



Floresta sintá(c)tica: um recurso para avaliação

Diana Santos
Susana Afonso

<http://cgi.portugues.mct.pt/treebank/PaginaFloresta.html>

27 de Junho de 2002

1



Motivação

- É necessário texto real para avaliar sistemas que queiram lidar com texto real
- É necessário saber o que está bem do ponto de vista humano para poder avaliar o desempenho de sistemas automáticos
- É necessário verificar qual o espaço de convergência e de divergência para uma dada aplicação

27 de Junho de 2002

2



Motivação interna

- É preciso aprender a fazer projectos cooperativos que envolvam mais de dois lugares e pessoas com vários perfis
- É preciso aproveitar
 - a colaboração já existente
 - os recursos já disponíveis
- É preciso avaliar a qualidade dos corpora anotados no projecto AC/DC (e/ou a saída do analisador sintáctico utilizado)

27 de Junho de 2002

3



Resultado: recurso

- 1400+ árvores (35000+ palavras) revistas
- 40000+ árvores (1 milhão de palavras)
 - Publicamente disponíveis (acesso e entrega)
 - Com informação morfológica, estrutura de constituintes e função sintáctica
 - Documentação sobre a codificação e sobre várias opções linguísticas
 - Lista de discussão, promessa de criação de novos formatos se necessário

27 de Junho de 2002

4



Resultado: experiência

- Fixação de uma forma de revisão e de disseminação dos resultados
 - vários passos (e formatos), verificação da consistência, discussão comum, marcação das versões
- Estimativa de trabalho futuro
 - 15/20 frases em 4 horas
- Estimativa de tempo de aprendizagem
 - dois-três meses

27 de Junho de 2002

5



Resultado: problemas

- Como avaliar?
- Como comparar automaticamente?
- Como proceder a testes inter-anotadores e obter um resultado razoável?
- Sugestão: avaliar a Floresta como está
 - usá-la na avaliação conjunta de tantas quantas áreas for possível

27 de Junho de 2002

6

Resultados: experiência de avaliação



- Muito importante definir o que avaliar
 - mais proveitoso definir sub-tarefas mais simples
- Importante procurar uma métrica intuitiva
 - 3 anotadores - se 1 difere, duas diferenças (0,2,3)
 - qual a unidade (frase, palavra, constituinte)
- Qual o ponto de vista das contagens?
 - 2:1, 1:2 como contar o número de frases “certas”