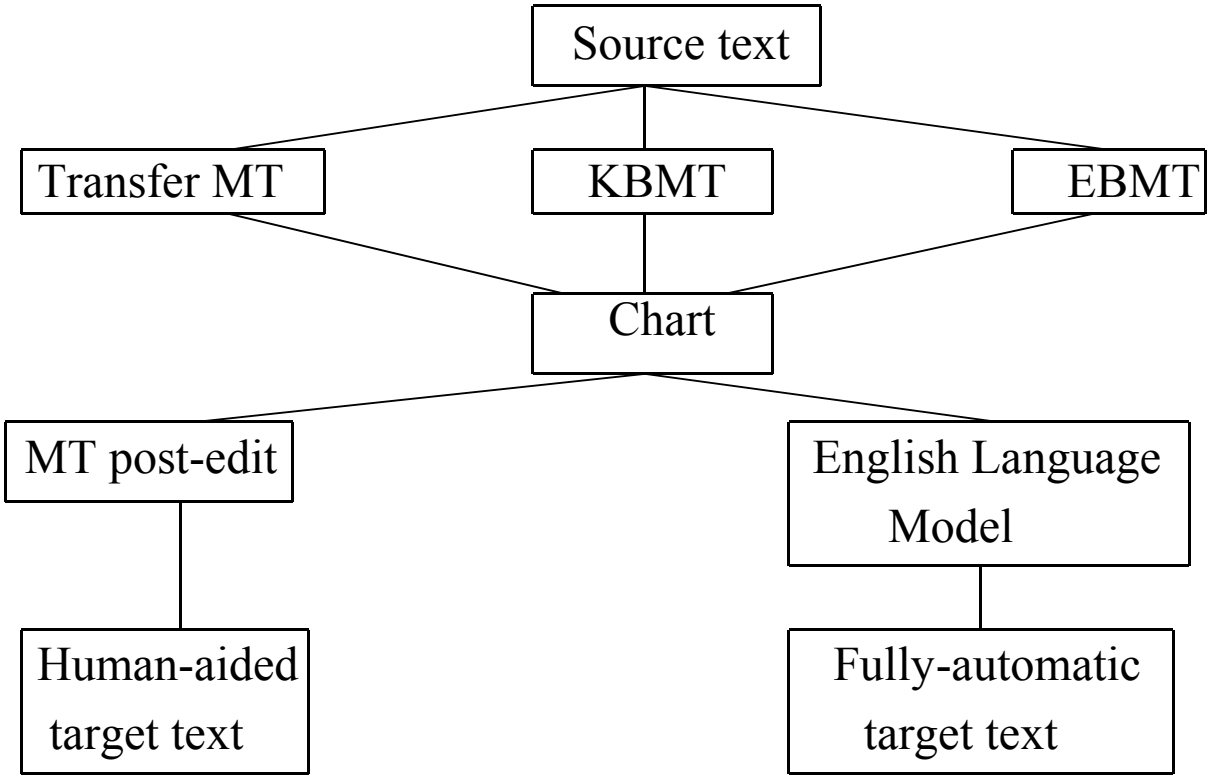
- [English] Advances in technology created new opportunities
- [Japanese] Because technology has advanced, opportunities have been created
- [or Japanese] Technology has advanced. There are new opportunities.

- **All methods of MT tend to retain SL structural features; however, theoretically SMT 'language model' approach could 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:**

Hybrid systems:
an example (Pangloss Mark III)



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)

MT and CAT in practice

- Demand:
  - dissemination: external publications, internal reports, operational manuals, localization
  - assimilation: internal ‘monitoring’, information ‘filtering’
  - interchange: correspondence, email, telephone
  - information access: information retrieval/extraction, web pages
- Methods
  - MT with human assistance (pre-editing, controlled language, post-editing)
  - translation aids (translation workstation, translation memories)
  - raw unedited MT (on-line, real-time)

Issues for corporations

- MT or translation aids (TM)
- General-purpose system or specialised system
- Controlled language (existing or developed in-house)
- Lexical resources (creation, maintenance)
- Translation memories (creation, use, maintenance)
- Management implications
- Control of terminology
- Consistency
- Standards; exchange formats
- Compatibility (hardware, software)
- Integration: technical authoring, publishing

Translation memories: problems

- Expensive to build in time and money
- Loss of context (beyond sentence), e.g. domain of document
  - translators may need to translate whole
- Redundant entries, ambiguous entries, conflicting examples
- No ‘learning’ from adaptations made by translators
- Combining of extracts made by translators [need semi-automation]
- Fuzzy matching ineffective
  - occurrence of different (hidden) formatting tags
  - recombination of fuzzy matches is longer than translation from scratch
- Whole sentence repetition is rare
  - limited value for administrative documents, minutes of meetings, marketing texts, most reports, web sites
- **Repetition of phrases and clauses much more common. Therefore need ‘example-based’ approach**

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: problems

- 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)

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)

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 -- simple clauses only possible
- 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

- **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**

- **improvements by use of SMT 'language model' for TL output?**

French-English (Lycos-Systran)

- M. le Président rappelle que le problème de la réduction du temps de travail a été étudié à la réunion de Munich. Différentes thèses s'affrontent: pour les syndicats, la réduction du temps de travail contribuera à supprimer le chômage, mais les employeurs pensent qu'elle supprimera des emplois soit en renchérissant les coûts, soit en accroissant la productivité. Il serait souhaitable de poursuivre aujourd'hui cette discussion en laissant de côté tous les présupposés idéologiques. Pour commencer, il convient de demander au représentant de la Commission européenne, qui a réussi à rejoindre Luxembourg malgré un mouvement de grève à Bruxelles, s'il souhaite compléter l'exposé qu'il avait présenté à Munich.

- Mr. the President points out that the problem of the reduction of the working time was studied with the meeting of Munich. Various theses clash: for the trade unions, the reduction of the working time will contribute to remove unemployment, but the employers think that it will remove employment either by increasing the costs, or by increasing the productivity. It would be desirable to continue this discussion today by leaving side all the presupposed ideological ones. To start, it is advisable to ask to the European Commission representative, which succeeded in joining Luxembourg in spite of a movement of strike in Brussels, if it wishes to supplement the talk that it had presented in Munich.

English-French (FreeTranslation)

- Position of nutritional research in the Federal Republic of Germany. This article contains a list of gaps in the nutritional research programmes of various research and industrial Federal German establishments prepared by a committee of the Ministries for Youth, Families and Health and the Ministry for Nutrition, Agriculture and Forestry. The topics are classified into food technology, food chemistry, food microbiology and hygiene, and nutrition medicine and physiology.

- La position de recherche de nutritional dans la République Fédérale d'Allemagne. Cet article contient une liste d'écarts dans les nutritional recherche émissions de divers recherche et industriel Fédéral Allemand établissements préparés par un comité des Ministères pour Jeunesse, Familles et Santé et le Ministère pour Nutrition, Agriculture et Forestry. de des la les la la la. Les sujets sont classifiés dans la technologie de nourriture, la chimie de nourriture, microbiology de nourriture et l'hygiène, et le médicament de nutrition et la physiologie.

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 and information analysis and extraction: requirements

- tasks for information analysis/filtering tasks
  - should be fully automated, with no pre- or post-editing
  - tuned for specific domains
  - should accept OCR input
  - should tolerate (and ideally correct) misspellings, missing diacritics, wrong transliteration, grammar mistakes, scanning errors
  - deal with mix of languages in same document
  - identify and retain all formatted information
  - provide facilities for easy updating of lexicon
  - specialist lexica for different domains
- additional tasks for information extraction
  - domain (scenario) templates for SL; presentation of completed template in TL
- additional tasks for 'translingual speech retrieval' (browsing radio broadcasts, information routing, automatic alerts)
  - generalised speech recognition
  - word detection; indexing of key terms

Some future directions and expectations (1)

- merging of MT and TM for enterprise dissemination systems
- data-driven vs. (and) theory-driven -- hybrid systems
- Internet as resource
- rapid development of systems
  - particularly for assimilation/interchange
- improvements in quality
  - particularly PC commercial and online systems
- special-purpose systems (domain and function)
  - particularly on Internet (but will no longer be free!)

- Reusability of resources (particularly dictionaries and translation memories)

Some future directions and expectations (2)

- Spoken language translation
- '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
- Further integration with other NLP systems
  - MT as option with summarization, information extraction, information retrieval, data retrieval, question-answering, Internet search tools
- bilingual (multilingual) communication as much as translation