

## Tradução Automática

Belinda Maia  
Universidade do Porto  
&  
PoloCLUP Linguateca

## A importância da Tradução Automática (TA)

- Importância social – necessidade de comunicação entre pessoas sem língua comum
- Importância política - multilinguismo
- Importância comercial e económica
- Interesse científico e filosófico

## Tradução e a Máquina - distinções

- MAHT > THAM – Tradução Humana Assistida pela Máquina
- CAT > TAC - Tradução Assistida pelo Computador
- HAMT > TAAH – Tradução Automática Assistida pelo ser Humano

## THAM – Tradução Humana Assistida pela Máquina

- Descrição mais vaga
- Implica utilização de várias tecnologias para facilitar a tradução
- Processamento de texto
- Armazenamento de textos e terminologia
- Recurso à informação na Internet

## TAC - Tradução Assistida pelo Computador

- “CAT tools” - SDL/TRADOS, STAR TRANSIT, DÉJÀ VU, WORDFAST etc que incluem
  - Memórias de Tradução
  - Terminologia
  - Extractores de texto
  - ... TA

## TAC - Localização

- = Tradução de manuais, documentação e a parte linguística de programação informática
- Programas de Localização – PASSOLO, CATALYST etc
- Teoricamente funcionalista - ou adaptação do texto etc às necessidades LOCAIS
- Na prática > uniformização de texto e tradução

## Programas de TAC e Localização

- Inventados por engenheiros impacientes com a lentidão de tradução humana
- Aproveitam textos e traduções já existentes > reduzem custos
- Garantem uniformidade de terminologia e expressões comuns
- Oferecem material para melhorar TA

## TA – usos actuais

- Tradução 'Gist' –ex. na Comissão Europeia
- Textos efémeros + utilizadores tolerantes > TA on-line
- TA em domínios específicos + léxicos especializados e sintaxe simples
- Programas específicos entre L1 < > L2

## TA com assistência humana

- Estrutura linear de frases
- Pré-edição de texto
- 'Linguagem controlada'
- Léxicos em domínios específicos
- Pós-edição
- Trabalhando com a 'caixa de vidro'
- Acesso limitado à 'caixa negra' – esp. léxicos

## A TA e o Tradutor Humano

- Os tradutores humanos tendem a
  - desprezar a TA
  - ter medo da TA
- Os tradutores nunca são convidados a avaliar a TA!
- MAS
- A maior ameaça à tradução profissional não é a TA
- É o INGLÊS!

## A TA em linha e o utilizador

- Facilita comunicação internacional – e a vários níveis = mais democrática
- Cria vontade de conhecer textos em línguas desconhecidas = mais tradução humana
- TA em domínios específicos ou preparada para fins bem-definidos – acelera o trabalho de tradução
- TA – entre certas línguas - utilizada cada vez mais por tradutores humanos

## Ferramentas desenvolvidas pela Linguateca

- METRA
- EVAL
- TrAva
- CORTA
- BOOMERANG
- ..... Pesquisa dos logs do METRA

## METRA

- <http://www.linguateca.pt/metra>
- Meta-ferramenta de TA
- Utiliza vários motores de TA disponíveis grátis em linha
- EN > PT
- PT > EN



## METRA3

Tradução de "Linguateca's Summer School will take place at the Faculdade de Letras between 10 and 14 July."

E qual a melhor tradução ao sua opinião ?

- A escola de Verão de Linguateca ocorrerá no Faculdade de Letras entre 10 e 14 julho. FreeTranslation
- Escola de Verão do Linguateca acontecerá no Faculdade de Letras entre 10 e 14 julho. Systran
- A escola de verão de Linguateca ocorrerá no Faculdade de Letras entre 10 e 14 julho. E T Server
- Linguateca acontecerá ao de de Faculdade Letras entre 10 e 14 julho. Amikai
- Nenhuma destas traduções

Atualmente o METRA3 utiliza vários motores de tradução: (Babelfish / FreeTranslation / WordLingo / Google / E-Translation/Amikai / Amikai)  
NOTA: Estes serviços são usados livremente disponibilizados de terceiros de tradução automática comercial que não são responsáveis da Linguateca. Alguns dos motores de tradução podem estar temporariamente indisponíveis.

METRA v3 - Fevereiro de 2006  
Filo da Faculdade de Letras - Comissão Linguística - Linguateca - Faculdade de Letras

## Resultados

- Centenas de visitas por semana!
- Muito útil para o ensino de linguística e tradução
- Grandes possibilidades para avaliação de TA entre EN <> PT
- Quantidade enorme de 'logs' para análise de TA, comparação de motores de TA etc.

## EVAL > TrAva

- Uma experiência pedagógica
- INPUT – frases 'naturais' tiradas de corpora, com um elemento de problema
- OUTPUT das 'caixas negras' – análise de problemas
- Avaliação de diferentes aspectos dos sistemas de TA

## TrAva – a ferramenta

- 4 motores de TA em linha EN > PT:
  - Free Translation
  - Systran
  - E T Server
  - Amikai
- Sistema de classificação dos erros de TA

## TrAva – escolha de ‘problemas’

- Homógrafos e itens lexicais polissémicos
- Unidades lexicais complexas
- Frases idiomáticas
- Colocações
- Complexidade sintáctica
- Ambiguidade sintáctica
- Resolução de anáfora

## TrAva – escolha do Input

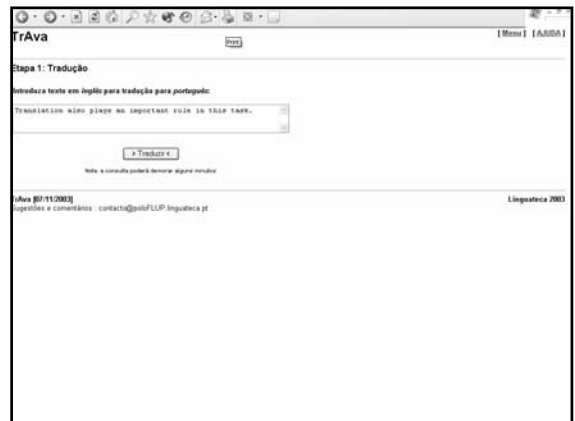
- Não utilização de ‘test suites’
- Nada de tentar ‘baralhar’ a TA
- Selecção de exemplos reais em corpora:
  - British National Corpus – BNC (100 milhões de palavras)
  - COMPARA
  - Outros textos
- Remover texto supérfluo

## Classificação do Input

- Léxico EN
- Códigos morfológicas (POS/parts-of-speech) utilizados pelo BNC
- Estructuras sintácticas = combinações de POS do BNC
- Observação dos resultados diferentes

## Classificação do Output

- Problema 1: Classificação do input
  - Classificação BNC POS – limitada a POS e insuficiente para análise sintáctica
  - Mas que tipo de classificação??
- Problema 2: Explicação do output
  - Erros lexicais – dicionários fracos
  - Erros devidos a falhas do sistema de análise sintáctico e/ou homógrafos
  - Falta de ‘conhecimento do mundo’ da TA





<b>[4] Frase Original:</b>	<b>The NFCC published an eleven-volume historical series entitled Eras of Nonconformity.</b>
Avulso:	indefinido
Origem:	BBC
Padrão FCS Frase Original:	<b>ATB AJC AB NN1</b>
Segmento Crítico:	<b>an eleven-volume historical series</b>
Tradução:	O NFCC publicou um onze-volume série histórica Eras intituladas de Nonconformity *
Fonte:	O NFCC publicou uma série histórica do onze-volume intitulada eras do nonconformity *
ET Server:	O NFCC publicou a série de eleven-volume histórica eras intituladas de Nonconformity *
Nota:	O NFCC publicou um onze-volume que séries históricas intitularam Eras de Nonconformity *
Número de traduções Errores:	2
Problemas identificados na tradução:	Adjectivo: Escolha Lexical Ordem: Interim do SN Coordenação: SNs
Concordância entre a tradução:	Traga A NFCC publicou a série histórica em onze volumes intitulada Eras de Não Conformidade.
<b>[5] Frase Original:</b>	<b>The sad reality for the newly elected Labour Minister is that Mr Patten's estimate was right.</b>
Avulso:	indefinido
Origem:	BBC
Padrão FCS Frase Original:	<b>ATB AJC AB NN1</b>
Segmento Crítico:	<b>the newly elected Labour Minister</b>
Tradução:	A realidade triste para o Ministro recentemente eleito de Trabalho é aquela estimativa de Patten de Sr era direita. *
Fonte:	A realidade sad para o ministro labour novo-eleito é que a estimativa do Sr. Patten era direita. *
ET Server:	A realidade triste para o ministro de trabalho recentemente eleito é esse Senhor que a estimativa de Patten era constantemente *
Nota:	A realidade triste para os recentemente elegeram Labura e Ministro é a estimativa daquele Mr Patten era certa. *
Número de traduções Errores:	2

## 'Lexical Bundles'

- Biber et al
- (também 'Multi Words Units'..)
- Ex. *I want to, you are supposed to, you were meant to, you want me to, in the course of, at the time when, as far as... is concerned ...*

<b>[6] Frase Original:</b>	<b>I want you to be concerned about your next door neighbour.</b>
Avulso:	indefinido
Origem:	MotoWEB
Padrão FCS Frase Original:	<b>I want you to VBI AB PRP</b>
Segmento Crítico:	<b>I want you to</b>
Tradução:	Quero ser preocupado sobre seu so lado vizinho. *
Fonte:	Eu quero ser preocupado sobre seu vizinho seguinte da porta *
ET Server:	Acesso Falhado
Nota:	Eu quero que você seja interessado sobre seu próximo vizinho de porta
Número de traduções Errores:	2
Problemas identificados na tradução:	Verbo: Escolha Lexical Conjunções: Conjunções Subordinadas
Concordância entre a tradução:	Modo verbal: indefinido no lugar de conjunção
<b>[7] Frase Original:</b>	<b>What do you want me to do?</b>
Avulso:	indefinido
Origem:	COMPARA
Padrão FCS Frase Original:	<b>DTO VMB you want me to VDI UNC</b>
Segmento Crítico:	<b>you want me to</b>
Tradução:	O que quer me fazer? *
Fonte:	Que você me quer fazer? *
ET Server:	Acesso Falhado
Nota:	O que quer você me fazer? *
Número de traduções Errores:	3
Problemas identificados na tradução:	Pronomes: Pronomes Pessoais (Função Sujeito) Conjunções: Conjunções Subordinadas
Concordância entre a tradução:	Modo verbal: indefinido no lugar de conjunção - Pronome pessoal: 1ª/2ª/3ª - Substivo

<b>[8] Frase Original:</b>	<b>I'm not sure if I'm meant to be a factory worker.</b>
Avulso:	indefinido
Origem:	?
Padrão FCS Frase Original:	<b>CJS I'm meant to VBI ATD NN1</b>
Segmento Crítico:	<b>I'm meant to</b>
Tradução:	Eu não estou seguro se sou quando dizer para ser trabalhador de fábrica. *
Fonte:	Eu não sou certo se sou significado para ser um trabalhador de fábrica. *
ET Server:	Não tenho certeza se sou significado para ser um trabalhador de fábrica. *
Nota:	Eu não estou seguro se eu sou quando dizer ser um trabalhador de fábrica. *
Número de traduções Errores:	4
Problemas identificados na tradução:	Verbo: Escolha Lexical
Concordância entre a tradução:	Novamente a preposição "Para" em excesso, desta vez em duas versões: FreeTranslation e ET Server.
Concordância entre a tradução:	A Conjunção "Se" obriga novamente a recorrer ao contexto, para determinar a necessidade ou não do uso do conjunção.
<b>[9] Frase Original:</b>	<b>You're meant to do that stuff in private.</b>
Avulso:	indefinido
Origem:	MotoWEB
Padrão FCS Frase Original:	<b>You're meant to VBI DTB NN1 PRP</b>
Segmento Crítico:	<b>You're meant to</b>
Tradução:	São quando dizer para fazer aquele material em privado.
Fonte:	Você é significado fazer esse material em confidencial. *
ET Server:	Você é significado para fazer esse material em particular. *
Nota:	Você é quando dizer fazer aquela matéria prima dentro privado. *
Número de traduções Errores:	3
Problemas identificados na tradução:	Verbo: Escolha Lexical
Concordância entre a tradução:	Preposição "Para" em demasia

## Homografia e Polissemia

- Exemplos
- *Round*
- *Run*

<b>[1] Frase Original:</b>	<b>Richard has round brown eyes.</b>
Avulso:	debe ser
Origem:	?
Padrão FCS Frase Original:	<b>AB</b>
Segmento Crítico:	<b>[round="round" &amp; part="AB"]</b>
Tradução:	Richard tem olhos marrons redondos.
Fonte:	Richard tem em volta dos olhos marrons *
ET Server:	Acesso Falhado
Nota:	Richard tem redondos olhos marrons.
Número de traduções Errores:	1
Problemas identificados na tradução:	Adjectivo: Escolha Lexical
Concordância entre frase de origem:	Adaptada.
<b>[2] Frase Original:</b>	<b>When we round the next bend, bring your boat.</b>
Avulso:	debe ser
Origem:	?
Padrão FCS Frase Original:	<b>VBI</b>
Segmento Crítico:	<b>[round="round" &amp; part="VBI"]</b>
Tradução:	Quando nós arredondamos o lago depois, traz seu barco. *
Fonte:	Quando nós arredondamos a curvatura seguinte, traga seu barco *
ET Server:	Acesso Falhado
Nota:	Quando nós arredondamos a próxima curva, traga seu barco. *
Número de traduções Errores:	3
Problemas identificados na tradução:	Verbo: Escolha Lexical
Concordância entre frase de origem:	Adaptada.

<b>[7] Frase Original:</b>	<b>I suppose you could make up a little compilation and send your friends round</b>
Tradutor:	debor.amos
Idioma:	COMPARSA
Parâmetro POS Free Origin:	AVD
Segmento Clítico:	ADV
FreeTranslation:	"Suponho-o pode compor uma compilação pequena e convidar seus amigos redondo?"
Textin:	"Eu suponho que você poderia fazer acima uma compilação pequena e convidar seus amigos redondo?"
El Server:	"Suponho que você poderia compilar uma pequena compilação e convidar os seus amigos a redondo?"
Nota:	"Eu suponho você poderia compor uma pequena compilação e poderia convidar seus amigos redondo?"
Número de traduções brutas:	4
Problemas identificados na tradução:	Resolução: MVAQADPpp
Tradução proposta:	Talvez ele ganhe fazer uma pequena compilação e convidar os amigos para irem lá a casa (Compar).
Comentário sobre frase de origem:	Frase abreviada.

<b>[8] Frase Original:</b>	<b>One man had blood seeping through a bandage we applied round his head.</b>
Tradutor:	debor.amos
Idioma:	COMPARSA
Parâmetro POS Free Origin:	AVD
Segmento Clítico:	ADV
FreeTranslation:	"Um homem teve sangue penetrar por uma compressa embaixo em torno de a sua cabeça. "
Textin:	"Um homem teve o sangue estar através de um bandage embaixo em volta de sua cabeça?"
El Server:	"Um homem teve sangue que gotejava através de uma bandagem embaixo de a volta de sua cabeça."
Nota:	"Um homem teve sangue que vazou por uma bandagem embaixo de volta a cabeça dele?"
Número de traduções brutas:	3
Problemas identificados na tradução:	Resolução: MVAQADPpp
Tradução proposta:	Havia um homem com sangue a ocorrer de uma ligadura que tinha estado à cabeça.
Comentário sobre frase de origem:	Tradução: COMPARA.

<b>[9] Frase Original:</b>	<b>My name is Glória and I run a beauty salon</b>
Tradutor:	cjenes
Idioma:	ITC
Parâmetro POS Free Origin:	7
Segmento Clítico:	ATB VVB
FreeTranslation:	"Chamo-me Glória e eu como um salão de beleza?"
Textin:	"Meu nome é Glória e eu funciono um salão da beleza?"
El Server:	"O meu nome é Glória e eu executo um salão de beleza. "
Nota:	"Meu nome é Glória e eu como um salão de beleza?"
Número de traduções brutas:	4
Problemas identificados na tradução:	Verbo: Escolha Lexical
Comentário sobre frase de origem:	Assume por defeito 'correr', 'funcionar'.

<b>[11] Frase Original:</b>	<b>In the short run this new body is seen by the Commission as a useful means</b>
Tradutor:	cjenes
Idioma:	ITC
Parâmetro POS Free Origin:	NNI
Segmento Clítico:	
FreeTranslation:	"Na corrida curta este novo corpo é visto pela Comissão como um meio útil?"
Textin:	"No short funcionado este corpo novo é visto pelo comissão como um útil significa?"
El Server:	"Acesso Falhado."
Nota:	"Na corrida curta este corpo novo é visto pela Comissão como um meios úteis?"
Número de traduções brutas:	3
Problemas identificados na tradução:	Substantivo: Escolha Lexical
Comentário sobre frase de origem:	'Run' substantivo, procedido de um adjetivo 'short'. Será interessante ver até que ponto os motores de tradução identificam esta vertente.
Comentário sobre a tradução:	Não conseguiu identificar 'run' como substantivo.
Comentário sobre frase de origem:	A tradução seria: "A curto prazo. "

<b>[7] Frase Original:</b>	<b>Parkside.</b>
Tradutor:	cjenes
Idioma:	7
Parâmetro POS Free Origin:	
Segmento Clítico:	was run down
FreeTranslation:	"Huskisson de william, Presidente da Tabua de Negocio, foi corrido para baixo e matado em Parkside."
Textin:	"William Huskisson, presidente da placa de comércio, foi funcionado para baixo e matado em Parkside."
El Server:	"Acesso Falhado"
Nota:	"William Huskisson, Presidente da Tabua de Comércio, foi corrido abaixo e matado a Parkside."
Número de traduções brutas:	3
Problemas identificados na tradução:	Verbo: Escolha Lexical
Comentário sobre frase de origem:	"O motor não conseguiu traduzir 'was run down' no sentido de 'foi atropelado'. Assume por defeito 'foi corrido abaixo'."
Comentário sobre a tradução:	Tradução possível: " ... fo atropelado e morto. "

<b>[8] Frase Original:</b>	<b>She fell silent and a tear started to run down her cheek.</b>
Tradutor:	cjenes
Idioma:	7
Parâmetro POS Free Origin:	
Segmento Clítico:	No run down
FreeTranslation:	"Cau silencioso e uma lágrima começou correr para baixo a sua bochecha."
Textin:	"Cau silencioso e um raio começou funcionar abaixo seu mordente."
El Server:	"Acesso Falhado"
Nota:	"Ela se calou caída e uma lágrima começou a correr abaixo a bochecha dela."
Número de traduções brutas:	3
Problemas identificados na tradução:	Verbo: Escolha Lexical
Comentário sobre frase de origem:	Apesar da tradução de Free Translation e do Amblai não estar muito correcta em termos de sintaxe, apresentam a essência da ideia 'correr pela face abaixo'. O Systan assume por defeito 'funcionar'.

## Um exemplo lexical ou sintáctico?

- 'She drove her children to school'
- 'She drove her children up the wall'
- 'She sang the baby to sleep'

## Acesso

- Pesquisador
  - Trabalha no seu espaço individual
  - Pode aceder ao seu trabalho
- Professor / administrador pode aceder ao trabalho do aluno para ajudar
- Os resultados aparecem todos no 'log' do METRA

## Links

- METRA :
  - <http://www.linguateca.pt/metra>
- TrAva :
  - <http://www.linguateca.pt/trava/>

## Pesquisa futura

- Melhorar TrAva
  - Reformular bases teóricas
  - Aplicação de 'glass box'?

## METRA -Pedagogical applications

- Contrastive linguistics (undergraduate)
  - Choose item of grammar or lexicon
  - Find examples in BNC or COMPARA
  - Submit to METRA
  - Compare MT to human translation
  - Reflect on results
- Better understanding of interconnection of lexicon + syntax + context

## Conclusions - teaching

- A successful experiment
- Future implications for research
- Useful pedagogically
  - As linguistic analysis
  - As appreciation of MT
- Development of training of working with MT

## Conclusions - research

- Further work needs to be done on:
  - Improving METRA
  - Classifications in TrAva
  - Searching METRA logs
- A lone researcher can use it to examine a restricted area
- BUT – a large team is needed to make practical use of the tools

## Other research

- Analysis of METRA logs
  - Who?
  - Why?
  - What?
  - Comparison of MT engines
  - On-line consultation of logs

## METRA logs

- A large number of 'hits' per day
- GBs of material – and rising!
- What can we do with it?
- We have ideas – but welcome suggestions!





## Useful bibliography

- ARNOLD, D, BALKAN, L., LEE HUMPHREYS, R. Lee, MEIJER, S., & SADLER, S. (1994) *Machine Translation - An Introductory guide*. Manchester & Oxford : NCC Blackwell. ISBN 1-85554-217-X - or at: <http://www.essex.ac.uk/linguistics/clmt/MTbook/>.
- COLE, Ron (ed) 1996 "Survey of the State of the Art in Human Language Technology" Chapter 8 - Multilinguality, at the Center for Spoken Language Understanding, Oregon. <http://cslu.cse.ogi.edu/HLTsurvey/ch8node2.html#Chapter8>
- MELBY, Alan K. 1995, *The Possibility of Language: A discussion of the Nature of Language, with implications for Human and Machine Translation*. Amsterdam: John Benjamins Pub. Co.

## Useful bibliography

- MELBY, Alan 1995 ' Why Can't a Computer Translate More Like a Person?' The 1995 Barker Lecture at ttt.org - Translation, Theory and Technology. at: <http://www.ttt.org/theory/barker.html>
- MLIM - Multilingual Information Management: Current levels and Future Abilities - report (1999) Chapter 4 at: <http://www-2.cs.cmu.edu/~ref/mlim/chapter4.html>