

Anotação inglesa do COMPARA com o CLAWS C7: critérios de revisão

Rosário Silva e Ana Frankenberg-Garcia

Última actualização: 4 de Novembro de 2008

A anotação gramatical automática do COMPARA foi efectuada usando o CLAWS -- Tagset C7. Presentemente, o corpus está a ser submetido a revisão humana¹ com base nas directrizes de Post-Editor's Guide to CLAWS7 Tagging (Wynne 1996).² O presente documento contém a descrição dos poucos casos em que a revisão da anotação do COMPARA divergiu das directrizes acima mencionadas.

As únicas divergências verificadas até à data resultam da nossa decisão de classificar da mesma maneira que foram classificados na parte portuguesa do corpus:

1. Unidades multi-palavras consideradas como nomes próprios
2. Nomes comuns que começam com letra maiúscula.

1. Unidades multi-palavras consideradas como nomes próprios

No COMPARA, as unidades multi-palavras consideradas como nomes próprios singulares foram classificadas como NP1 e as unidades multi-palavras consideradas como nomes próprios plurais foram classificadas como NP2. Este critério diverge daquele que foi utilizado por Wynne (1996), em que cada elemento da unidade multi-palavra foi classificado separadamente. Veja-se os exemplos seguintes:

1.1 Nomes de pessoas

no COMPARA	nas directrizes CLAWS em Wynne (1996)
[Raimundo Silva] NP1	[Raimundo] NP1 [Silva] NP1
[Philip Swallow] NP1	[Philip] NP1 [Swallow] NP1

¹ As informações relacionadas com o progresso da revisão encontram-se disponíveis em Revision Progress.xls.

² <http://www.natcorp.ox.ac.uk/docs/claws7.html>

1.2 Nomes de produtos, empresas, instituições, jornais, etc.

no COMPARA	nas directrizes CLAWS em Wynne (1996)
[Three Nuns Empire Blend] NP1	[Three] MC [Nuns] NN2 [Empire] NN1 [Blend] NN1
[Red Cross] NP1	[Red] JJ [Cross] NN1
[Saturday Evening Post] NP1	[Saturday] NPD1 [Evening] NNT1 [Post] NN1

1.3 Nomes de lugares (ex. países, cidades, vilas e aldeias, moradas) e nomes geográficos

no COMPARA	nas directrizes CLAWS em Wynne (1996)
[Oxford Street] NP1	[Oxford] NP1 [Street] NNL1
[Pacific Ocean] NP1	[Pacific] NP1 [Ocean] NNL1

1.4 Títulos de livros, peças, filmes, etc.

no COMPARA	nas directrizes CLAWS em Wynne (1996)
[Madame Bovary] NP1	[Madame] NNB [Bovary] NP1
[Book of Psalms] NP1	[Book] NN1 [of] IO [Psalms] NN2
[Charles Boon Show] NP1	[Charles] NP1 [Boon] NP1 [Show] NN1

1.5 Nomes religiosos

no COMPARA	nas directrizes CLAWS em Wynne (1996)
[Holy Ghost] NP1	[Holy] JJ [Ghost] NN1
[Hail Marys] NP2	[Hail] NN1 [Marys] NP2
[Church of Christ] NP1	[Church] NN1 [of] IO [Christ] NP1
[Holy Wars] NP2	[Holy] JJ [Wars] NN2

1.6 Nomes de animais, espécies, etc.

no COMPARA	nas directrizes CLAWS em Wynne (1996)
[Black Orpingtons] NP2	[Black] JJ [Orpingtons] NP2
[Old Dutch Bantams] NP2	[Old] JJ [Dutch] JJ [Bantams] NP2

2. Nomes comuns que começam com letra maiúscula

No COMPARA, os nomes comuns singulares que começam com letra maiúscula foram considerados nomes próprios e classificados como NP1. Da mesma forma, os nomes comuns plurais que começam com letra maiúscula foram classificados como NP2.

Quando os nomes com letra maiúscula (tanto singulares como plurais) se encontram no início da frase, usamos o contexto para solucionar qualquer tipo de ambiguidade. Este critério difere daquele utilizado por Wynne (1996), em que os nomes comuns com letra maiúscula não são considerados nomes próprios. Veja-se os exemplos seguintes:

no COMPARA	nas directrizes CLAWS em Wynne (1996)
[Father] NP1 [Son] NP1 [and] CC [Holy Spirit] NP1	[Father] NN1 [Son] NN1 [and] CC [Holy] JJ [Spirit] NN1
[Romanticism] NP1	[Romanticism] NN1
[Burger Kings] NP2	[Burger] NN1 [Kings] NN2