

# Avaliação de Tradução Automática

## Alguns Esforços de Abordagem

Luís Sarmento: [Linguatca@CLUP](mailto:Linguatca@CLUP)  
Belinda Maia: [CLUP/FLUP](mailto:CLUP/FLUP)

## Problemática da Avaliação de TA

- Assunto muito complexo
- Envolve muitas perspectivas
  - Linguistas
  - Engenheiros
  - Sistemas Comerciais
  - Sistemas Experimentais/Académicos
  - ...
- Uma pesquisa no Google ilustra que a avaliação de TA para o Português é uma área que está ainda a nascer
- É necessário conhecer melhor os problemas!
- Necessidade de estabelecer bases comuns para a desenvolvimento cooperativo

## Objectivos

O Pólo de Porto da Linguatca iniciou actividades na área da Avaliação da Tradução Automática em Janeiro de 2003, tendo por objectivos:

1. Estudo do estado actual dos Motores de TA
2. Estudo dos problemas associados à TA de Português
3. Definição colectiva de metodologias de Avaliação TA
4. Construção cooperativa de *Test-Suites* e *Base-Lines*
5. Organização de Sessões Regulares de Avaliação de TA

## Breve resumo de Actividade

- As actividades de desenvolvidas até agora na área da Avaliação de TA abrangem duas fases:
  - Fase 1:
    - observação do desempenho dos Motores de TA
    - desenvolvimento ferramentas de suporte à observação e experimentação: METRA e Boomerang
    - identificação empírica de alguns dos problemas de TA
  - Fase 2:
    - desenvolvimento de um candidato a sistema de classificação de erros a partir das observações realizadas
    - criação de uma ferramenta de recolha de frases problemáticas para TA: EVALI
    - apresentação dos dados recolhidos
    - discussão pública e análise de resultados...

## Fase 1

- Necessidade de olhar de uma forma mais abrangente para os Motores de TA
- Tentar identificar as potencialidades e limitações da TA oferecida
- Fazer algumas comparações empíricas
- Disponibilidade:
  - Motores Web
  - Demonstrações On-line de Sistemas Comerciais
  - Alguns sistemas disponíveis em Licença GNU
- Investigar a tecnologia usada pelos Motores de TA disponíveis

## TA na Web para Português (1)

- Amikai
- BabelFish
- FreeTranslation
- Google
- IBM Websphere
- Intertran
- Itranslator
- Linguatca E-Translation Server
- LogoMedia
- Systran
- T-Mail
- VISL- Machine Translation
- WorldLingo

## TA na Web para Português (2)

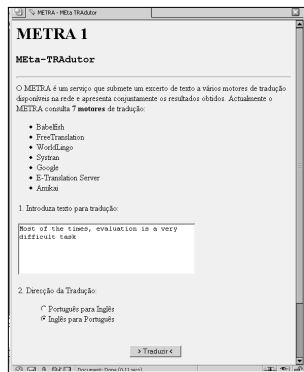
- Interface (Demo) Web
- Alguns usam a mesma família de tecnologias de TA
- Alguns usam o **mesmo** Motor de TA mas possivelmente com diferenças de preparação...
- Alguns Motores traduzem PT -> EN e EN -> PT outros apenas num dos sentidos
- Mais Motores disponíveis mas não possuem recursos para o Português

**Boa oportunidade para uma primeira comparação**

## METRA – Meta TRAdutor

- Uma primeira experiência de estudo e comparação
- Encaminha pedido de Tradução PT-> EN ou EN ->PT:
  - Babelfish
  - FreeTranslation
  - WorldLingo
  - Systran
  - Google
  - E-Translation Server
  - Amikai
- Aguarda pelos resultados e apresenta ao utilizador em simultâneo todas as traduções obtidas
- <http://poloclup.linguateca.pt/ferramentas/metra/>

## METRA – Meta TRAdutor (1)



## METRA – Meta TRAdutor (2)

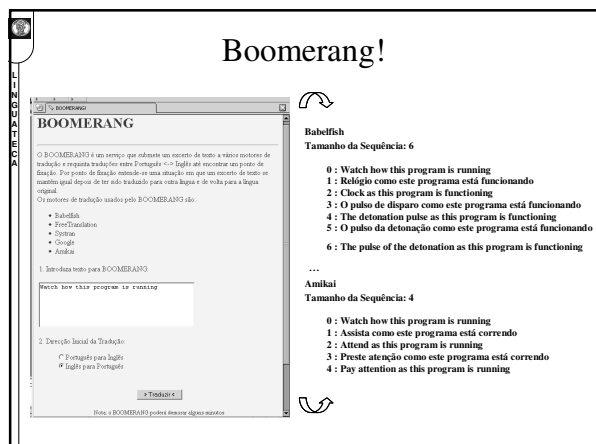


## METRA – Meta TRAdutor

- Permite
  - o desenvolvimento de alguma sensibilidade para a comparação.
  - a deteção de algumas limitações nas versões disponíveis dos motores
    - Falhas Lexicais
    - Muitos Problemas na Análise de Compostos/Homónimos
    - Muitos Problemas de Concordância na Síntese
    - Erros Ortográficos!
    - ...
  - a deteção de frases cuja tradução é problemática ou aparentemente boa
    - "Eu fui passar" -
    - "I can hardly restrain myself from shouting." -
    - "We won't be able to compact the lexicon with new entries now nor in the near future" +
  - iniciar um abordagem sistemática à classificação dos erros de tradução que é possível encontrar

## Outra Ferramenta: Boomerang!

- Reutiliza a estrutura montada para o METRA
- Requisita traduções entre Português e Inglês até encontrar um ponto ponto fixo
- Explora as não linearidades da TA
- Permite observar:
  - operações de substituição de palavras
  - tendência afunilamento/simplificação
- <http://poloclup.linguateca.pt/ferramentas/boomerang/>



## Algumas Observações

- Depois de alguma experimentação com estas ferramentas foi possível observar que:
  - Actualmente os motores de TA geram muito erros de tradução ao nível da frase: não parece razoável neste ponto testar estruturas mais abrangentes
  - Expressões idiomáticas são naturalmente difíceis para a TA (mas também para o Tradutor Humano)
  - A quantidade de problemas numa TA está normalmente relacionada com o tamanho da frase
- Relativamente aos utilizadores:
  - A detecção do problemas de tradução é feita melhor pelos utilizadores cujo idioma nativo é o das frases traduzidas
  - Há um risco efectivo de os utilizadores requisitarem a tradução de frases pouco naturais ou muito improváveis na língua estrangeira.

## Fase 2

- Objectivos:
  - Desenvolver maior compreensão pelos problemas associados a TA para Português
  - Desenvolver um sistema de categorização de erros para TA de Português
- Processo:
  - Recolher um grande número de frases problemáticas para TA e as respectivas traduções
  - Recolher as observações do maior número de utilizadores sobre o tipo de erros de TA que encontram ou sobre os problemas que **não** encontram
  - Recolher o máximo de informação adicional para posterior análise de nível superior (ex: categorizar as frases originais por POS)
- Método:
  - Ferramenta de Recolha Cooperativa de Frases Problemáticas (EN->PT):

**EVAl1**

## EVAl1

- Foi lançado o repto à comunidade para a construção colectiva de um base de dados de frases problemáticas para TA
- A recolha seria distribuída e realizada via Web
- Participantes:
  1. Alunos de Mestrado TT da FLUP
  2. Alguns participantes associados à Linguateca
- Problema imediato: diferentes níveis de conhecimentos Linguístico e Experiência com TA...

## EVAl1 – A Recolha

- Algumas linhas orientadores para a **Recolha**:
  - evitar frases que sejam evidentemente problemáticas mesmo para um tradutor humano: não é lícito exigir à máquina melhor tradução do que a um humano.
  - para cada problema encontrado devem ser preferidas na recolha as frases mais “simples” onde esse problema ainda seja observável: *Princípio da Parcimónia*
  - deve ser preferidas frases “reais” i.e. não devem ser colocadas à TA frases que poderão nunca surgir na língua estrangeira ou cuja estrutura é pouco natural: evitar “biasing”
- Recolha foi feita sobre frases em Língua Inglesa retirada preferencialmente de Corpora

## EVAl1 – A Recolha na FLUP

- Internamente encontrava-se disponível para pesquisa o BNC
- Desenvolveu-se um interface BNC optimizado para:
  1. Pesquisa por Expressões Regulares
  2. Pesquisa por Padrões POS
- Divisão do BNC em 10 partes para tornar as pesquisas mais rápidas

## EVAL1 – A Categorização

- Pretendia-se categorizar as frases originais e os Erros de/na Tradução
- Categorização da Frase Original:
  - Frases em Inglês
  - Categorização “Natural”: POS
  - Sistema de Categorização POS/BNC
  - Categorização POS apenas da parte relevante para a tradução (discutível...)
  - Objectivo: detectar padrões acima do léxico!
- Categorização dos Erros de/na Tradução:
  - Exige a Definição de um Sistema
  - Muito mais difícil, ainda menos consensual (e com outros problemas)

## EVAL 1 – O processo (1)

- 1) Recolha
- 2) Interface EVAL: submissão a TA

## EVAL 1 – O processo (2)

1. Categorização da Frase Original por POS + Descrição do Elemento em Teste
2. Categorização das Frase(s) Traduzida(s) recorrendo a uma lista de erros ao Nível Lexical e Sintático

## EVAL 1 – Sistema de Monitorização (1)

1. As tarefas de recolha e classificação mostraram-se bastante complexas
2. Necessidade de acompanhar os colaboradores em tempo real
3. Ganhar maior percepção relativamente aos problemas de tradução
4. Recolher algumas estatísticas sobre o desempenho dos motores de tradução
5. Identificar o problemas do sistema de classificação que adoptamos

## EVAL 1 – Sistema de Monitorização (2)

1. Listar frases por categoria(s)
2. Estatísticas de Utilização

## EVAL 1 – Resultados (1)

- Mais de 1000 frases recolhidas com segmentos anotados POS.
  - > 50 % provêm do BNC: é possível completar a anotação POS
- Alguns Resultados da Contabilização:

POS	POS	POS	POS	POS	POS	POS	POS	POS	POS
NP	NP	NP	NP	NP	NP	NP	NP	NP	NP
VP	VP	VP	VP	VP	VP	VP	VP	VP	VP
ADP	ADP	ADP	ADP	ADP	ADP	ADP	ADP	ADP	ADP
PP	PP	PP	PP	PP	PP	PP	PP	PP	PP
IN	IN	IN	IN	IN	IN	IN	IN	IN	IN
CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
CD	CD	CD	CD	CD	CD	CD	CD	CD	CD
DT	DT	DT	DT	DT	DT	DT	DT	DT	DT
AD	AD	AD	AD	AD	AD	AD	AD	AD	AD
PRN	PRN	PRN	PRN	PRN	PRN	PRN	PRN	PRN	PRN
PRT	PRT	PRT	PRT	PRT	PRT	PRT	PRT	PRT	PRT
CONJ	CONJ	CONJ	CONJ	CONJ	CONJ	CONJ	CONJ	CONJ	CONJ
NUM	NUM	NUM	NUM	NUM	NUM	NUM	NUM	NUM	NUM
SYM	SYM	SYM	SYM	SYM	SYM	SYM	SYM	SYM	SYM
UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK	UNK

Pistas para novas pesquisas por POS no BNC mas Existe evidente défice de cobertura

## EVAL 1 – Resultados (2)

- Listagem de “Problemas”
  - Top 25:

Tipo de Problema	N
Verbos: Escolha Lexical	319
Substantivos: Escolha Lexical	133
Adjectivos: Escolha Lexical	87
Preposições: Escolha Lexical	78
Verbos: Tempos Simples	76
Ordem: Interior do SN	68
Verbos: Tempos Compostos	59
Adverbios: Escolha Lexical	53
Verbos: Locuções Verbais	37
Determinantes: Escolha Lexical	36
Ordem: Interior da Frase	33
Determinantes: Artigos	31
Ordem: Interior do SV	30
Elisão: Determinante - Artigo	26
Concordância: Número	25
Ordem: Outra	25
Concordância: Género	22
Elisão: Sujeito - Pronome	21
Conjunções: Conjunções Subordinadas	21
Verbos: Verbos Reflexivos	21
Pronomes: Escolha Lexical	18
Coordenação: SVs	17
Conjunções: Escolha Lexical	16
Resolução: Outra	14

### Problemas:

- Algumas Categorias Possivelmente Ambíguas
- Dificuldade dos Colaboradores em Distinguir Casos: Treino!
- Necessidade de sistemas de Classificação Simples mas Abrangentes: **Sistemas Mais Específicos**

## EVAL 1 – Resultados (3)

- Tentativa de Correlacionar alguns POS Frase Original com Erros na Frase Traduzida
- Tarefa complexa mas interessante para a preparação de um futuro *test-suite*
- Necessidade de Análise mais profunda
- Análise Caso-a-Caso

1) Verbos: Escolha Lexical : 319

POS1 POS2 POS3 POS4 POS5	N
VBE TOI	14
VMB	12
PNP VTI CDS PNP	9
SNL VFO	9
"did you" Oag	8
VMB PNP VTI CDS	7
VMB PNP VTI VFD	4
AUS VFO	4
"is it" Oag	4
no have	3
VTI	3
was down	3
SNL VTN	3
TON VTB POS HNB	3
was not down	3
have you Oag	3
PNP VTD ALC VTI	3
stop had not stop	3
PNP AUV VTB	2
You're never to VCE DPO HSI PNP	2
"are they" Oag	2
AUTO HSI VFD	2
PNP VMB AUV CDS	2
PNP VTD ALC VTB	2
VTB	2
PNP PNI HSI	2

## EVAL 1 – Próximos Passos

- Novos Sistemas de Classificação de Erros:
  - Mais específicos para combinar simplicidade com abrangência
  - mais centrada nos erros de Tradução
  - Sintáctica + Sintagmática
  - aprendizagem com as limitações encontradas
  - discussão com a comunidade
- Aumentar a Integração entre BNC + Motores TA
  - Possibilitar a categorização POS integral
  - Melhorar o fluxo de trabalho
- Treino de colaboradores
  - Produção de Documentação sobre o Sistema
  - Produção de Documentação sobre os resultados já obtidos
  - Melhorar a precisão da recolha
  - Treinar Futuros Avaliadores

## CORTA

### CORpus de Traduções Automáticas

- Disponibilização dos Resultados da Recolha à Comunidade para análise conjunta
- Disponível em:
  - <http://poloclup.linguatca.pt/ferramentas/corta/>
- Pesquisa Por Expressões Regulares (CQP) nas Frases Originais e Traduções

## Alguns Projectos/Ideias Para o Futuro

- EVAL 1.1
  - Desenvolvimento de Novos Sistemas de Classificação de Erros
  - Maior compreensão sobre o estado actual dos Motores de TA
  - Continuação de Recolha de Frases Problemáticas para construção de *Test-suites Segmentadas*
  - Recolha de “Estruturas Seguras” para Escrita Controlada
- Construção Cooperativa de Corpus Paralelo de Traduções H+A:
  - {1 Frase Original, M x Tradução Humana, N x Tradução Automática}
  - Frases Recolhidas no EVAL
  - Interface Web de distribuição de frases por vários Tradutores Humanos