



---

# PLIN021 SÉMANTICKÁ ANALÝZA V PRAXI

ZUZANA NEVĚŘILOVÁ

2020–24

# SÉMANTIKA DISKURZU

Psí granule a kafe.

- Co si přejete?
- Co se to tady vysypalo?
- Co máte nejraději?
- Cos dnes jedl?
- Co po tobě ten pes chtěl?
- Co po tobě ten člověk chtěl?
- Co po tobě ten člověk mrštil?
- ...



<https://playgrounds.ai/models/dalle-mini>

# ANALÝZA PROMLUVY: KRABICOVÝ MODEL

- Jazykové a mimojazykové signály

A: Už jsi ten motor smontoval?

– Proveč lano tím okem na horní straně motoru.

– Jo, mimochodem, koupils už ten benzín?

B: Jasně, koupil, když jsem sháněl disk do sekačky.

– Zapomněl jsem vzít kanystr, tak jsem koupil nový.

A: Byl drahý?

B: Ne, ale bude se mi hodit do auta.

A: Fajn.

– Už to máš provlečené?

# SÉMANTIKA DISKURZU

## Prostředky koherence (konektory)

- časová souslednost (jednota času, místa a děje)
- porušení časové souslednosti je vyjádřeno explicitně: "ještě předtím"
- výrazy jako „Nejprve ... , potom ...“, „Oproti tomu ...“, “také”
- elipsa: Koupila jsem si auto a Marie [si koupila auto] taky.

# ELIPSA, VÝPUSTKA (ELLIPSIS)

- Petr šel na večírek, kde [Petr] potkal Pavlu.
- Koupila jsem si auto a Marie [si koupila auto] taky.
- Mám zavolat já tobě, nebo ty [máš zavolat] mně?
- [Mám vám dát na ty brambory] máslo?
- Nevím proč [bych měla tuhle knížku číst].

# PROMLUVOVÉ OBJEKTY

- seznam objektů promluvy (promluvový objekt, PO; discourse entity):
- množina prvků znalostní báze (knowledge base, KB), které byly **zmíněny** a mohou být odkazovány pomocí zájmen
- pokud prvek nebyl zmíněn, a přesto může být odkazován, byl **evokován**
- jmenná fráze typicky vyjadřuje nějaký PO

Karlovi<sub>i</sub> někdo ukradl auto<sub>j</sub>, které<sub>j</sub> [on]<sub>i</sub> měl zaparkované před domem<sub>k</sub>.

[on]<sub>i</sub> Zavolal na policii<sub>i</sub>, [oni]<sub>i</sub> přijeli, [oni]<sub>i</sub> sepsali to<sub>m</sub>.

Za měsíc mu<sub>i</sub> [oni]<sub>i</sub> napsali, že [oni]<sub>i</sub> případ<sub>m</sub> odkládají.

# ODKAZY V DISKURZU

- **exofora** (odkaz mimo text)  
Co je *to*?
- **endofora** (odkaz do textu)  
v takovém případě
- **anafora** (zpětný odkaz) – antecedent (dříve evokovaný PO)  
Anežka na *sebe* hodila kabát a vyrazila.
- **katafora** (dopředný odkaz)  
Protože [*on*] byl chytrý, vydal se David nejprve za svým šéfem.
- **koreference**: Václav Klaus, Klaus, bývalý prezident, on, čorlpero
- druhy anafor:
  - **deixe**: Petr si ukrojil chleba a pak *ho* snědl.
  - **synonymum**: Petr si ukrojil chleba a pak *krajíc* snědl.

# TERMINOLOGIE

- Anafora (anaphor) – ukazatel na antecedent
  - Antecedent – entita v promluvě
  - Koreference (coreference) – stav, kdy anafora a antecedent společně odkazují na objekt v reálném světě
  - Koreferenční řetězec (coreferential chain) – více frází je spojených koreferencí
- **This book** is about anaphora resolution. **The book** is designed to help beginners in the field and **its** author hopes that **it** will be useful.



# TYPY ANAFOR

- Pronominální
  - Osobní zájmena
  - Přivlastňovací zájmena
  - Reflexivní zájmena
  - Ukazovací zájmena
  - Vztažná zájmena
- Lexikální jmenná fráze
- Slovesná, adverbiální
- Nulová anafora = elipsa
- **Ne všechna zájmena jsou anafory.**  
**Ono** je **to** jedno.  
Deixe (ukazování) – mimojazyková skutečnost
- **Ne všechny anafory znamenají korefenci.**  
Substituční test:  
**Every man** has **his** own destiny.
- **Některé anafory odkazují k podobným entitám** (identity-of-sense anaphora).  
The man who gave his **paycheck** to his wife was wiser than the man that gave **it** to his mistress.

---

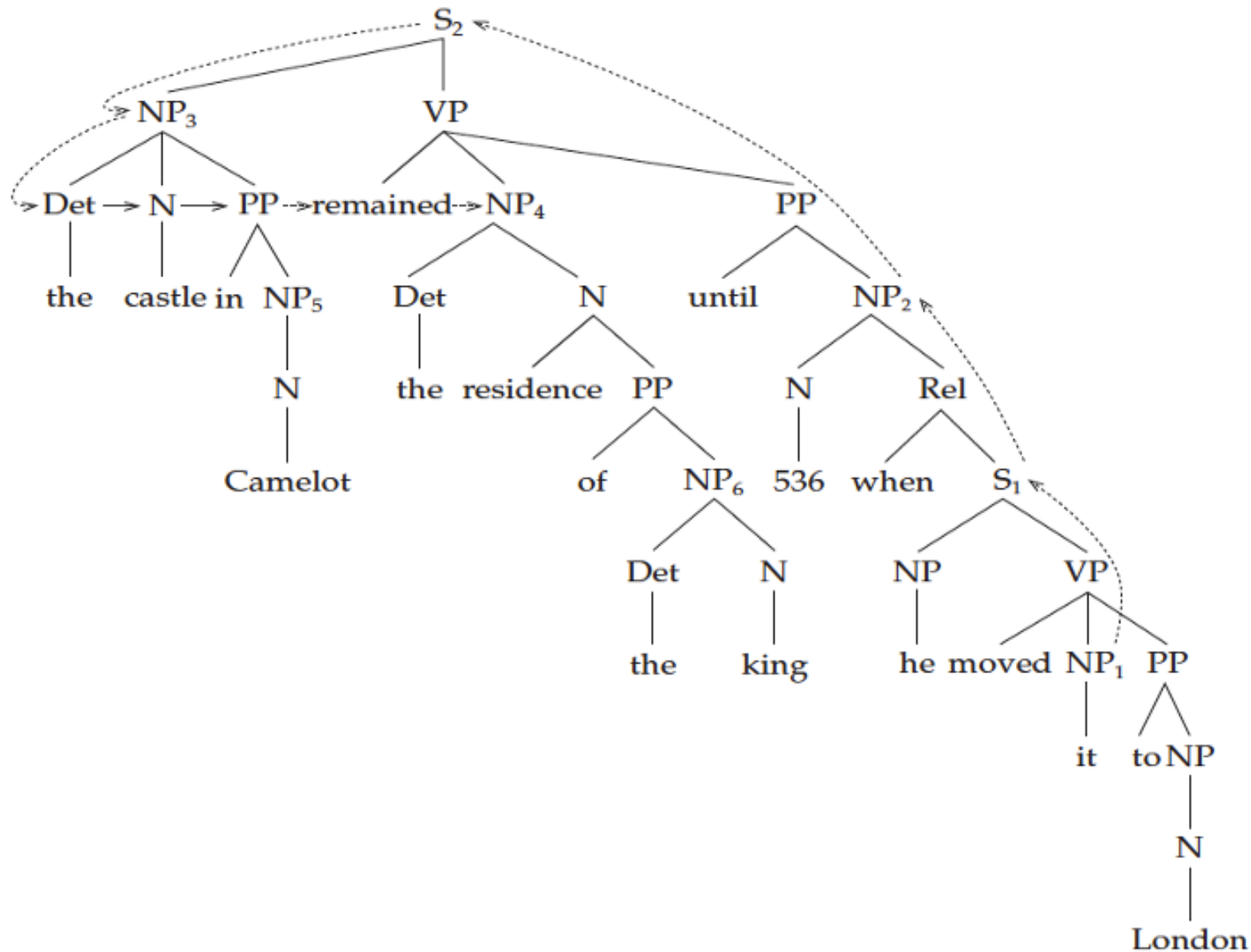
# ROZPOZNÁNÍ ANAFOR, REZOLUCE ANAFOR (ANAPHORA RESOLUTION)

HOBBSŮV ALGORITMUS



# HOBBSŮV ALGORITMUS

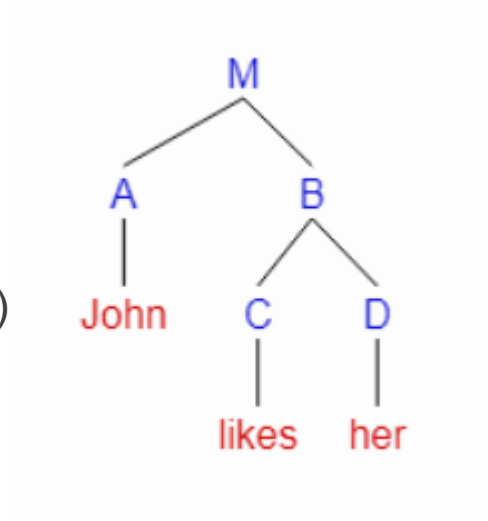
- Postupujeme od pronominální anafory doleva a nahoru
- Omezení (constraints)
  - C-command
  - Gramatická shoda
  - Sémantická omezení
- Preference
  - Podmět, přímý předmět
  - Syntaktický paralelismus
  - Centrum



# OMEZENÍ A PREFERENCE

- Gramatická shoda (rod, číslo)
- C-command
  - S výjimkou reflexiv
- Sémantické rysy

George removed the disk from the computer and then disconnected **it**.



- **Podmět**  
The customer lost patience and called the waiter.  
**He** ordered two 12-inch pizzas.
- **Centrum (topic, téma)**  
Tilly's mother had agreed to make her a new dress for the party. She worked hard on the dress for weeks and finally it was ready for Tilly to try on. Impatient to see what it would look like, Tilly tried on the dress over her skirt and ripped **it**.
- **Syntaktický paralelismus**  
The programmer successfully combined Prolog with C, but he had combined **it** with Pascal last time.



# MITKOV'S ANTECEDENT INDICATORS



# MITKOV'S ANTECEDENT INDICATORS

- Poslední tři věty
  - Všechny možné antecedenty
  - Každý kandidát získá pozitivní nebo negativní skóre na základě indikátorů
  - Skóre: statistická analýza (x, y)
  - Skóre: pravidla
    - Pozitivní pravidlo: první NP
    - Negativní pravidlo: NP jako součást PP, neurčitá NP
  - Algoritmus vybere antecedent s nejvyšším skóre



# NEURÁLNÍ MODEL PRO REZOLUCI ANAFOR



# NEURÁLNÍ MODEL PRO REZOLUCI ANAFOR (NEURALCOREF)

- Fráze:  $i$ , které zmiňují nějakou entitu (mentions)
- Možné antecedenty:  $Y(i) = \{\varepsilon, 1, \dots, i - 1\}$
- $\varepsilon$ :
  - fráze  $i$  nezmiňuje entitu
  - Fráze  $i$  zmiňuje entity, ale ta není koreferencí
- Koreference: relace mezi  $i$  a  $y_i$

Yes, **I** noticed that **many friends**, around **me** received **it**.  
**It** seems that **almost everyone** received **this SMS**.

Mention	Potential antecedents						
I	∅						
many friends	∅	I					
me	∅	I	many friends				
it	∅	I	many friends	me			
It	∅	I	many friends	me	it		
almost everyone	∅	I	many friends	me	it	It	
this SMS	∅	I	many friends	me	it	It	almost everyone



## DATASETY

Na čem lze vyhodnotit  
rezoluci anafor?

- SemEval-2010 Task 1 (Multilingual Coreference Resolution)
- EVALITA 2011 Anaphora Resolution Task
- ELG Anaphora Resolution Dataset (Wikipedia)
- Anaphora Resolution and Underspecification (ARRAU) corpus
- OntoNotes 5.0
- CoNLL 2012
- GAP (Gender Ambiguous Pronouns)
- WSC (Winograd Schema Challenge)

# LITERATURA

- [https://wiki.apertium.org/wiki/Anaphora\\_resolution\\_module](https://wiki.apertium.org/wiki/Anaphora_resolution_module)
- Kenton Lee, Luheng He, Luke Zettlemoyer: Higher-order Coreference Resolution with Coarse-to-fine Inference. NAACL 2018 · <https://paperswithcode.com/paper/higher-order-coreference-resolution-with>
- Uryupina, Olga; Poesio, Massimo (2021). Anaphora Resolution Dataset. Version 1.0.0. European Language Grid. [Dataset (Text corpus)]. <https://doi.org/10.57771/hk5e-df59>
- Marta Recasens, Lluís Màrquez, Emili Sapena, M. Antònia Martí, Mariona Taulé, Véronique Hoste, Massimo Poesio, and Yannick Versley. 2010. [SemEval-2010 Task 1: Coreference Resolution in Multiple Languages](#). In Proceedings of the 5th International Workshop on Semantic Evaluation, pages 1–8, Uppsala, Sweden. Association for Computational Linguistics.
- Massimo Poesio and Olga Uryupina: EVALITA 2011 Anaphora Resolution Task. <http://www.evalita.it/2011/tasks/anaphora>
- Rhea Sukthanker, Soujanya Poria, Erik Cambria, Ramkumar Thirunavukarasu: Anaphora and Coreference Resolution: A Review <https://arxiv.org/abs/1805.11824>
- [IA161 Natural Language Processing in Practice](#) (autumn 2024): <https://nlp.fi.muni.cz/en/NLPInPracticeCourse/AnaphoraResolution>