

A rede de tradução e como operacionalizá-la com base em corpos paralelos: Primeiros passos

1

DIANA SANTOS

Apresentação no encontro Per-Fide (Braga, 17 de Setembro de 2010)

Percurso

2

- Reminiscências de uma velhinha
- Tese de mestrado:
 - Aluna bem-comportada
- Salto quântico: doutoramento
 - e se olhássemos para a realidade da tradução?
- Realidade à esquina
 - Apoio ao que os outros querem fazer
 - Construção de recursos, disponibilização, (des)orientação
 - Discursos políticos, arguição...
- Passados 25 anos... desenterrar o machado?

No meu tempo...

3

- A disciplina chamava-se “Natural language understanding”
- O que estava a dar era “inferência baseada em caminhos” (em redes semânticas)
- As discussões eram sobre
 - Tipos de programação: funcional ou lógica
 - Enquadramentos ou redes semânticas
 - Revisão de crenças ou propagação de valores
- A minha visão era: IA é filosofia aplicada, ou engenharia filosófica

Tradução automática

4

- Duas propostas originais
- Expressões com várias palavras dependentes do par de línguas: “fazer as malas” – “pack”, implementação invocando a gramática da língua em questão
- Separação entre transferência motivada pela língua da origem, e escolha das várias traduções possíveis com base na língua de destino (quase-sinónimos)

Doutoramento (intenção)

5

- E a semântica? E a semântica formal? Não precisarei delas para traduzir?
- Caí na esparrela do “formalismo”, de que só escapei no fim da tese
- Arranjei os orientadores mais formais que podia ter encontrado: Lauri para teoria de jogos e Amílcar matemático, e comecei a ler sobre lógicas temporais... ontologias de acontecimentos, etc.

Banho de realidade

6

- Lista de problemas no contraste do tempo e aspecto entre as línguas português e inglês
 - aconselho a que seja lido, para quem pensa que os factos existem independentemente das teorias
 - ou que as teorias permitem obter uma lista de “factos” fidedignos que podem ser encontrados na realidade
- Impasse. Então o Lauri comprou *A Pérola* e pediu-me para formalizar a primeira frase, e/ou mais umas frasesinhas
- E fez-se luz! Pelo menos alguma...

Trabalho árduo... mas reconfortante

7

- Após ter traduzido mentalmente e espremido todos os exemplos apresentados nos artigos de linguística...
- Agora tinha “milhões” de exemplos que precisava de classificar e analisar, mas
 - Esses exemplos não tinham sido produzidos por mim
 - Dei-me conta que o meu inglês não melhorava a ler artigos de linguística, mas sim melhorava a olhos vistos a ver as soluções e os problemas que os tradutores tinham tido
 - Basicamente, passei a VER

Um modelo para o português emergiu

8

- Claro, aos bocadinhos, pegando exemplos daqui e dali
- Isso permitiu compreender a forma como o português via o inglês, e de certa forma reformular o meu modelo do inglês
- E através do contraste surgiu a ideia da rede de tradução (generalização da rede aspectual do Moens (1987))
- Muito importante indicar que foi um processo iterativo/hermenêutico, e que várias vezes o mesmo exemplo foi reorganizado/reclassificado

Processo qualitativo

9

- De certa forma, foi um processo qualitativo, porque embora a distribuição quantitativa tivesse sido um primeiro ponteiro para observar alguns casos, todos os exemplos foram analisados por mim para chegar às pequenas redes de tradução que os ilustram, e depois generalizados para coisas mais gerais, através da consulta a outros exemplos
- Nem sempre consegui produzir uma explicação
- Nunca assumi que a tradução era perfeita, mas assumi sempre que o tradutor era inteligente (e portanto erros ou divergências eram dados muito interessantes)

Daí para teorizar...

10

- ... foi um pequeno passo
- E eis-me a tecer considerações sobre a tradução, sobre a semântica, sobre a metodologia linguística, ...
- Mas para implementar... que eu sei que é o que realmente me iria dar gozo, e que me faria aprender muito mais sobre a tradução e o tempo e aspecto... até hoje! ☹
- Mas não tenho tempo para desculpas.

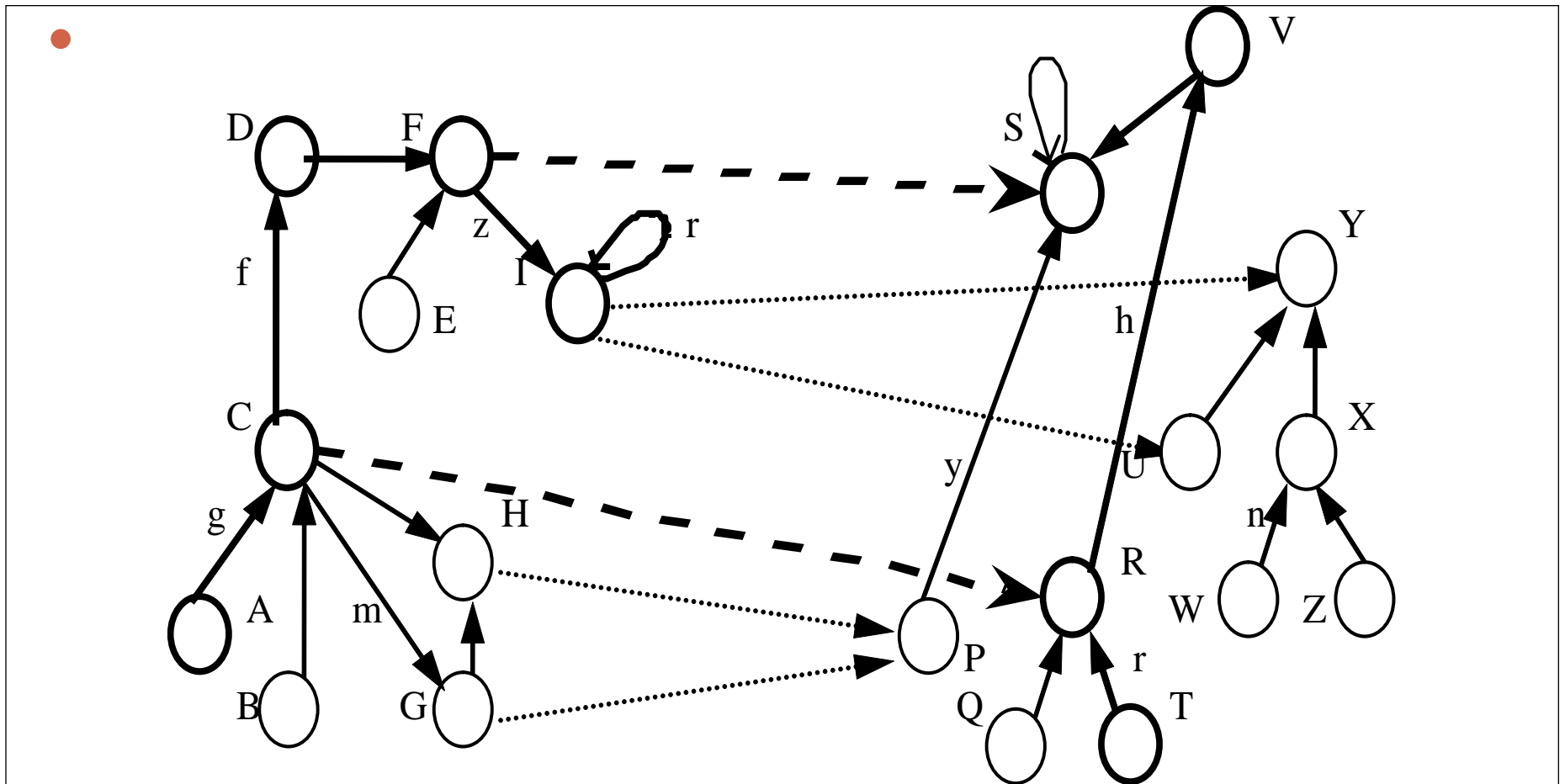
O que é preciso fazer?

11

- Analisar um conjunto de frases alinhadas de acordo com os parâmetros temporais-aspetuais presentes nas redes aspetuais de cada língua
 - usando o PALAVRAS (Bick, 2000) como base
 - e o CLAWS (Rayson & Garside, 1998) para o inglês
- Tentar inferir/induzir arcos de tradução
- Pesquisar os arcos e a classificação aspetual das palavras e das frases com base nos exemplos encontrados
- Isto daria um primeiro varredor de traduções mais artilhado

Um par original-tradução numa rede de tradução

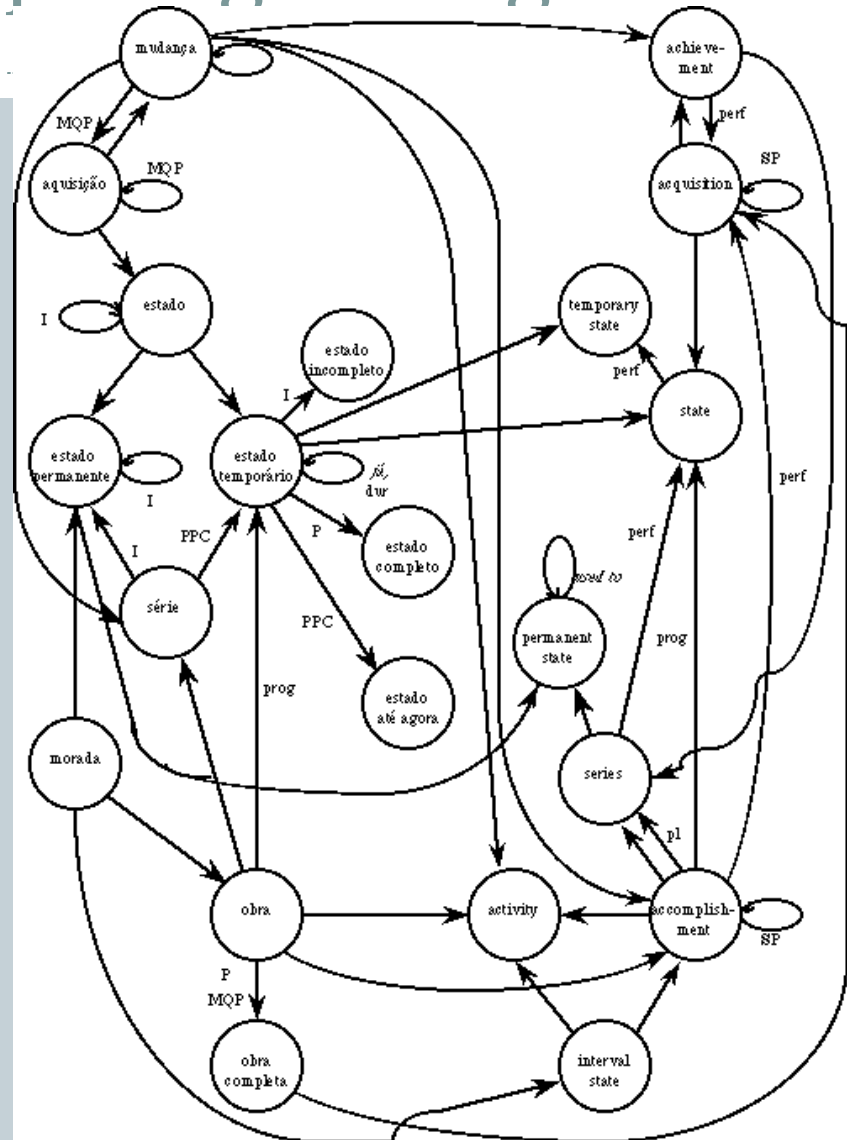
12



A rede de tradução português-inglês

13

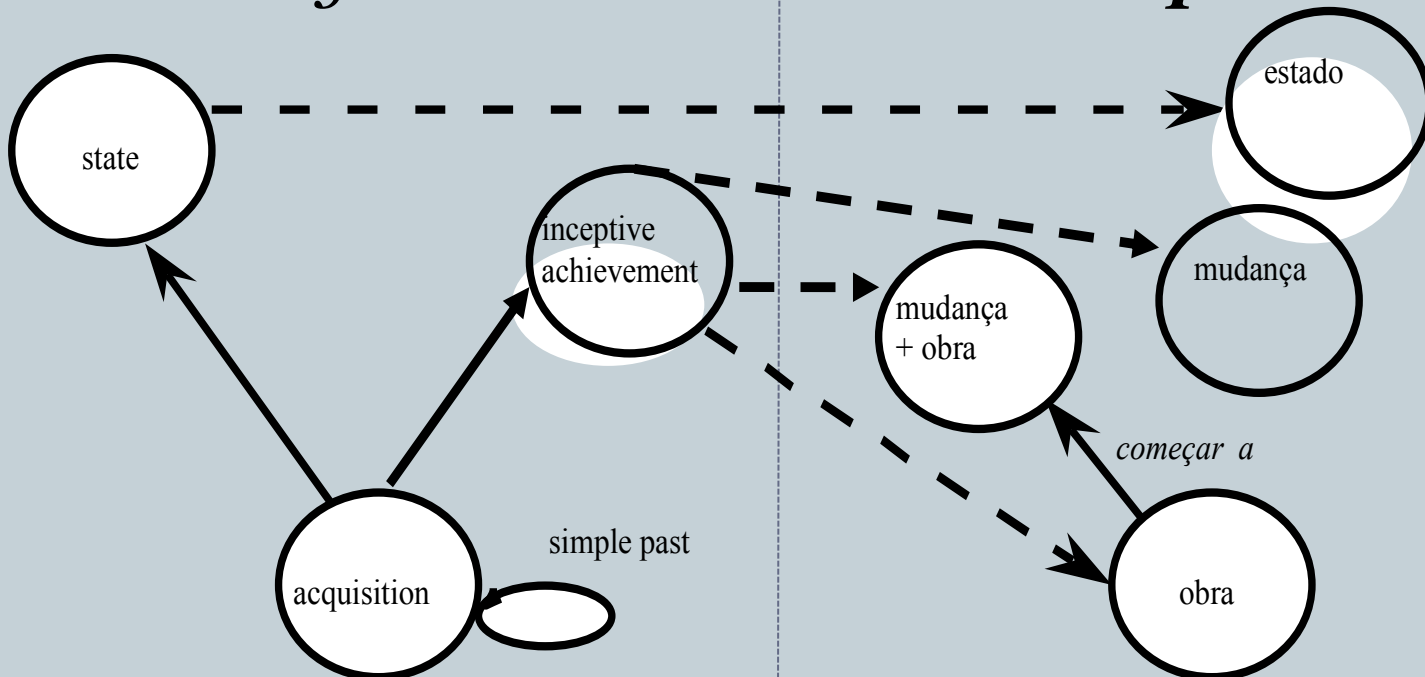
- Formalização do tempo e aspeto das duas línguas e observando a tradução de uma para a outra
- Na direção oposta, há uma rede de tradução diferente!



Exemplos da tradução de *be* (acquisition)

14

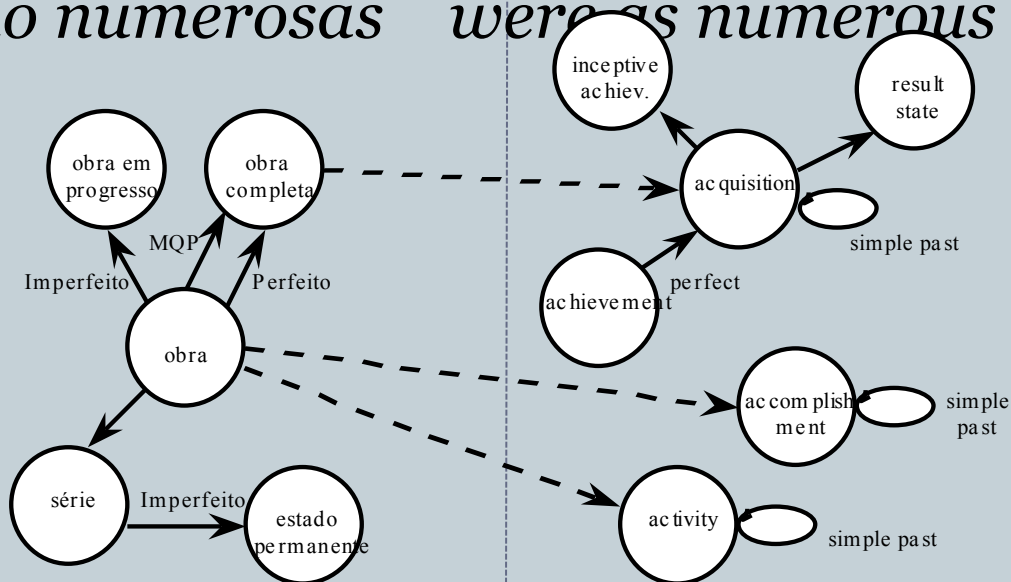
- *But Kino **was** in motion*
- *And suddenly he **was** afraid of his talking*
- *Now uncertainty **was** in Kino*
- *mas Kino **começou** a mexer-se*
- *E, de repente, **assustou-se** com os próprios pensamentos*
- *A dúvida **apoderou-se** de Kino*



Exemplos da tradução de Obras

15

- ***Passei** entre as mesas empilhadas*
- *O caso **dera** que falar*
- *Onde agora, como não onde o frade se **distraíra**, as pedras eram tão numerosas*
- ***I walked** among the stacks of tables*
- *The episode soon **became** famous*
- *where now, unlike when the friar had **become** distracted, stones were as numerous*



Últimas palavras

16

- A melhor maneira de conhecer uma língua e perceber o seu funcionamento é observar como é que ela se porta quando traduzida e ao traduzir.
- Estudar a tradução humana pode não ser diretamente relevante para a tradução automática, mas é-o certamente indiretamente!
- Já não estou interessada em tradução automática, porque uma parte essencial da boa tradução passa pela personalidade/compreensão do tradutor, mas continuo interessada na análise automática de traduções!

Revisão de alguma literatura relacionada

- Obtenção automática de diferentes pares original – tradução a partir de corpos (mais ou menos) comparáveis (Monteanu & Marcu, 2005, Haghighi et al. 2008)
- Obtenção automática de paráfrases (Barzilay & Elhadad, 2003)
- Estudo de quase-sinónimos (ver o meu curso na ESSLLI para ponteiros e exemplos)

Referências

- Barzilay, Regina & Noemie Elhadad. “Sentence alignment for monolingual comparable corpora”. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*. ACL, NJ, 25-32.
- Bick, Eckhard. *The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Aarhus University Press, 2000.
- Haghighi, Aria, Percy Liang, Taylor Berg-Kirkpatrick & Dan Klein. "Learning Bilingual Lexicons from Monolingual Corpora", *ACL* 2008

Referências (cont.)

- Moens, Marc. "Tense, Aspect and Temporal Reference". PhD thesis, University of Edinburgh, 1987.
- Munteanu, Dragos Stefan & Daniel Marcu. "Improving Machine Translation Performance by Exploiting Non-Parallel Corpora". *Computational Linguistics* **31**, 4 (Dec. 2005), pp. 477-504.
- Rayson, Paul & Roger Garside. "The CLAWS Web Tagger". *ICAME Journal* **22**, 1998, pp. 121-123.