A New Era in Machine Translation Research

John Hutchins

University of East Anglia
Norwich, England
Outline

- Rule-based MT
  - transfer, interlingua
- Lexicalist tendency
- New formalisms
- General-purpose NLP systems
- Lexicon creation
- Corpus-based MT
  - statistics-based MT
  - example-based MT
  - bilingual corpora
- Connectionist MT
- Text generation
- Controlled language MT
- Domain- and user-specific MT
- Eras of MT
- "Third generation" systems
- Use of MT systems
Formalisms

Constraint-based formalisms
Unification grammars

Lexical-Functional Grammar
Definire Clause Grammar
Generalized Phrase Structure Grammar
Head-driven Phrase Structure Grammar
Principles-based MT

Reversibility

General-purpose NLP systems
CLE (Core Language Engine)
PLNLP (Programming Language for Natural Language Processing)
ELU (Environnement Linguistique d'Unification)

Lexicon construction
grammatical information
syntactic and semantic constraints
non-linguistic information
lexical acquisition
collaboration (e.g. EDR)
Bilingual corpora

Text alignment

Translation memories
(translation workstations)

Connectionism, parallel processing
(`learning' systems)
Carnegie-Mellon, UMIST, Matsushita

Post-editing feedback
Tovna, MAPTRAN, PIVOT (NEC)
Use of MT and the future

increasing usage
  (more than one million pages per annum)
  - multinational companies
  - translation agencies

non-professional usage
  cheap PC-based packages
    (Globalink, Microtac, PC-Translator)

MT on networks
  Minitel, PC-VAN, Niftyserve, CompuServe
  Systran, Logos

Future system types:
  for translation services/agencies
  for non-professional translation
  for information gathering
  for monolinguals
  for spoken language

  domain-specific, sublanguage
  user-specific, custom-built
  controlled input