



PLIN021 SÉMANTICKÁ ANALÝZA V PRAXI

ZUZANA NEVĚŘILOVÁ

2020-2024



OD ZKOUMÁNÍ VÝZNAMŮ K POČÍTAČOVÉ LINGVISTICE

- Význam zkoumali lidé nejméně od antiky.
- Některé myšlenky 20. století významně ovlivnily počítačové zpracování významu.

OD ZKOUMÁNÍ VÝZNAMŮ K POČÍTAČOVÉ LINGVISTICE

- Význam zkoumali lidé nejméně od antiky.
- Některé myšlenky 20. století významně ovlivnily počítačové zpracování významu.



**Proto se na ně
podíváme**

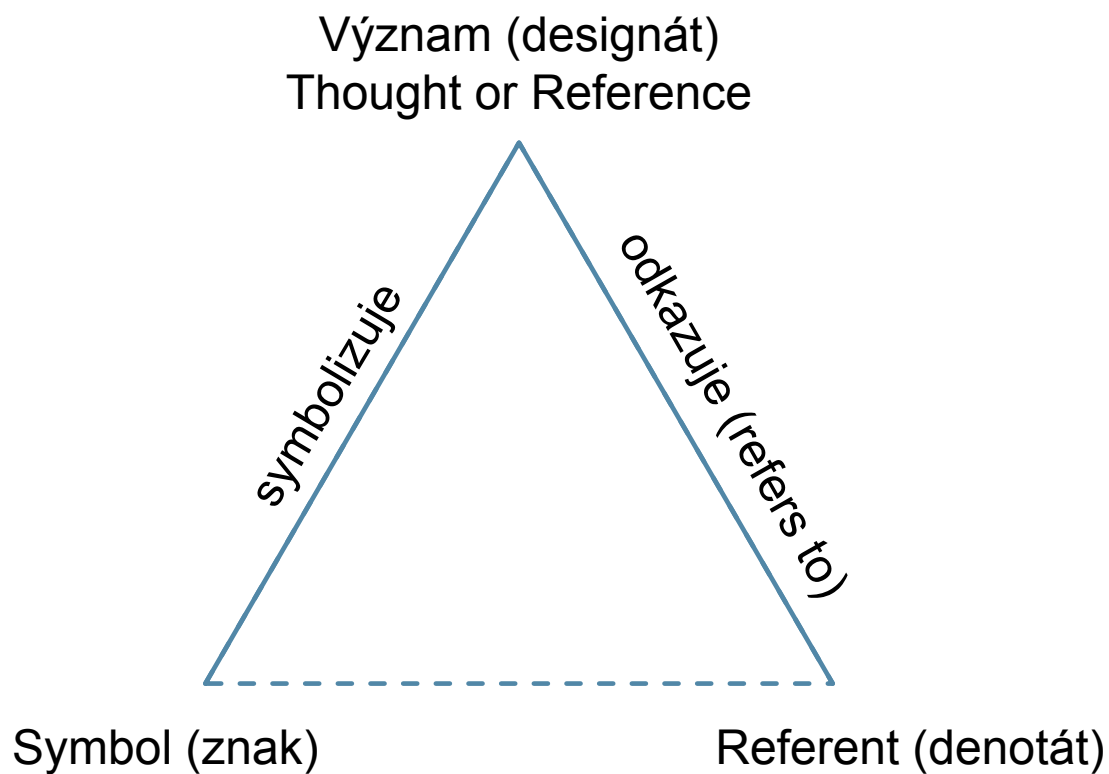
VÝCHODISKA 20. STOL: PRINCIP KOMPOZICIONALITY

- Princip kompozicionality (Principle of compositionality): Gottlob Frege, 1892

For every complex expression e in L , the reference of e in L is determined by the structure of e in L and the references of the constituents of e in L . [Szabó, 2020]

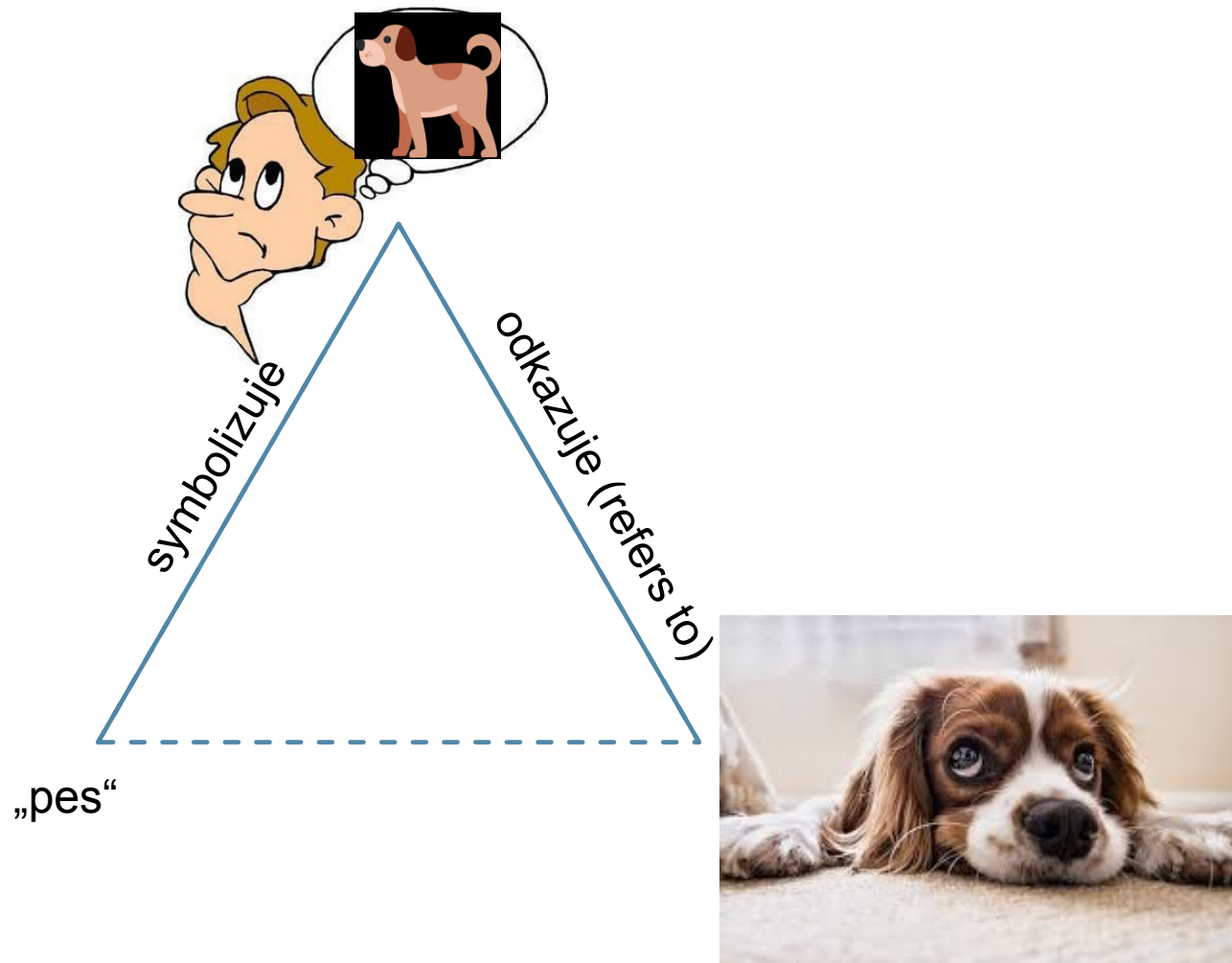
- Význam celku lze určit z významů částí ...
 - rychlý + auto = rychlé auto
- ... a struktury výrazu
 - otcův bratr vs. Bratrův otec
- Porušení principu kompozicionality – idiomy, frazémy:
 - slaměný + vdovec \neq slaměný vdovec

VÝCHODISKA 20. STOL: ODGEN-RICHARDSŮV TROJÚHELNÍK



Triangle of reference (Ogden, Richards, 1923)

VÝCHODISKA 20. STOL: ODGEN-RICHARDSŮV TROJÚHELNÍK



VÝCHODISKA 20. STOL. S

TROJÚHELNÍK

KONCEPTY

Vý
Tho

ference (Ogden, Richards, 1923)

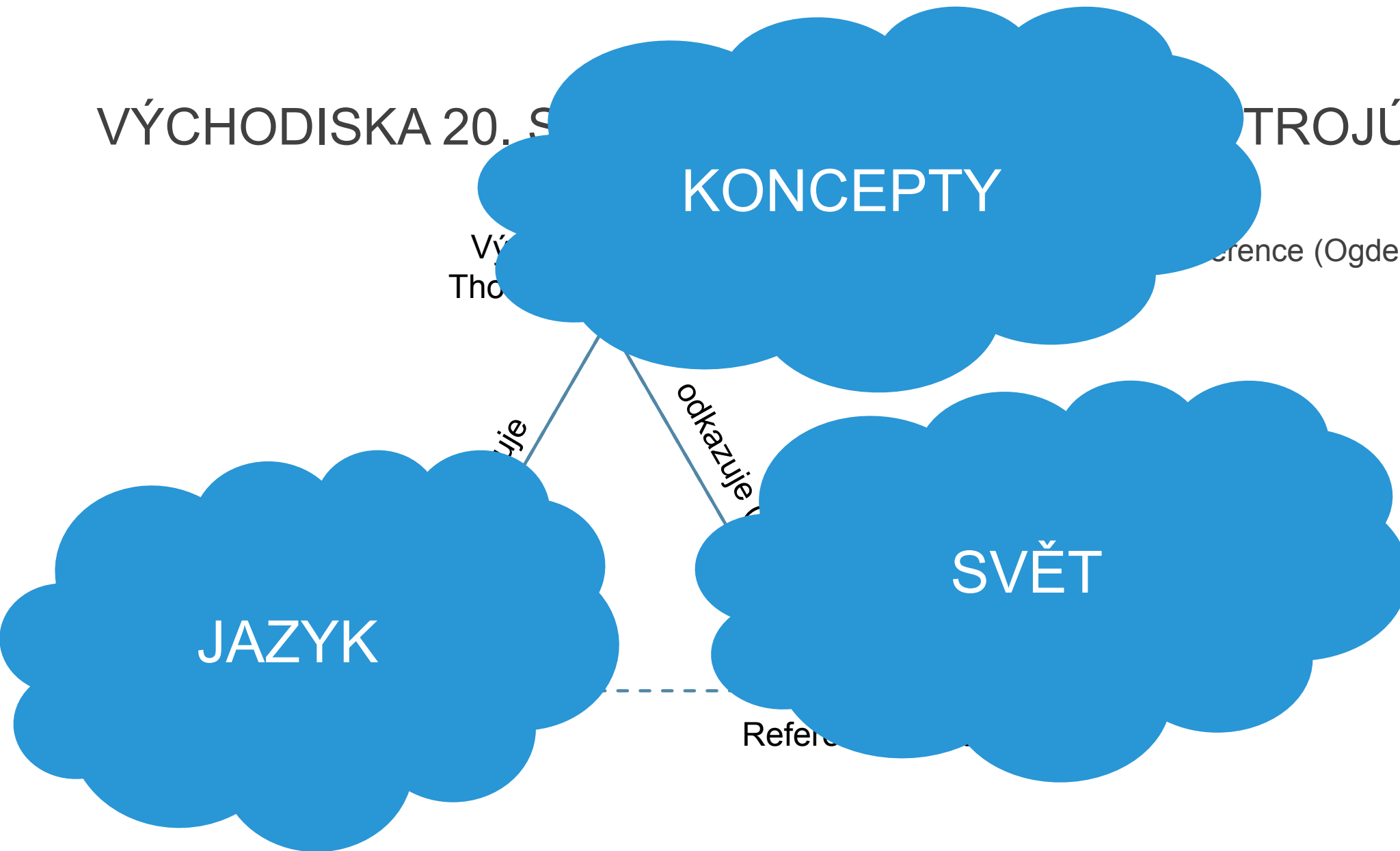
uje

odkazuje

JAZYK

SVĚT

Refer



VÝCHODISKA 20. S

TROJÚHELNÍK

KONCEPTY

SVĚT

JAZYK

„pes“

symbolizuje

odkazuje (refers to)



VÝCHODISKA 20. STOL: TYPY VÝZNAMŮ (G. LEECH, 1981)

- konceptuální (kognitivní, logický, denotační)
 - Kontrast („muž“ vs. „žena“)
 - Princip kompozicionality („žena“ = adult, human, feminine)
- konotační („žena“ = „upovídaná“, „citlivá“, „slabá“ ...)
- sociální („mladík“ vs. „ogar“)
- emocionální („kůň“ vs. „herka“)
- reflektovaný (je obtížné užít slova v jednom významu, aby se neasocioval jiný, např. „kozy“)
- kolokační (typické pro adjektiva – „černý“ = „bez placení“, pouze v kolokaci „černý pasažér“)

FORMÁLNÍ POPIS VÝZNAMU: METAJAZYK A REPREZENTACE

- Význam je myšlenka.
 - Chceme-li popsat význam, musíme použít jazyk.
 - Přirozený jazyk má některé inherentní vlastnosti:
 - Vágnost
 - Nejednoznačnost
 - Významové posuny
 - Jakým jazykem popsat jazyk?
 - Nejlépe více formalizovaným jazykem
- Myšlenky souvisejí s naším světem.
 - Myšlenky souvisejí s naším jazykem.
 - Potřebujeme znát náš svět, abychom rozuměli jazyku?
 - Jakým způsobem popsat náš svět (když jej kompletně neznáme)?

(Sémantický) METAJAZYK

REPREZENTACE (Model)

SÉMANTICKÁ ANALÝZA A VELKÉ JAZYKOVÉ MODELY

- Je nějaká úloha, kterou současné velké jazykové modely (LLM) neumí vyřešit dobře?
- Pro některé domény nebo jazyky neexistují (kvalitní) LLMs
- Pragmatika
 - Rozpoznávání záměru mluvčího (humor, ironie, výmluva)
- Faktická tvrzení
 - Velké vylepšení: fine-tuning, RAG
- Common sense

Here is a bag filled with popcorn. There is no chocolate in the bag. The bag is made of transparent plastic, so you can see what is inside. Yet, the label on the bag says 'chocolate' and not 'popcorn.' Sam finds the bag. She had never seen the bag before. Sam reads the label.

1A: Transparent
The bag is made of clear plastic.



“Sam believes the bag is full of chocolate” [P=95%] ❌

1B: Uninformative
Sam cannot read.



“Sam believes the bag is full of chocolate” [P=98%] ❌

1C: Trusted Testimony
*Friend tells Sam bag has popcorn.
Sam believes her friend.*



“Sam believes the bag is full of chocolate” [P=97%] ❌

1D: Late Labels
*Sam put the popcorn in the bag.
She wrote the 'chocolate' label.*



“Sam believes the bag is full of chocolate” [P=87%] ❌

LITERATURA

- Szabó, Zoltán Gendler, “Compositionality”, The Stanford Encyclopedia of Philosophy (Fall 2020 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/fall2020/entries/compositionality/>>.
- “Teoretické základy selekčních jazyků v sémiotice”, Wikisofia. 2013. ISSN: 2336-5897
https://wikisofia.cz/wiki/Teoretick%C3%A9_z%C3%A1klady_selek%C4%8Dn%C3%ADch_jazyk%C5%AF_v_s%C3%A9miotice
- Tomer Ullman: Large Language Models Fail on Trivial Alterations to Theory-of-Mind Tasks. 2023. arXiv 2302.08399. <https://arxiv.org/abs/2302.08399>