

Tradução automática: a tecnologia e sua aplicação

Diana Santos
Primeira Escola de Verão da Linguateca

Mas qual é o problema?

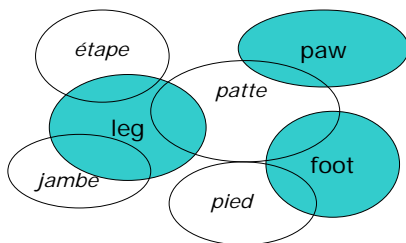
- Duas línguas
- Passar uma mensagem com um dado sentido e com uma dada audiência (explícita ou implícita) usando um dado sistema
 - Para outra
- A comunicação é uma arte, é uma técnica, e é uma ciência

11-08-2006

Tradução automática

2

Questões de sobreposição lexical



Jurafsky & Martin (2000:806)

11-08-2006

Tradução automática

3

É preciso saber... para traduzir

- *dangerous cyanide and chlorine fumes*
- *carbon and nitrogen tetraoxyde*

Boitet 1984

- fumos de cianido são perigosos
- não existe tetraóxido de carbono

11-08-2006

Tradução automática

4

É preciso sentir... para traduzir

- I'll come tomorrow
 - asu oukagai itashimasu
 - asu oukagai shimasu
 - asu ikuyo
 - asu ikuwa
 - asu ikimasu
- posição social, relação de intimidade entre os intervenientes, sexo do falante

Tsujii 1986

11-08-2006

Tradução automática

5

A minha opinião sobre a tradução

- A estrutura da frase e a organização do texto não são preservadas
- Os tradutores não preferem tradução literal
- A tradução literal não é sem problemas
- A tradução exige um processamento significativo dos textos
- Cada frase não tem só um sentido
- Um modelo de tradução deve contar com ruído (Santos, 2000)

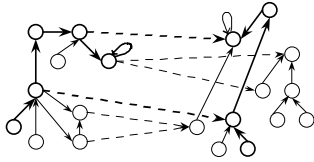
11-08-2006

Tradução automática

6

O modelo da rede de tradução

- Descrição independente das duas línguas
- Pontes entre diferentes categorias



11-08-2006

Tradução automática

7

Rede de tradução

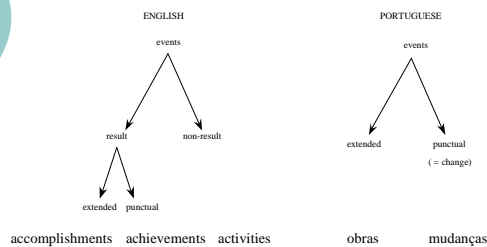
- Categorias diferentes em línguas diferentes
- “Falhas aspectuais” (aspectual gaps)
- Vagueza (vários pedaços de significado associados a um mesmo “operador”)
 - Nem todos podem ser traduzidos simultaneamente...

11-08-2006

Tradução automática

8

Exemplo concreto: classes aspectuais

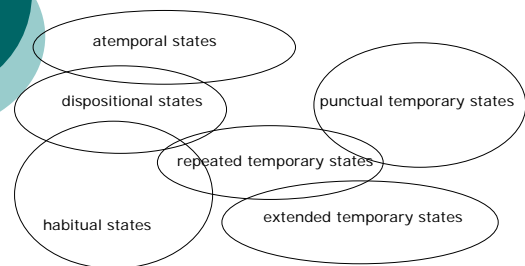


11-08-2006

Tradução automática

9

Exemplo concreto: estados



11-08-2006

Tradução automática

10

Exemplo concreto: falha aspectual

- Àquele “sentido” não é possível associar o mesmo tipo de padrão aspectual nas duas línguas
 - *Ficou na estrada a vê-los afastarem-se*
 - *he stood in the road watching them leave*
 - *In his house K. squatted on his sleeping mat, brooding.*
 - *E K. em casa, acocorado na esteira, meditava longamente.*

11-08-2006

Tradução automática

11

Exemplo: duração temporal

- *Kino's people had sung of everything that happened or existed.*
- *A gente de Kino cantara tudo o que acontecera ou existira*

passado presente
- *He was trapped as his people were always trapped*
- *Estava peado, como todos os da sua raça sempre tinham estado*

11-08-2006

Tradução automática

12

Exemplo prático

- Tradução de *sit* (e seus associados) no COMPARA
- *I saw him sit there all afternoon*
- *No prédio fronteiro, viu o calafate sentado à mesa*

11-08-2006

Tradução automática

13

O que é o COMPARA?

- O maior corpus paralelo revisto do mundo
- Textos literários traduzidos inglês-português
- Em permanente extensão e revisão
- Alinhado à frase
- Classificado sintacticamente
- Será muito badalado na presente Escola

11-08-2006

Tradução automática

14

Breve historial

- Anos 40-50 muita agitação, a aplicação
- 1964: Relatório ALPAC (AMTCL -> ACL)
- Sistemas comerciais
- IA vs. linguística (CMU vs Eurotra)
- 1990: Sato & Nagao: Memory-based translation [Nagao 1994]
- 1990: Brown et al: TA baseada em estatística
- 1993-2000: Verbmobil

11-08-2006

Tradução automática

15

Mais história

- 1988: Interlíngua=esperanto, dupla transferência
- 1990: modelo para tratar EVPs em TA
- 1991: Gale & Church, alinhamento
- 1992: Pierre Isabelle: bitextos para tradutores
- Anos 90 “translation browsers”, varredores de tradução

11-08-2006

Tradução automática

16

Tipos de escolas de TA

- Transferência (codifica regras de tradução)
- Interlíngua (tenta chegar ao sentido)
- Tradução directa (transforma o texto)
- Tradução baseada em exemplos (analogia)
- Métodos:
 - simbólicos (regras)
 - estatísticos (parâmetros estimados em corpora)

11-08-2006

Tradução automática

17

Formas de ver a TA (Nigel)

- Interessadas no processo: como fazer a tradução (directa, transf, interl)
- Interessadas no resultado: quais as medidas a maximizar, dados os requisitos (estatística)
 - argmax_T (fluência (T) x fidelidade (S,T))
- Maior respeito pela tradução humana?

11-08-2006

Tradução automática

18

Tradução baseada em exemplos

- Pressupõe um conjunto de exemplos já armazenados (ou uma forma de os obter)
- Problema: escolher qual o exemplo a usar, usando uma medida de semelhança
 - usando tesouros
 - usando corpora bilingues

The man eats vegetables - come, alimenta-se de
Acid eats metal - ataca, corrói

11-08-2006

Tradução automática

19

Tradução com factor de fiabilidade

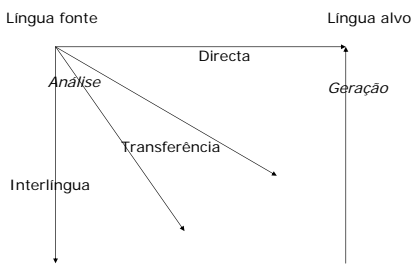
- Invocado por Sumita & Ida (1991) como uma vantagem de tradução baseada em exemplos comparada com a baseada em regras
- Factor de fiabilidade também dependente de conhecimento humano (traduções prévias)

11-08-2006

Tradução automática

20

Uma visão gráfica

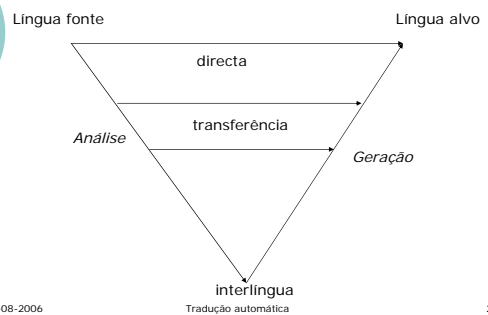


11-08-2006

Tradução automática

21

A visão gráfica tradicional (errada)



11-08-2006

Tradução automática

22

TA usando métodos estatísticos

- Todas as frases de uma língua são possíveis traduções de uma frase noutra língua (as probabilidades é que variam): $\text{Prob}(T|S)$
- Dada uma tradução T , procurar S que maximiza $\text{Prob}(S|T) = \text{Pr}(S) * \text{Prob}(T|S) / \text{Prob}(T)$
- $\text{Prob}(le\ chien\ est\ battu\ par\ Jean|John\ (6)\ does\ beat\ (3,4)\ the\ (1)\ dog\ (2)) = \text{Prob}(fert=1|John) * \text{Prob}(Jean|John) * \text{Prob}(fert=0|does) * \text{Prob}(fert=2|beat) * \text{Prob}(est|beat) * \text{Prob}(battu|beat) * \dots$

11-08-2006

Tradução automática

23

Aplicações associadas à tradução

- alinhamento
- desambiguação
 - escolha
- correcção
- sugestão
- memórias de tradução
- recolha de informação cruzada
- procura/acesso a corpora paralelos

11-08-2006

Tradução automática

24

Recursos associados à tradução

- dicionários bilíngues
- corpora paralelos
- regras de tradução
- memórias de tradução
- corpora comparáveis

Algumas dicotomias que podem não fazer sentido na tradução

- sintaxe vs. semântica
- “knowledge-lite” vs. “knowledge-heavy”
- automático vs. humano
- teoria vs. prática
- quantitativo vs. qualitativo
- estatístico vs. linguístico
- conhecimento do mundo vs. conhecimento linguístico

Algumas dicotomias que podem não fazer sentido na tradução (cont.)

- léxico vs. sintaxe
- texto técnico vs. texto literário
- termos vs. palavras da linguagem geral

O factor humano na TA

- As pessoas não gostam de ajudar o computador
 - tradução interactiva não funciona
- As pessoas gostam de corrigir o computador
 - pós-edição e conseqüente responsabilidade final