

*Tomas By (Lissabon, Portugal)*

## **The TiGer dependency bank in Prolog format**

### **Abstract**

The TiGer dependency bank is a conversion of 1867 sentences from the TiGer Treebank to a dependency format based on the one used in the PARC 700 dependency bank. This means that the tokens are lemmatised, and that the word order is not represented, so that multiple occurrences of the same word type in one sentence are ambiguous. The number of such ambiguous word tokens in the TiGer dependency bank is around 9%, compared to 15% in PARC 700. Other problems with this format include a spurious distinction between 'full' tokens and attribute tokens (inherited from LFG), the absence of a distinction between words and empty nodes in the dependency tree, and a unnecessarily distributed representation of the token attributes.

This paper reports a conversion of the TiGer dependency bank to a Prolog representation that represents the word order explicitly, uses the surface forms of the words rather than base forms, makes a clear distinction between words and empty nodes, and stores the token attributes in one place rather than spread out across the file. Together with the Prolog format data, graphical representations of the dependency trees are provided in PDF. Hopefully this work will make it easier to use the TiGer dependency bank data for parser evaluation of German.