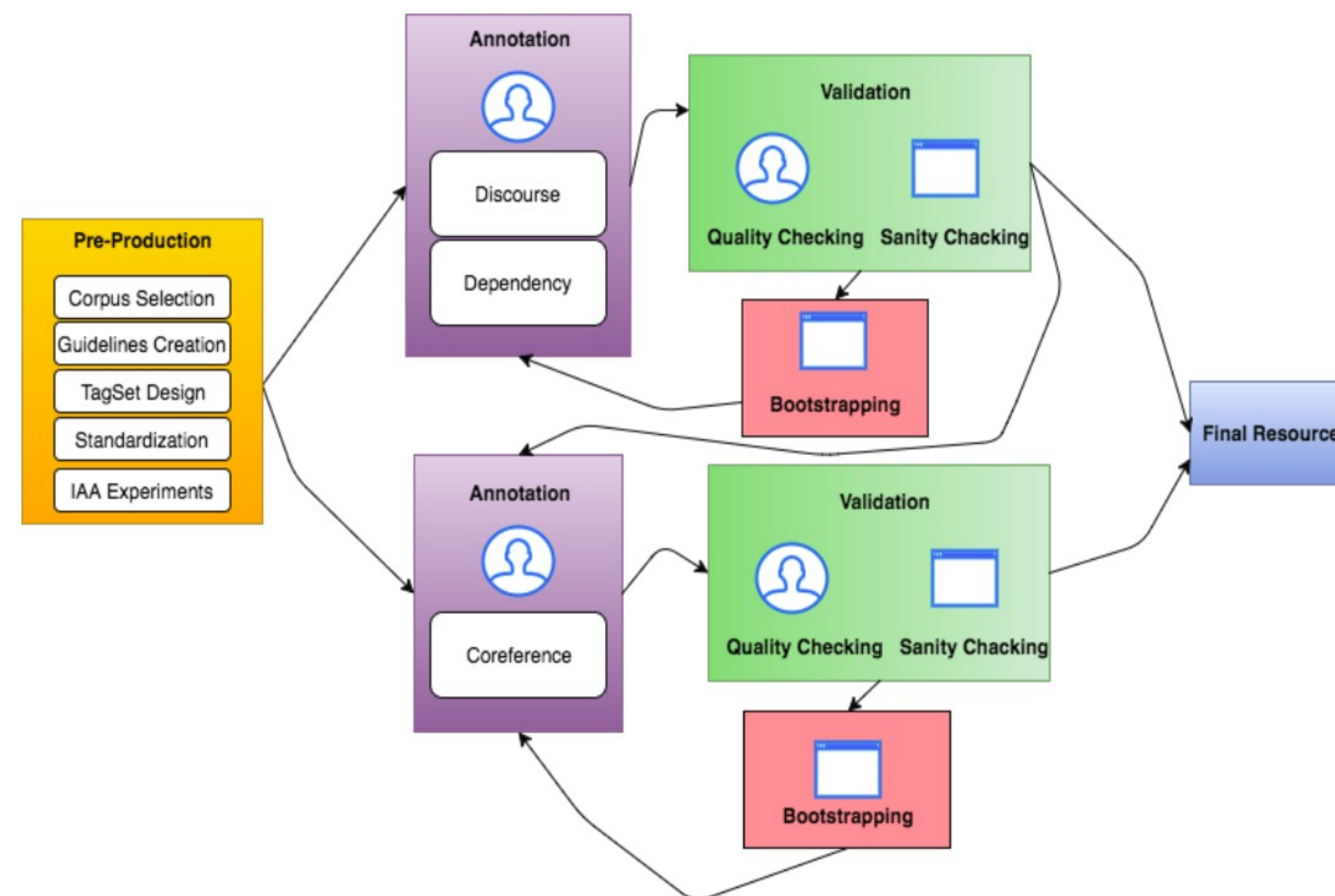




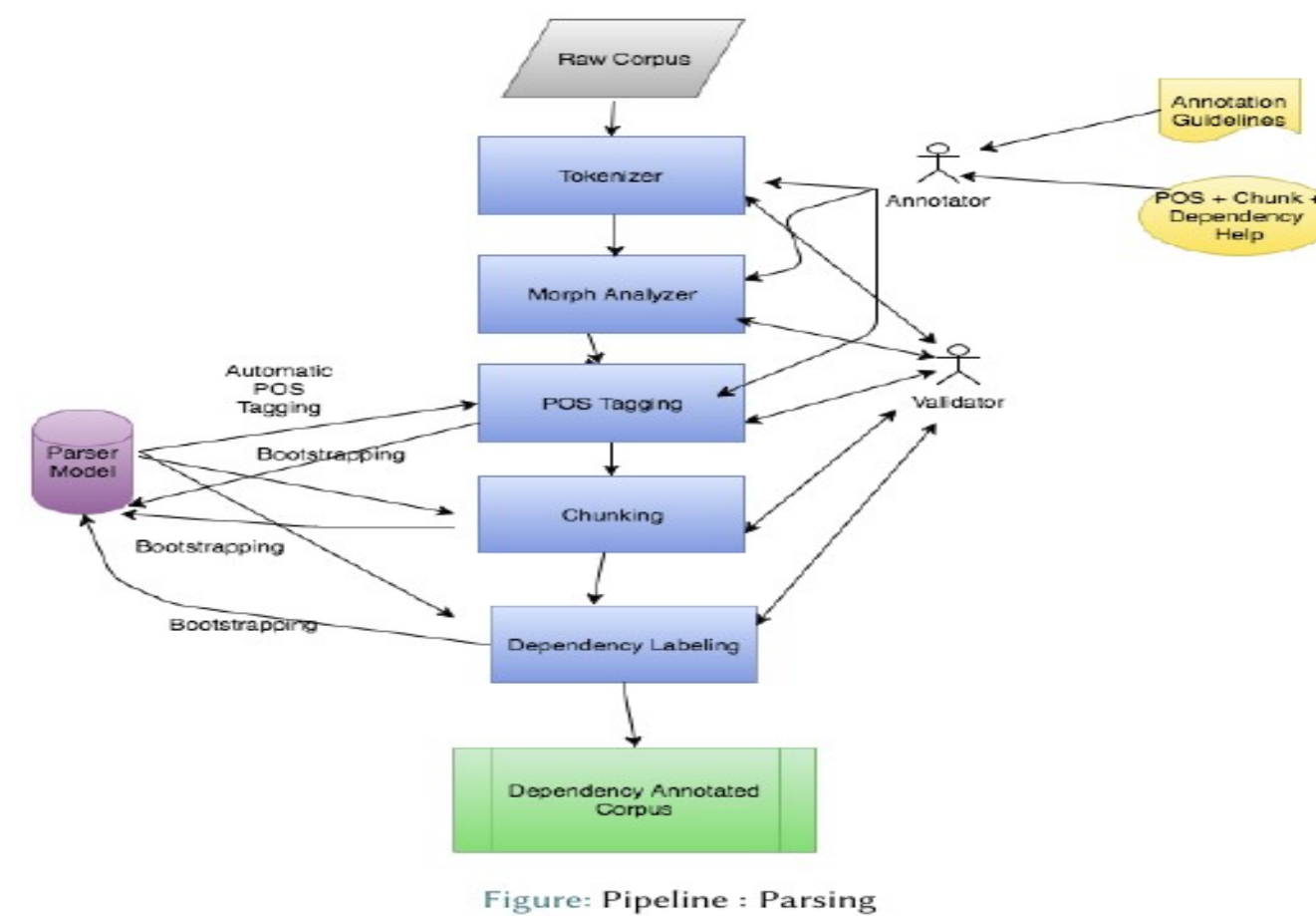
Resource Creation Annotation

For developing really intelligent systems one needs richer knowledge sources. Large amount of raw corpora, sufficient quantity of annotated corpora etc play an important role there. Once crucial information is marked and validated in a corpus, it can be used in various ways to extract the annotated information. It also makes a basic resource for ML. The goal here is to create 'right' quantities of high quality of different types of corpora (annotated, unannotated as the need



Dependency Annotation

Dependency Annotation is done on the raw corpus that includes annotation at various levels from Tokenization to Dependency Labeling. Dependency Annotated Data (Treebank) is prerequisite for most of the NLP applications.



Coreference Annotation

Coreference annotation is done on top of the dependency annotated data which helps in Machine Translation, Dialogue Systems, Q & A Systems, Summarization, many other NLP applications.

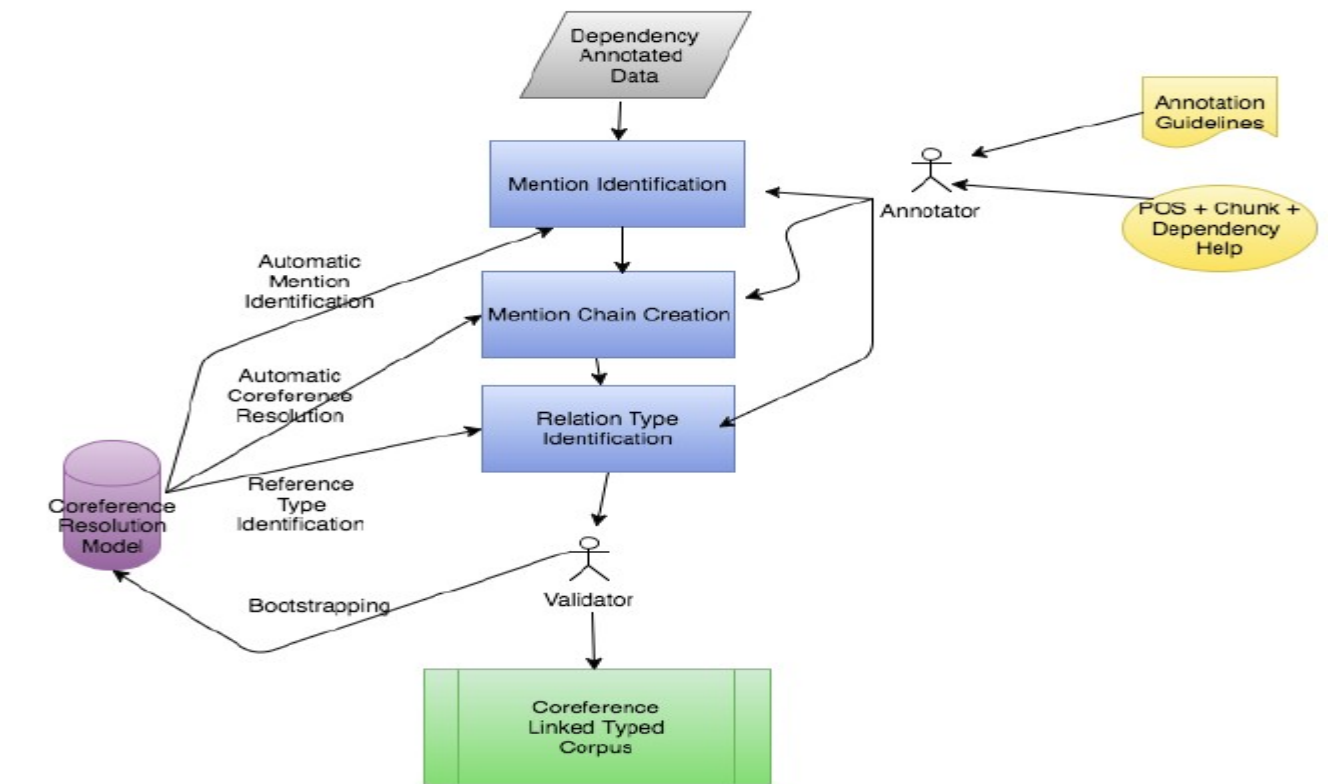


Figure: Pipeline : Coreference

SWAYAM Lectures Translation

Making English video lectures available in Indian languages (ILs) involving Transcription, Translation and Subtitling for English and ILs

