

COMPARA's English annotation with CLAWS

C7: revision criteria

Rosário Silva & Ana Frankenberg-Garcia

Last update: 4 November 2008

The automatic grammatical annotation of COMPARA was carried out using CLAWS -- Tagset C7. The corpus is currently undergoing human revision¹ using the Post-Editor's Guide to CLAWS7 Tagging (Wynne 1996).² This document contains a description of the few cases in which COMPARA's wordclass tagging revision has diverged from the aforesaid CLAWS7 guidelines.

The only divergences up to the present result from our decision to classify the following in line with the way they have been classified in the Portuguese part of the corpus:

1. Multiword units that read as proper nouns
2. Common nouns that begin with a capital letter.

1. Multiword units that read as proper nouns

In COMPARA, multiword units that read as a singular proper noun have been marked NP1 and multiword units that read as plural proper noun have been marked NP2. This diverges from Wynne (1996), where each element of the multiword unit has been tagged separately. Compare the following examples:

1.1 Names of people

in COMPARA	in Wynne (1996) CLAWS guidelines
[Raimundo Silva] NP1	[Raimundo] NP1 [Silva] NP1
[Philip Swallow] NP1	[Philip] NP1 [Swallow] NP1

¹ Information regarding the revision progress is available on Revision Progress.xls.

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

1.2 Names of products, companies, institutions, newspapers, etc.

in COMPARA	in Wynne (1996) CLAWS guidelines
[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 Names of places (e.g. countries, towns and villages, addresses) and geographical names

in COMPARA	in Wynne (1996) CLAWS guidelines
[Oxford Street] NP1	[Oxford] NP1 [Street] NNL1
[Pacific Ocean] NP1	[Pacific] NP1 [Ocean] NNL1

1.4 Titles of books, plays, films, etc.

in COMPARA	in Wynne (1996) CLAWS guidelines
[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 Religious names

in COMPARA	in Wynne (1996) CLAWS guidelines
[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 Names of animals, species, etc.

in COMPARA	in Wynne (1996) CLAWS guidelines
[Black Orpingtons] NP2	[Black] JJ [Orpingtons] NP2
[Old Dutch Bantams] NP2	[Old] JJ [Dutch] JJ [Bantams] NP2

2. Common nouns that begin with a capital letter

In COMPARA, singular common nouns that begin with a capital letter have been considered proper nouns and marked NP1. Likewise, plural common nouns that begin

with a capital letter have been marked NP2. When capitalized common nouns (both singular and plural) occur in sentence initial position, we use context to resolve any ambiguity. This diverges from Wynne (1996), where common nouns written with an initial capital letter are not considered to be proper nouns. Compare the following examples:

in COMPARA	in Wynne (1996) CLAWS guidelines
[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