Introduction

- **Objective**: creation and management of the Basque Dependency Treebank
  - Main characteristics:
    - Makes the annotation process faster
    - Avoids possible mistakes
    - Implemented in Java and multiplatform
    - Friendly interface and language independent
  - Main areas:
    - The corpus area
    - The tagging area
    - The tree visualizer area

The Basque Dependency Treebank

- **General project**: annotation of corpora at syntactic, semantic and pragmatic levels in Catalan, Spanish and Basque
  (http://www.dlsi.ua.es/projectes/3lb)
- Grammatical relations specifying dependencies between modifiers and their nucleus
- **Tagset**:
  - Adaptation of Carroll et al. (1998, 1999)
  - Difference: Arguments not lexicalised (phonetically empty pro)

Abar-Hitz

- **Previously analysed tools**
  - Annotation tools:
    - WordFreak (Morton and LaCivita, 2003)
    - Our annotation formalism not supported
  - Tree management tools:
    - TreeTrans (AGTK)
    - Based in constituents
  - TrEd (Prague Dependency Treebank)
  - Dependency tags in nodes (as attribute) but not in the connectors between nodes
  - The Graph Tree Editor Tool
  - Dependency tags in nodes but not in the connectors between nodes
  - TreeScape
  - Draws not editable trees
  - CM-ED (Arruarte et al., 2001)
  - Concept map editor adapted into ESALT, a tree visualizer that follows a dependency-based formalism

Example

Noizean behin, Lisboako kaleetan zehar kresal usain gazia sumatzen da

“From time to time, the salty scent of seawater can be perceived in the streets of Lisbon”

Conclusions and Future work

Conclusions

- Makes the annotation process faster and avoids mistakes
- Massively used by three linguists in the annotation of a treebank of 50,000 word-forms
  - One half of the corpus revised with Abar-Hitz
  - The other half tagged with Abar-Hitz

Future Work

- EULIA (Artola et al., 2004) a tool for creating, consulting, visualizing and modifying documents in XML will be integrated in Abar-Hitz
- Abar-Hitz will give the output, the syntactic analysis, in an XML document that will be compared to the document produced by the parser