The AnCora corpus has 500,000 words in 17,000 sentences, annotated in a CFG style with syntactic annotations and argument structures. We transformed them into a HPSG annotation style with the aim of creating a statistical HPSG parser for Spanish.

Our approach is based on the construction of the Enju parser, they transformed the Penn Treebank into a HPSG style corpus and learned a statistical HPSG model from it.

The aim is to build a HPSG corpus that contains morphological, syntactic and simplified semantic features. Our approach to semantics is based on the PropBank style argument structure annotations found in the corpus.

The leaves are extracted as lexical frames, keeping only combinational features.