

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 (ÚFAL MFF UK)



Valency and MWEs

Valency: ability of words to combine themselves with other lexical units

FGD Valency theory

MWEs: "idiosyncratic interpretations that cross word boundaries" (Sag et al., 2002)

Verbal MWEs

VNICs – *break a leg, hit it off*
LVCs – *take a walk, give a kiss*

Valency in the PDTs

For Czech → PDT-Vallex

<http://ufal.mff.cuni.cz/lindat/PDT-Vallex.html>

For English → EngVallex:

<http://ufal.mff.cuni.cz/lindat/EngVallex.html>

PDTs ← (ID -> val. frame) → LEXICONS

to give: ACT(.1) PAT(.4) ADDR(.3)

MWEs in the PDTs

Need for formal specification:

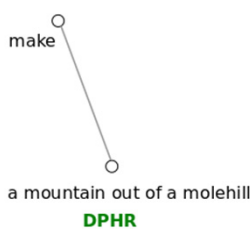
- morphological, syntactic and semantic properties

Multilayer specification of verbal MWEs:

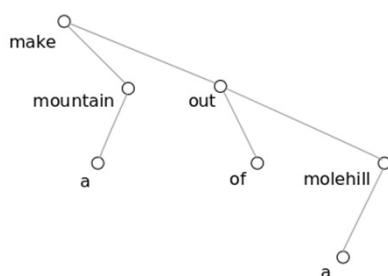
- Surface: more nodes, reflecting dependency syntax
- Deep: VNICs: verb + a single DPHR node
LVCs: verb + CPHR node (+ its dependents)

make a mountain out of a molehill

Deep Tree



Surface Dependency Tree



MWEs in the PDTs Valency Lexicons

DPHR -> formal syntax specification in the valency frame:

• *make a mountain out of a molehill*

DPHR(mountain.Obj.sg[a],out[of,molehill.Adv.sg[a]])

• *běhá mi mráz po zádech get shiver down my back*

DPHR(mráz.S1,po-1[záda.P6])

• *chodit kolem horké kaše beat about the bush*

DPHR(kolem-1[kaše.S2[horký.#]])

Translation Patterns in the PCEDT

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

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]

They are barking up the wrong tree, because it is basically their fault they can't attract new employees.

→ *Pláčí však na špatném hrobě, protože ...*

[they-cry however on a-wrong grave]

Eng VNIC (DPHR) → Cz non-VNIC

Mr. Pickens made considerable political hay with his troubles in Japan.

→ *Pickens se ... v Japonsku způsobil značný politický rozruch.*

[caused significant political turmoil]

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

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

which [determined the-direction]

Eng non-VNIC → Cz VNIC (DPHR)

Newsweek, trying to keep pace.

→ *... Newsweek, který se snaží držet krok s ... časopisem Time*

[is trying to-keep pace]

Friends told her she was pushing too hard.

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

too-much [presses onto the-saw]

Conclusion and Future Work

Treatment of VNICs

in the PDTs

in PDT-Vallex and EngVallex

PCEDT – corpus analysis

VNICs mostly **NOT** translated as VNICs

E.g. light verb constructions,
compositional phrases

Future: (more) MWEs in translation