

# Tradução Automática Baseada em Exemplos

Alberto Manuel Brandão Simões  
ambs@di.uminho.pt

Orientação

José João Almeida  
Diana Santos

Casa da Torre — Vila Verde  
6 de Janeiro de 2005



- Não consensual!!
- Sistema de tradução que:
  - se baseia em traduções já realizadas (exemplos);
  - utiliza/pesquisa os exemplos em tempo-real;
- Na verdade, também:
  - usa regras de conciliação de traduções;

matematicamente...

$translate : sentence \times tmdb \longrightarrow sentence$

$translate(s, db) \stackrel{\text{def}}{=}$

let  $I1 = split(s)$

$I2 = \langle match(db, x) \mid x \in I1 \rangle$

in  $recombine(I2)$



- Armazena:
  - pares de segmentos (unidades de tradução);
  - pares de padrões (segmentos com place-holders);
  - dicionário probabilístico de tradução;

matematicamente...

$$\begin{aligned} tmdb &\cong \begin{aligned} &transMemory : tm \times \\ &transPattern : tp \times \\ &dictionary : dic \end{aligned} \\ tm &\cong alignUnit \overset{m}{\leftrightarrow} alignUnit \\ tp &\cong pattern \overset{m}{\leftrightarrow} pattern \\ dic &\cong word \rightarrow wordInfo \\ wordInfo &\cong \begin{aligned} &trans : word \rightarrow real \times \\ &occur : int \end{aligned} \end{aligned}$$



Como é...

- pouco provável que a frase já esteja traduzida;
- provável que porções da frase já estejam traduzidas;

tenta-se generalizar...

- criação de classes de palavras  
(palavras usadas em contextos semelhantes)
- reconhecimento de entidades mencionadas  
(substituição de nomes próprios por placeholders)
- etiquetação de PoS  
(classificação de categorias morfológicas)
- detecção de partes semelhantes em frases  
(criação de “gramáticas” de exemplos)

Mais uma vez, como é...

- pouco provável que a frase já esteja traduzida;
- provável que porções da frase já estejam traduzidas;

temos de fazer pesquisas por segmentos “pequenos”...

- detecção de segmentos usando:
  - chunkers;
  - informação sintáctica;
  - detecção de termos e tempos verbais compostos;

matematicamente...

$split : sentence \rightarrow segment^*$

$split(s) \stackrel{\text{def}}{=} \dots$

...

- dado um segmento:
  - procurar exemplo semelhante;
  - procurar padrão;
  - traduzir palavra a palavra;
- calcular “fiabilidade” da tradução obtida:
  - se baixa, retornar ao módulo de segmentação

matematicamente

$match : tmdb \times segment \longrightarrow segment$

$match(db, s) \stackrel{\text{def}}{=} \dots$

...

- baseado em:
  - conjunto de regras;
  - analisador morfológico;
- regras inferidas usando:
  - corpora etiquetado;
  - florestas sintáticas;

matematicamente

$recombine : segment^* \longrightarrow sentence$

$recombine(s) \stackrel{\text{def}}{=} \dots$

...



- Construção de:
  - um protótipo de sistema de EBMT;
    - prototipagem dos vários componentes;
    - várias aproximações/técnicas;
  - protótipos de ferramentas de generalização:
    - avaliação/comparação das várias técnicas;
    - avaliação/comparação de técnica vs língua;
  
- Propostas de metodologias de avaliação:
  - para sistemas de EBMT;
  - para os vários componentes;





*align* : *corporaparelos*  $\longrightarrow$  *dic*

```
enorme => {  
  count => 45,  
  trans => {  
    large      => 0.417,  
    huge       => 0.227,  
    enormous   => 0.132,  
    (null)     => 0.107,  
    deep       => 0.034,  
    flinging   => 0.030,  
    fasten     => 0.021,  
    pheasants  => 0.011,  
  },  
},  
coruja => {  
  count => 36,  
  trans => {  
    owl      => 0.968,  
    vacuum    => 0.020,  
    forward   => 0.010,  
  },  
},
```



## Usando

- detector de unidades multi-palavra;
- alinhador à palavra;

## Obter

please wait	-> favor aguarde
bug report	-> relatório
to create	-> para criar
recent files	-> ficheiros recentes
fishing boat	-> barco de pesca
regular expression	-> expressão regular

Associar uma “probabilidade” de tradução a um par de frases.

$Q$	English	Portuguese
0.261751	vocês os dois - eu mando-vos uma coruja .	Thanks , said Harry , I ' ll need something to look forward to .
0.673549	"Have you got your own broom ?	- Tens a tua vassoura ?
0.659810	"Harry glanced down at his broom .	Harry deitou um olhar à sua vassoura .
0.599914	"Stick out your right hand over your broom , "called Madam Hooch at the front , "and say ' Up !	- Estendam a vossa mão direita ao longo da vassoura -, gritou Madame Hooch , - e digam De pé .



==> difficult situation

Using 6 occurrences (0.732864 seconds)

situation difficile - 0.8025

situation très difficile - 0.8025

situation aussi difficile - 0.8025

==> sentenced to death

Using 1 occurrences (0.214145 seconds)

condamné à mort - 0.4433333333333333

==> final version

Using 7 occurrences (0.843922 seconds)

version définitive - 0.5075

définitive - 0.09

définitif - 0.0875

no parlamento europeu existe uma grande crise de identidade

no parlamento europeu...

    european parliament - 62%

existe uma grande...

    there was a great - 37%

existe uma...

    there exists a - 51%

grande crise de...

    crisis here - 34%

grande crise...

    crisis - 60%

de identidade...

    identity - 73%

european parliament there exists a crisis identity



- alinhamento de corpora etiquetado:
  - EMS (para PT);
  - TnT (para EN);
- alinhamento de corpora com entidades mencionadas anotadas;
  - RENA (REM do projecto IKF);