

## Aplicações de uma floresta sintáctica

Diana Santos  
www.linguateca.pt

## Para que serve

- qual a função
- qual a utilidade
- qual o resultado/impacto
- quais as consequências
  
- Primeiro, uma abordagem descritiva
- Depois, uma abordagem crítica

## Para ilustrar uma (teoria da) gramática

- Uma coisa é ter uma teoria que descreve uma (ou todas as) línguas
- Outra coisa muito diferente é
  - ter uma aplicação dessa teoria que cobre **texto real**
  - por oposição a texto fabricado para exemplificar uma teoria
- É diferente porque
  - uma gramática (ou teoria da gramática) não especifica geralmente como chegar a um dado resultado – operacionalização
  - o texto tem sempre um número crescente de pormenores – não se pode falar de uma gramática completa (no sentido de que todos os fenómenos já foram descritos)
  - não se pode analisar parcialmente uma frase/não se costuma

## Para criar dados para futuro processamento

- se se conseguir obter um número significativo de casos
- pode-se desenvolver modelos (ou treinar sistemas) que usam esses casos para analisar mais texto
- pode-se criar regras ou hipóteses para exploração linguística mais detalhada em corpora maiores
  - se a maior parte dos casos de comparação representam ironia, pode-se usar essa regularidade para extrair grandes quantidades de candidatos a casos de ironia de corpora
  - se a maior parte das orações relativas explicativas aparecem associadas ao sujeito, pode-se usar essa “regra” para extrair candidatos a sujeitos
  - se a coordenação de adjectivos é frequentemente usada após sintagmas preposicionais em alguns casos e não noutros...

## Para avaliar sistemas

- desenvolvidos independentemente
  - comparando os resultados
  - obtendo dados específicos
  - comparando abordagens de anotação (estudos de mutilação)
- desenvolvidos ou melhorados durante o processo
  - o PALAVRAS
  - sistemas de REM
  - sub-sistemas do AC/DC (atomização e separação de palavras)
- para fazer uma avaliação conjunta em sintaxe computacional
  - para obter dados aos quais já está associada uma distribuição
  - para comparar tarefas específicas sobre as quais há ou é possível chegar a um consenso

## Para fazer investigação em sintaxe... e semântica

- “mãos na massa”
- dado um conjunto de frases sistematicamente analisadas e compreendidas pelos falantes
  
- comparar com as intuições
- descobrir casos complicados
- estudar a interacção de fenómenos pertencentes a esferas diferentes
- identificar casos excepcionais
  
- ensinar sintaxe

## Para fazer investigação em informática

- Que tipos de formalismos são melhores para descrever o resultado
- Que tipo de gramáticas são necessárias
- Qual o melhor sistema para indexar e validar a informação
- Que tipo de necessidades têm os utilizadores de uma floresta
- Que tipo e forma de resultados são preferíveis
- Sistemas de ajuda à revisão da anotação
- Sistemas de visualização

## Descrição de casos conhecidos

- Penn Treebank
  - indução de gramáticas
  - avaliação de análise sintáctica: ParsEval
- SUSANNE
  - descrição minuciosa da língua
  - avaliação de análise sintáctica: GR-scheme
  - criação e avaliação de novas medidas: LAM
- Czech TD
  - construção e melhoria de dicionários
- NEGRA
  - novas metodologias de anotação e criação

## Crítica

- A maior parte das pessoas que usam as florestas não têm a noção do trabalho que lá está incluído, nem do que ainda falta ser feito
- A maior parte das pessoas que trabalham com florestas passam o tempo a criá-las ou a melhorá-las, não a usá-las
- As florestas são um investimento para o futuro, mas geralmente não acompanhado:
  - ainda não existem os utilizadores
  - os futuros utilizadores muito raramente exprimem os seus desejos e/ou necessidades (e quando os exprimem, precisam sempre de corpora muitíssimo maiores!)
  - os gramáticos (fora da equipa) estão aparentemente completamente desinteressados na existência de uma floresta sintáctica ou não para a sua língua