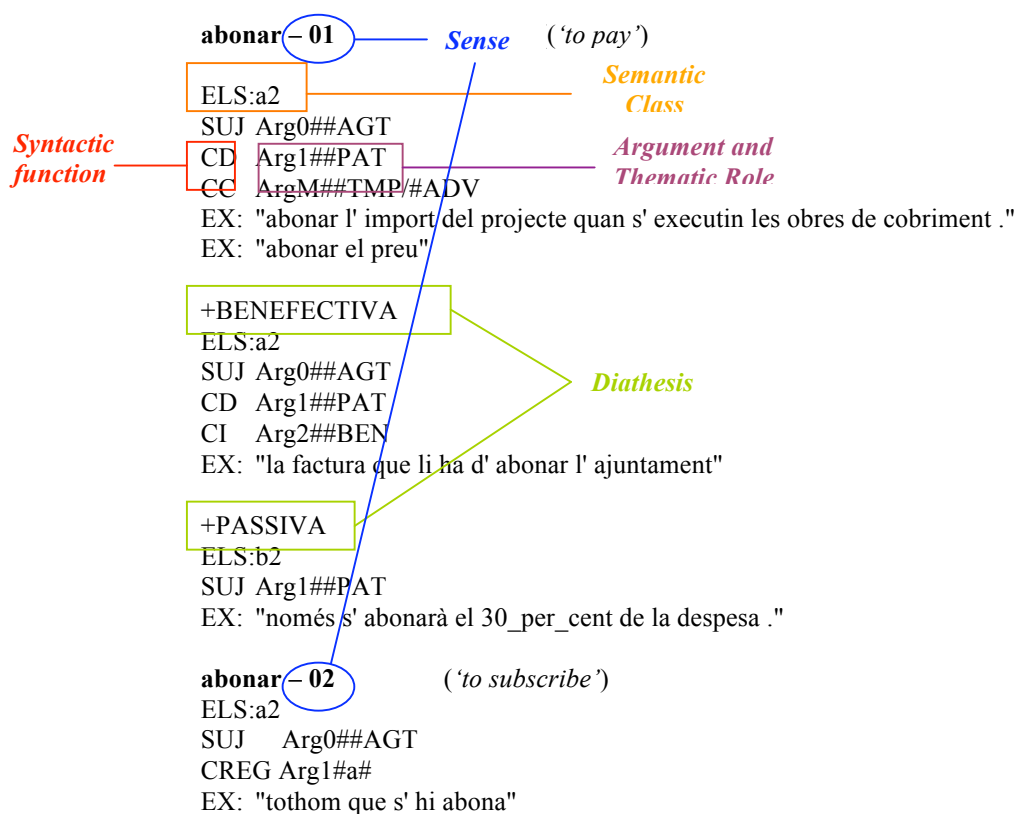# AnCora: Verbal lexicons

AnCora verbal lexicons (named **verbal-lexicon.ca** for Catalan and **verbal-lexicon.es** for Spanish in this shared task distribution) were obtained deriving, for each sense of each verb, all the syntactic schemata in which a verbal predicate appears in AnCora corpora. From this information, the mapping from syntactic functions to thematic roles, and the corresponding argument position, was manually declared in the lexicons.

In these verbal lexicons, each verbal predicate may be divided into different senses (01, 02, 03, …), where each sense is related to one or more semantic classes (Lexical Semantic Structures, ELS in Catalan/Spanish), basically differentiated according to the four event classes ('accomplishments', 'achievements', 'states' and 'activities'), and on the diatheses alternations in which a sense can occur. In the figure below the full information associated with the entry *abonar* ('to pay, to subscribe' in English) in the Catalan lexicon is shown.

Following *PropBank*, the arguments selected by the verb are incrementally numbered (Arg0, Arg1, Arg2, Arg3, Arg4) expressing their degree of proximity in relation to its predicate. The adjuncts are labelled as ArgM. The list of thematic roles consists of 19 different thematic labels: AGT (Agent), CAU (Cause), EXP (Experiencer), SCR (Source), PAT (Patient), TEM (Theme), ATR (Attribute), BEN (Beneficiary), EXT (Extension), INS (Instrument), LOC (Locative), TMP (Time), MNR (Manner), ORI (Origin), DES (Goal), FIN (Purpose), EIN (Initial State), EFI (Final State) and ADV (Adverbial).

The Catalan and Spanish lexicons contain a total of 2,142 and 2,580 verbs, respectively.