

**UFSC / PPGEP / UNL**

---

***Tradução Automática:  
Superando as Barreiras entre  
Línguas Européias e Chinesas***

**Maio, 2002**

**Paper 4**

**Irineu Theiss**

---

# Automatic Translation: Overcome Barriers between European and Chinese Languages

---

WONG Fai, MAO Yuhang, DONG QingFu, QI YiHong  
Tsinghua University (China)

Speech and Language Processing Research Center,  
Tsinghua Univ., Beijing 100084, P.R. China

## ABSTRACT

This paper describes the differences between European languages and Chinese language, the urgency and possibility of realizing machine translation between the languages of Chinese and Latin's. Three different levels of realizing machine translation are enumerated: online dictionary, word-to-word translation and text-to-text translation. The difficulties and the ways to realize machine translation are studied and analyzed. It is pointed out that the rule-based machine translation is most promising and feasible.

**Keywords:** machine translation, online dictionary, word segmentation, and ambiguity resolution

---

# Introdução

---

- A comunicação entre pessoas que falam línguas diferentes depende de tradução
- A tradução em tempo real (simultânea) nem sempre é possível e seu custo pode ser proibitivo

**A TRADUÇÃO AUTOMÁTICA se torna atrativa por ser por ser rápida e barata**

---

# Intercâmbio China X Europa

---

- Observa-se um grande crescimento no intercâmbio cultural, científico, tecnológico e comercial
- A eficiência desse intercâmbio sofre o impacto das diferenças linguísticas
- É urgente construir um “tradutor automático” para facilitar o intercâmbio

**MAS...**

---

## ... há dificuldades naturais

---

- Transferir o conhecimento associado a diferentes línguas para o computador não é tarefa fácil
  - Entre línguas européias, por exemplo, existe muita similaridade:
    - projetar o tradutor automático para essas línguas foi fácil, mas executá-lo e colocar no mercado é trabalho de décadas
    - o sistema SYSTRAN é um exemplo de sucesso
-

# Diferenças Linguísticas

---

- Para línguas de origens diferentes, há grandes diferenças na estrutura de vocabulário e gramática
  - É o caso do Chinês (mandarim ?) em relação às línguas ocidentais
  - É fundamental levar em conta tais diferenças ao projetar um Tradutor Automático
-

# Principais Características

---

- Línguas Europeias: milhares de palavras construídas com um alfabeto limitado
  - **Mandarim: palavras são feitas com milhares de caracteres**
  - Numa frase, as palavras são separadas por espaços
  - **Na sentença, não há espaço entre palavras e caracteres**
-

# Principais Características (cont.)

---

- Ocorrem mudanças morfológicas das palavras
  - Não há mudanças morfológicas
  - Há diferença de gênero nos nomes
  - Não há diferença de gênero
-

# Principais Características (cont.)

---

- Línguas Europeias: usa-se artigos diante dos nomes numa frase
  - **Mandarim: pouco uso de artigos**
  - Advérbio de tempo e lugar no final da frase
  - **Advérbio de tempo e lugar no início da frase**
  - Pouco uso de classificadores para modificar nomes
  - **Usualmente, nomes são modificados através de classificadores**
-

# Principais Características (cont.)

---

- Modificadores de sujeito e objeto vêm depois
  - Normalmente vêm antes
  - Expressão de tempo e voz é explícita
  - Expressão de tempo e voz implícita
-

# Diferentes Níveis de Tradução Automática

---

- Abordagem de acordo com a forma e precisão requerida
  - Dicionário on-line :
    - para buscar a tradução de uma simples palavra
    - não é suficiente para traduzir um texto
-

# Diferentes Níveis de Tradução Automática

---

## ➤ Tradução palavra-a-palavra:

- essa abordagem permite uma tradução expedita de uma frase, um parágrafo ou mesmo um texto
  - faz apenas a análise morfológica (verbo, gênero, número)
  - não faz análise sintática
  - em alguns casos, o resultado da tradução pode ficar incompreensível
-

# Diferentes Níveis de Tradução Automática

---

## ➤ Tradução palavra-a-palavra (cont.):

- na tradução do mandarim para uma língua europeia, pode ser usado um algoritmo para segmentar as palavras chinesas. Numa frase ambígua, se a segmentação não for apropriada pode-se chegar a uma tradução errônea (ou falsa interpretação da frase)
  - esse tipo de abordagem (palavra-a-palavra) pode ser usado para traduzir termos comerciais ou ligados à tecnologia
-

# Diferentes Níveis de Tradução Automática

---

## ➤ Tradução texto-a-texto:

- o maior desafio da tradução via computador é traduzir uma frase inteira
  - o sistema deve fazer não só a análise morfológica, mas também a análise semântica e a análise contextual
  - projetar a base de conhecimento é a tarefa fundamental
  - essa base conterá não só informações sintáticas, mas também semânticas
-

# Diferentes Níveis de Tradução Automática

---

## ➤ Tradução texto-a-texto (cont.):

- A resolução de ambigüidades é outra questão importante
  - Para a tradução via computador é absolutamente necessário descobrir os padrões de formação das frases nas diferentes linguagens
  - Em resumo, vários processos são requeridos: segmentação de palavras, análise morfológica, resolução de ambigüidade, construção de frases, compatibilização de padrões de sentenças e geração da língua alvo
  - O modelo de tradução baseado em regras é o mais promissor
-

# Sistemas já Desenvolvidos

---

- Sistema de tradução Chinês para Inglês
    - A tecnologia-chave nesse sistema é a segmentação de palavras para o Mandarim e a resolução de ambigüidade para frases ambíguas
-



# Sistemas já Desenvolvidos

- Sistema de tradução Inglês para Mandarim



# Sistemas já Desenvolvidos

---

- PCT 2.1 – Português/Chinês – word-by-word
  - MACAU – coexistência do Chinês com o Português – há grande demanda de tradução de documentos oficiais do Português para Chinês, pois a maioria dos documentos oficiais e Leis estão na Língua Portuguesa
  - Projeto de pesquisa conjunto da Universidade de MACAU e da Universidade Tsinghua
-

# Sistemas já Desenvolvidos

- PCT 2.1 – Português/Chinês – word-by-word



Figure 3 PCT 2.1 Main Interface

# Sistemas já Desenvolvidos

- PCT 2.0 – Português/Chinês – sentença por sentença – em desenvolvimento

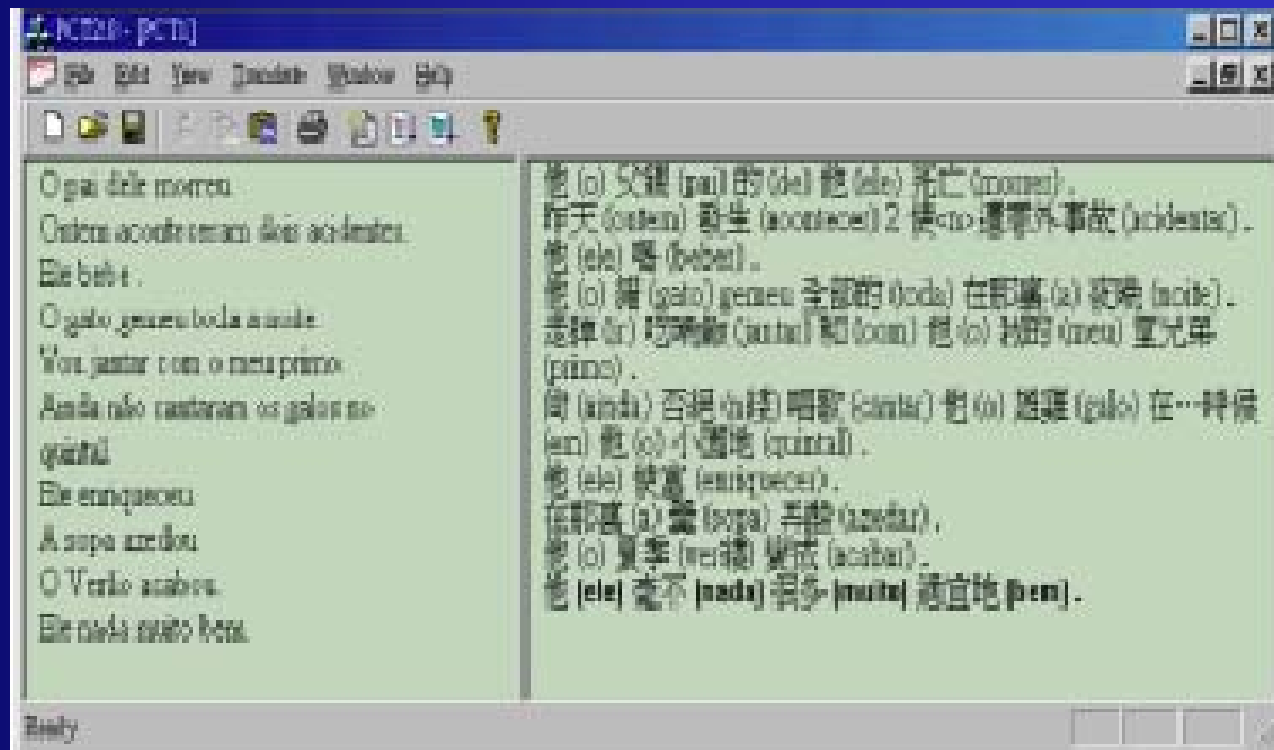


Figure 4 PCT 2.0 Translation System

# Sistemas já Desenvolvidos

---

- FET 1.0 – Francês/Chinês – word-by-word
    - Projeto conjunto com o Ministério da Educação da França
  - CFMT 1.5 – Chinês/Francês – sentença por sentença – projeto em desenvolvimento
-

# Conclusão

---

- Há uma demanda urgente por ferramentas para tradução, em função do explosivo crescimento da Internet
  - Com a abertura econômica da China, cresce o número de negócios com empresas estrangeiras
-

# Conclusão

---

- Tradução via computador é assunto do campo da Inteligência Artificial
  - Perguntas que precisam ser respondidas:
    - Como é o processo de geração da linguagem na mente humana?
-

# Conclusão

---

- Se for uma espécie de database, então o pensamento e o conhecimento humano são infinitos
  - Mas se é baseado em regras, então podemos dizer que as regras para formar sentenças a partir de palavras e frases são finitas
-

# Conclusão

---

- O que precisa ser feito é preparar um bom dicionário e um conjunto de regras para as línguas fonte e destino
  - E, aplicando algoritmos, será possível criar sistemas mais eficazes de tradução via computador
-

Perguntas ?

---

**Obrigado !**

---