Identification of Multiword **BBN NEs** in **Dependency** Annotation of **WSJ**

ÚFAL

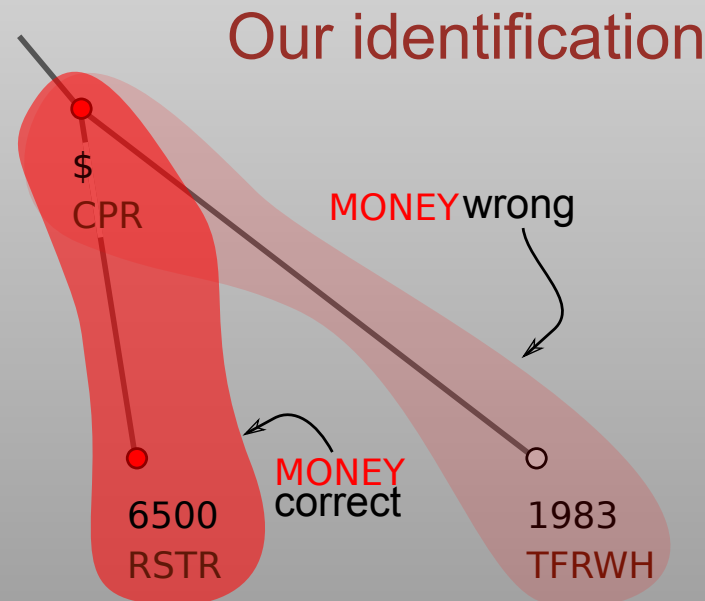Eduard Bejček, Pavel Straňák   Charles University in Prague

- **Input:** dependency annotation of WSJ with BBN Named Entities
  (Wall Street Journal → Penn Treebank → Prague Czech-English Dependency Treebank)

- **Procedure:** building a "lexicon" of MWNEs from the train data; identification of MWNEs in the test data

- **Results:** - missed, "out-of-vocabulary" entities (as expected); - analysis of other, more interesting mistakes

- **Examples:**   BBN annotation   Our identification

1) *Following the acquisition earlier this year...*

DATE
correct?

2) *...with the farm price index hitting record peaks earlier this year...*

$
CPR

MONEY wrong

...compared with $ 6500 in 1993.

6500
RSTR

MONEY
correct

1983
TFRWH