



An Analysis of Annotation of Verb-Noun Idiomatic Combinations in a Parallel Dependency Corpus

Zdeňka Urešová, Jana Šindlerová, Eva Fučíková and Jan Hajič

Charles University in Prague
Institute of Formal and Applied Linguistics

*(Supported by grant the Grant No. GPP406/13/03351P of the Grant Agency of the Czech Republic
Data provided by LINDAT/Clarin, Infrastructural Project MEYS CR No. LM2010013)*



2nd general meeting, 10–11 March 2014, Athens, Greece

Translation of Verbal Idioms in the PCEDT



1. Eng VNIC (DPHR) → Cz VNIC (DPHR):

*It is easy to see why the ancient art **is on the ropes**.*

→ *Je snadné pochopit, proč toto starobylé umění **visí na vlásku**.*

[hangs on a-hair]

2. Eng VNIC (DPHR) → Cz non-VNIC

*Petroleum futures were generally higher with heating oil **leading the way**.*

→ *Termínové obchody ... byly ... u topného oleje, který **určoval směr**. PAT*

which [determined the-direction]

3. Eng non-VNIC → Cz VNIC (DPHR)

- *Friends told her she **was pushing too hard**.MANN*

→ *Přátelé jí říkali, že moc **tlačí na pilu**.*

too-much [presses onto the-saw]



Corpus Analysis

Total: 92890 (aligned) verbs

Total: 1300 VNICs (Czech side)

- 88 Cz VNIC (DPHR) → Eng VNIC (DPHR)
 - Plus 30 annotated as 'light verb construction' on either side
- 278 Cz non-VNIC → Eng VNIC (DPHR)
- 570 Cz VNIC (DPHR) → Eng non-VNIC