



---

# PLIN021 SÉMANTICKÁ ANALÝZA V PRAXI

ZUZANA NEVĚŘILOVÁ

2020/21

# FORMALISMY PRO ZACHYCENÍ VÝZNAMU (POUŽÍVANÉ V NLP)



(Sémantické) sítě,  
ontologie

Rozlišení významu  
Odvozování



Sémantické rámce

Sloty, (slovesná) valence  
Výběrová omezení



(Masivní) propojení znalostí  
o jazyce a o světě



Vektorová  
reprezentace

Užití slova v kontextu

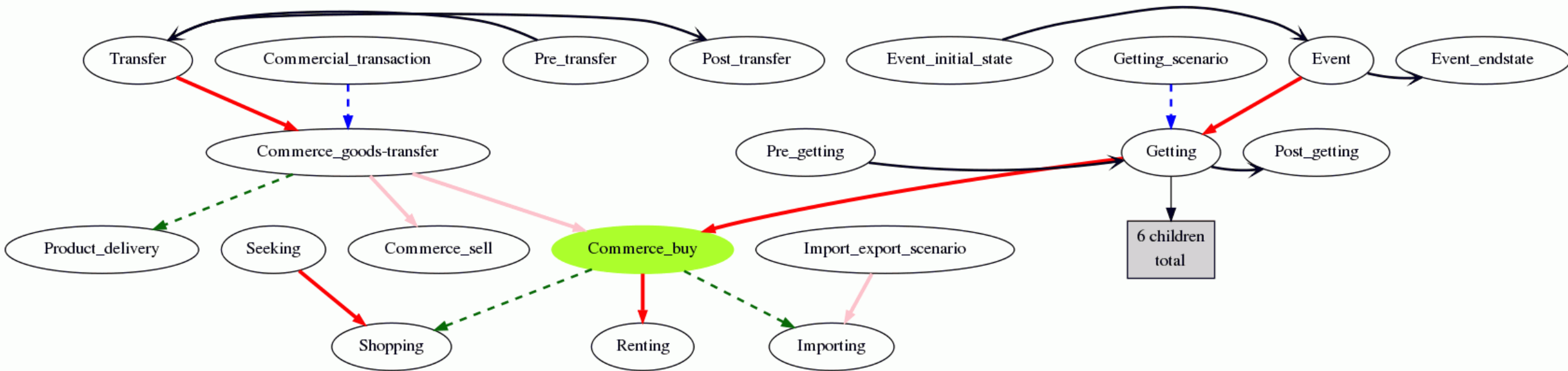
## Víme opravdu všechno?

- Propojení rámců do větších celků
- Větná koherence: elipsy, anafory, ...
- Porušení norem
  - Hovorový jazyk, „Internet language“
  - Kreativita, hravost, neologismy
- Tvrzení, logika, vztah k realitě
  - Logická analýza

# PROPOJENÍ SLOVESNÝCH RÁMCŮ

- **FrameNet** (Charles Fillmore, 1997–současnost)
- Rámec = reprezentace typické situace
  - Prvek rámce (Frame Element):  
Účastníci, propriety, role
  - Lexikální jednotky (Lexical Unit)
- 1 200 sémantických rámců
  - Některé jsou spolu **propojeny**
- 13 000 lexikálních jednotek
- <https://framenet.icsi.berkeley.edu/fndrupal/>
- Dostupný také přes Natural Language Toolkit (NLTK)
- Zobrazení dat pomocí strukturovaného textu
- Vizualizace propojení rámců
- Framenety pro korejštinu, japonštinu, francouzštinu, němčinu, portugalštinu, španělštinu, švédštinu

# PROPOJENÍ SLOVESNÝCH RÁMČŮ



John **sold** a car to Mary = Mary **bought** a car from John

# RÁMCE: DOPLNĚNÍ IMPLICITNÍ INFORMACE

*Koupila jsem ojetou felicii. Byly to vyhozené peníze.*

koupit:

- má\_činitele člověk/instituce/skupina
- má\_benefaktora člověk/instituce/skupina
- má\_předmět výrobek/nemovitost/zvíře/rostlina/přírodnina
- má\_část činitel dá peníze
- má\_část benefaktor dá předmět

Poměr explicitní : implicitní informace v textu je až 1:8. (Graesser, 1981)

# ZPÁTKY V ČASE: SKRIPTY, SCÉNÁŘE (ABELSON)

V restauraci (prvky skriptu mohou být rámce)

- host (člověk, není v zaměstnání, má u sebe *peníze*, sedí na *židli*, jí *jídlo*)
- číšník (člověk, je v zaměstnání)
- kuchař (člověk, je v zaměstnání)
- místnost (obsahuje *židle*, stoly, příjemnou teplotu)
- jídlo (uvařil *kuchař*, donesl *číšník hostovi*)
- peníze (zaplatil *host číšníkovi* za *jídlo*)

Pepovi u večeře zazvonil telefon. Chvíli poslouchal, pak položil telefon a opustil restauraci.

Předpokládáme, že mezi „položil telefon“ a „opustil restauraci“ se stalo:

- Číšník donesl účet.
- Pepa zaplatil.
- Pepa se oblékl

Usuzování z rámců  
Vybočení ze  
stereotypu

# CO VÍME O VÝZNAMU

- Význam je idea v hlavě, nelze jej definovat klasicky.
- Významy lze nejlépe interpretovat pomocí pravděpodobnosti, s jakou se užití blíží prototypu.
  - Prototypy: Rosch
  - Scénáře Schank a Abelson
  - Rámce: Fillmore
- Významy jsou spojeny se vzory (Hunston a Francis, 2000), konstrukcemi (Goldberg), „frazémy“ (Melčuk) či slovy.
  - Korpusová lingvistika dokáže tyto vzory (preference užití) zjistit a studovat
- Významy vně kontextu neexistují, existují jen významové potenciály, které se kontextem aktivují (Hanks, 2000).

# VÝZNAMOVÉ POTENCIÁLY (HANKS)

PATTERN: [[Human]] translate ([[Document]]) (from [[Language 1]]) (into [[Language 2]])

IMPLICATURE: [[Human]] expresses the meaning of [[Document]] in [[Language 1]] in the words and phraseology of [[Language 2]]

Corpus Patterns Analysis



# VÝZNAMOVÉ POTENCIÁLY (HANKS)

- norma = užití slova podle určitého syntagmatického vzoru
- slova se užívají jednak v souladu s normou (očekáváním), jednak mluvčí normu porušují
- vzory se skládají z užitého slova a z lexikálních množin, se kterými se slovo užívá
- lexikální množiny mohou být obrovské, např. [[Human]]
- čím menší lexikální množina, tím silnější je její vliv na význam vzoru
- kreativní užití jazyka – porušení normy nepřesahuje 10 % případů v korpusu
- pokud ano, je to nejspíš dosud neobjevená norma
- i porušení normy má jistá pravidla – „dvojitá šroubovice“
  - pravidla, jak slova používat normálně
  - pravidla, jak normu porušit

...možná bych prodloužila poslední verš... Whitesun a stále nedokončeným C-3PO. Cliegg atraktivní módní přehlídce. Tvůrkyně ošacení zablokovali oba směry magistrály. Někteří centrem byl zdemolován vůz TV Nova. Jinak zde hráli v sále hospody a hodně lidí zůstalo jsem byl naprosto unešena. Jsou perfektní. psala, dorazil a věcičky jsou nádherný Vše dnes mi s podprsenkami dorazil... Skvěle moc děkuji, těhotenské rifle jsou super -

**nesedí** mi tam nejen vizuálně, že je příliš krátký  
**sedí** na létajícím křesle, protože mu chybí jedna  
**seděla** na premiéře v první řadě v atraktivní róbě  
**sedí** na schodech k opeře. Mávají rudými vlajkami  
**sedí** ještě menší skupina demonstrantů. Jde většinou  
**sedět** venku, protože tentokrát počasí přálo.  
**Sedí** jako ulité. Jen délka ,ale tu jsem za 15  
**sedí** tak jak má a já jsem mooc spokojená Moc  
**sedí** . Jsem moc spokojená! Budu se těšit zase  
**sedí** úplně perfektně!!! A taky děkuju za dáreček

## PRAVIDLA ELIPSY

[[Human]] ] sedět [[Location]]  
 | sedět ] [[Manner]]  
 | sedět [[Human, dative]] [[Location]] ]

# VÝZNAM A LOGIKA

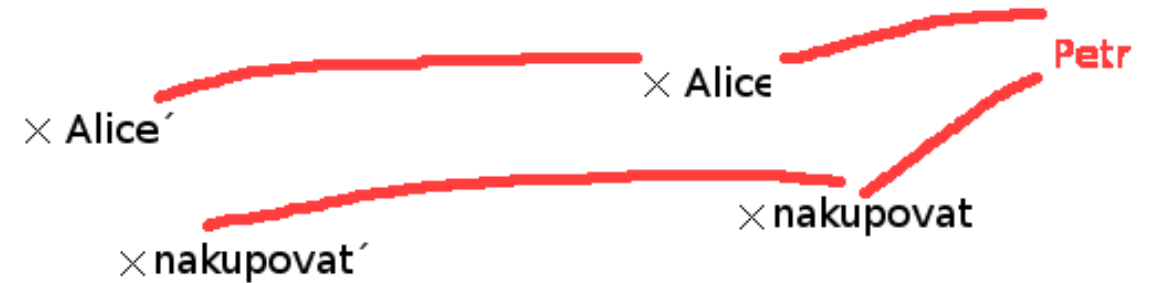
- Logika se zabývá pravdivostní hodnotou.
- Některá tvrzení jsou pravdivá nezávisle na stavu světa:
  - $2 + 2 = 4$
- Většina tvrzení může být pravdivých v určitém čase a světě:
  - Alice šla nakupovat.
- Usuzování „přes různé světy“:
  - Alice počítala, kolik je  $2+2$ .  
Alice počítala, kolik je 4.

## Formalismy:

- Výroková logika (příliš jednoduchá: výroky a logické spojky)
- Predikátová logika (proměnné, kvantifikátory, predikáty s argumenty)
- Modální logiky (možnost, nutnost)
- Fuzzy logiky
- Intenzionální logiky

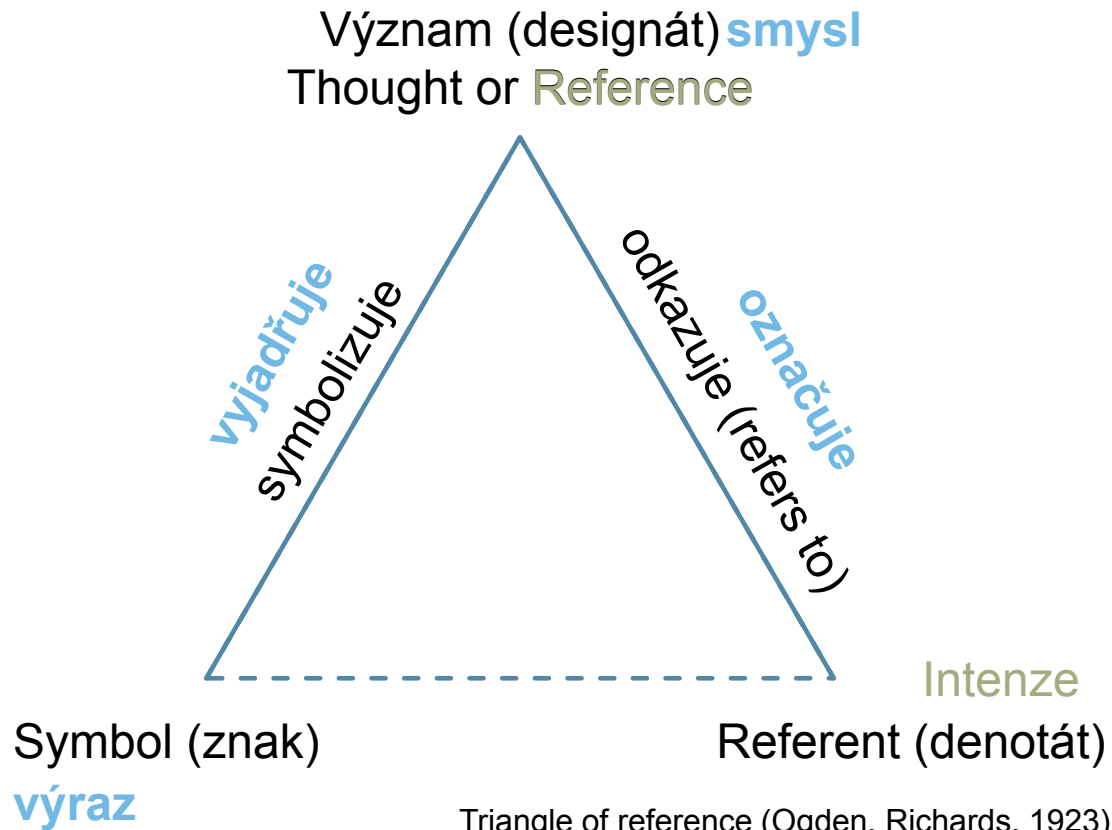
# MENTÁLNÍ PROSTORY (FAUCONNIER, 1994)

- Většina tvrzení může být pravdivých v určitém čase a světě:
  - Alice šla nakupovat.
- Některá tvrzení platí v jiném „prostoru“:
  - Petr řekl, že Alice šla nakupovat.



# TRANSPARENTNÍ INTENZIONÁLNÍ LOGIKA

Význam je cesta k intenzi.



Triangle of reference (Ogden, Richards, 1923)

- Intenze
  - Empirický výraz
  - Funkce z možných světů
- Extenze
  - Analytický výraz
  - Matematický objekt
- Konstrukce

# NEJEDNOTNÝ POPIS VÝZNAMU

## Víme opravdu všechno?

- Propojení rámců do větších celků
- Větná koherence: elipsy, anafory, ...
- Porušení norem
  - Hovorový jazyk, „Internet language“
  - Kreativita, hravost, neologismy
- Tvrzení, logika, vztah k realitě
  - Logická analýza

## Umíme (něco z toho) aplikovat?

**Ultimátní cíl NLP:** Komunikace v přirozeném jazyce

- Chatboty
- Asistenti
- Rozhraní člověk–stroj

# LITERATURA

- Hanks, P. (2000). **Do word meanings exist?** *Computers and the Humanities*, 34:205-215. 10.1023/A:1002471322828.
- Hanks, P. (2010). **Elliptical arguments: a problem in relating meaning to use.** In Paquot, M. and Granger, S., editors, *eLexicography in the 21st century : New challenges, new applications. Proceedings of eLex 2009, Louvain-la-Neuve, 22-24 October 2009*, volume 7 of *Cahiers du Cental*, pages 109-124, Louvain-la-Neuve, Belgium. Université Catholique de Louvain, Presses universitaires de Louvain
- Charles J Fillmore. 1976. **Frame semantics and the nature of language.** *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, 280:20-32. <http://www.icsi.berkeley.edu/pubs/ai/frame semantics76.pdf>
- Laparra, E. and Rigau, G. (2009). **Integrating wordnet and framenet using a knowledge-based word sense disambiguation algorithm.** In *RANLP*, Borovets, Bulgaria
- DUŽÍ, Marie; MATERNA, Pavel. *TIL jako procedurální logika: průvodce zvědavého čtenáře transparentní intensionální logikou.* Bratislava: Aleph, 2012. 428 s. (Noema: edice časopisu Organon F; sv. 7). ISBN 978-80-89491-08-7. <http://www.cs.vsb.cz/duzi/aleph.pdf>