Transition Based Dependency Parser for Amharic Language Using Deep Learning

Mizanu Zelalem Degu, Jimma University,

Million Meshesha(PhD), Addis Ababa University

Introduction

- Amharic is one of under-resourced languages that needs linguistic tools to be developed, one of them is dependency parser.

- Researches, like malt parser, were attempted to develop universal dependency parser, but have been discovered weak in parsing morphological rich languages, like Arabic, Hebrew and Amharic.

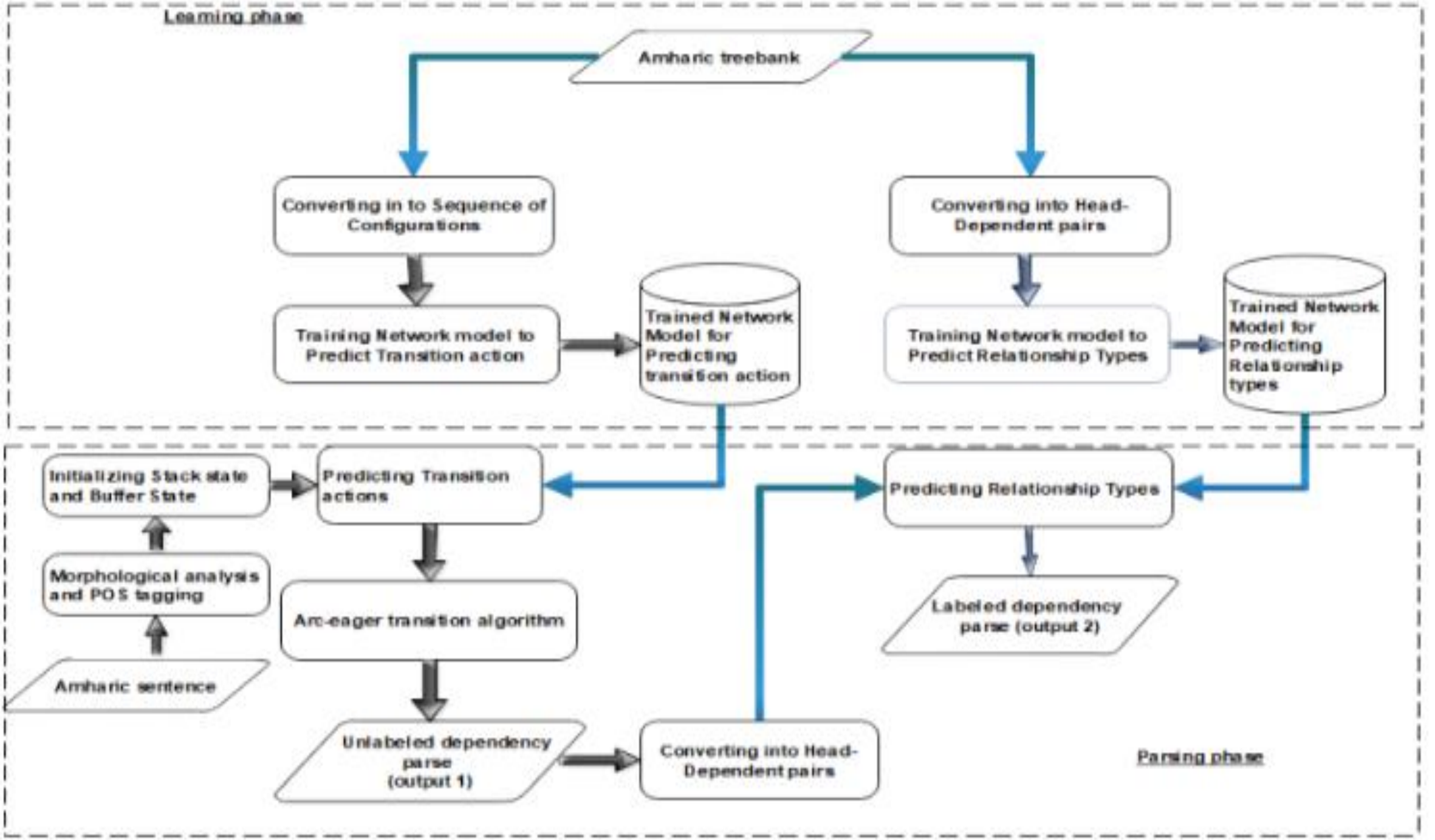
Cont.

- The first dependency parser developed for Amharic was a dependency grammar (a rule based approach using XDG), developed by M. Gasser.

- A motivational work on developing dependency parser for Amharic language was done by Seyoum et al. They developed The first Amharic tree-bank which contains 1074 sentences.

Methodology

- The aim of this study was to develop a dependency parser for Amharic language using the Arc-Eager transition rules.
  - The architecture of the system is shown in the next slide

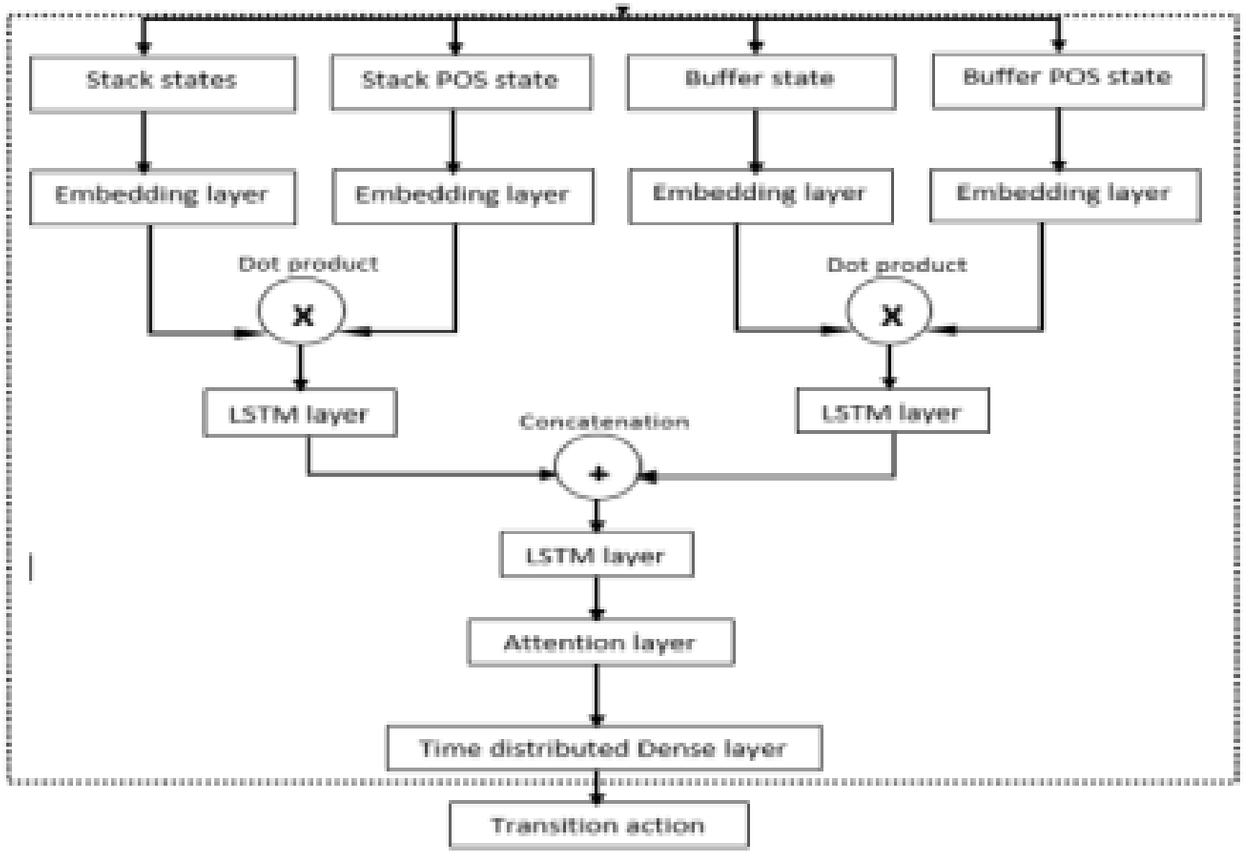

Learning phase
Amharic treebank
Converting in to Sequence of Configurations
Converting into Head-Dependent pairs
Training Network model to Predict Transition action
Trained Network Model for Predicting transition action
Training Network model to Predict Relationship Types
Trained Network Model for Predicting Relationship types
Initializing Stack state and Buffer State
Predicting Transition actions
Predicting Relationship Types
Morphological analysis and POS tagging
Arc-eager transition algorithm
Labeled dependency parse (output 2)
Amharic sentence
Unlabeled dependency parse (output 1)
Converting into Head-Dependent pairs
Parsing phase

Cont.

As depicted in the architecture, in the training phase, two network models are trained.

1. For predicting one of the four arc-eager transition actions
2. For predicting relationship types

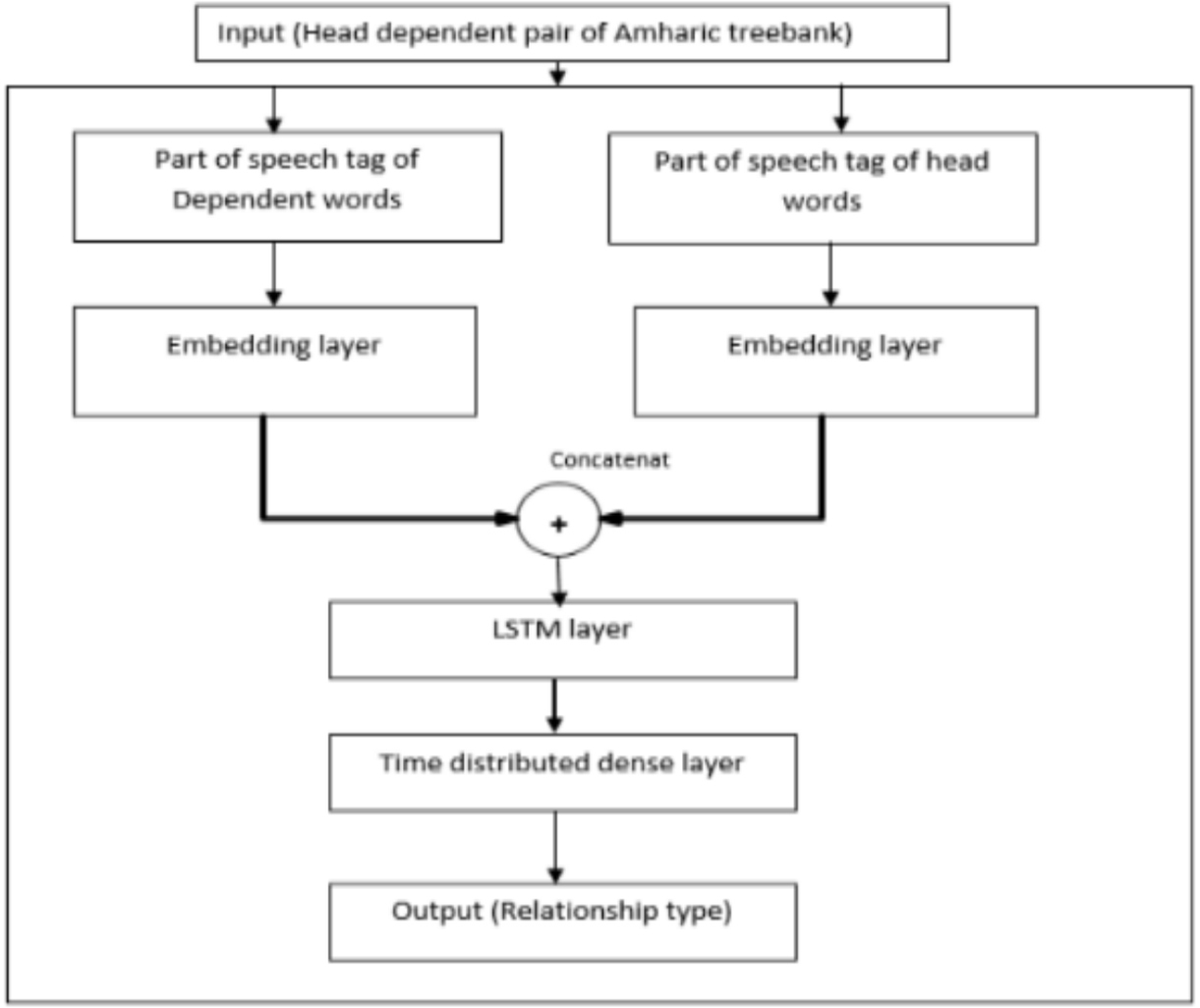
1. Network model for predicting Transition actions.



- A normal tree-bank cannot be an input for the network model.
- Therefore we designed an algorithm to convert tree-bank into sequences of arc-eager transition configuration

2. Network model for Predicting Relationship type

- The importance of this Network model is to reduce the number of classes in constructing labeled dependency structure.
- It helps to reduce classes from 2n+2 to n, where n is number of relationship types in the language.
- Input for this network model is list of head-dependent pairs from Amharic tree-bank



Evaluation of The parser

- A tree-bank with 1574 Amharic sentences is used for training (70%) and testing (30%) the parser.

1. Evaluation of transition action predictor.

- Converting the tree-bank into arc-eager transition configuration gives 26,242 configurations.
- The predictor correctly classified 94.7% of transition configurations.

2. Evaluation of relationship type predictor

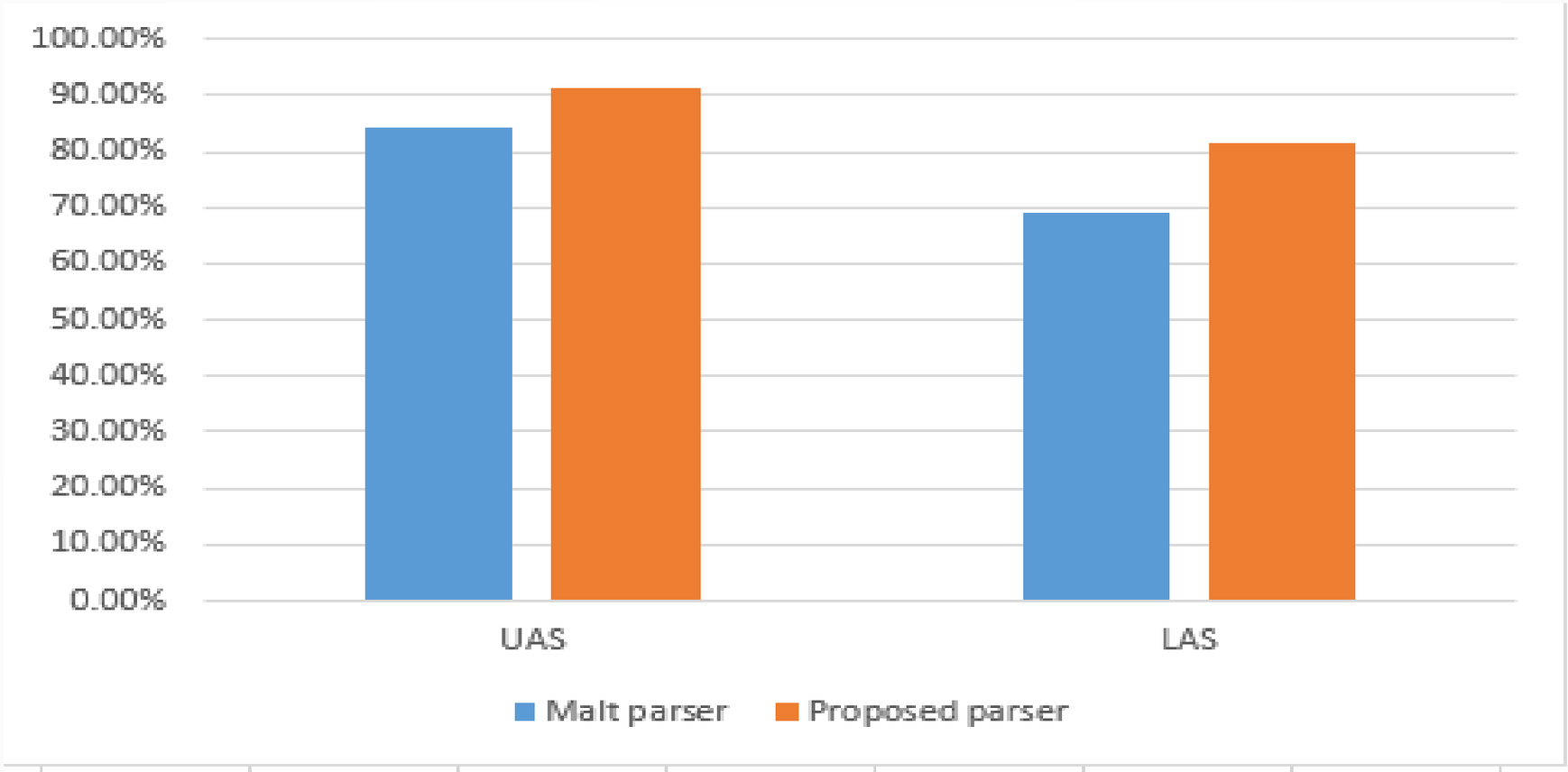
- Converting the tree-bank into head-dependent pair gives 15,534 head-dependent pairs.
- The predictor correctly classifies relationship type of 82.4% head-dependent pairs.

Cont.

3. Attachment scores

- Unlabeled Attachment score (UAS) = 91.54 %
- Labeled Attachment Score (LAS)= 81.4 %

4. Comparison of the parser with Malt parser



Conclusion

In this paper, we implemented a dependency parser system for Amharic language. The parser is developed based on arc-eager transition action and additional network model for predicting relationship types. The use of the second network model is to increase the number of examples for each class (relationship types) from the tree bank and increase the accuracy of labeled attachment score. The system is evaluated on Amharic treebank and results 91.54% and 81.4% for unlabeled and labeled attachment scores respectively. From the experiment, we noted that, the system can perform better even when the size of the tree bank increases.