

PLIN037 Sémantika a počítače

OP VK Mezi bohemistikou a informatikou
www.projekt-inova.cz

Zuzana Nevěřilová
xpopelk@fi.muni.cz

Centrum zpracování přirozeného jazyka, B203
Fakulta informatiky, Masarykova univerzita

19. dubna 2016

Sémantika diskurzu

Rozpoznání anafor

Sémantika diskurzu (*discourse semantics*)

Honza_i ztratil peněženku_j. [*On*]_i Hledal ji_j v zadním voze tramvaje_k nejmíň půl hodiny. Nakonec ho_i odtam_k vyhodili, přestože [*On*]_i ji_j nenašel. [*On*]_i Byl z toho? pěkně naštvaný.

Sémantika diskurzu (*discourse semantics*)

Honza_i ztratil peněženku_j. [*On*]_i Hledal ji_j v zadním voze tramvaje_k nejmíň půl hodiny. Nakonec ho_i odtam_k vyhodili, přestože [*On*]_i ji_j nenašel. [*On*]_i Byl z toho? pěkně naštvaný.

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

typický PO: jmenná fráze

Sémantika diskurzu (*discourse semantics*)

Honza_i ztratil peněženku_j. [*On*]_i Hledal ji_j v zadním voze tramvaje_k nejmíň půl hodiny. Nakonec ho_i odtam_k vyhodili, přestože [*On*]_i ji_j nenašel. [*On*]_i Byl z toho? pěkně naštvaný.

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**

typický PO: jmenná fráze

Odkazy v diskurzu: rozpoznání anafor, rezoluce anafor (anaphora resolution, AR)

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 (různé výrazy pro tentýž PO): Václav Klaus, Klaus, bývalý prezident, on

Odkazy v diskurzu: rozpoznání anafor, rezoluce anafor (anaphora resolution, AR)

druhy anafor:

- deixis: Petr si ukrojil chleba a snědl *ho*. (pronominální anafora)
- synonymum: Petr si ukrojil chleba a pak *krajíc* snědl.
- hyperonymum: Petr si ukrojil chleba a pak *jídlo* snědl.

v jazycích s volným slovosledem jsou časté nevyjádřené podmínky:

Petr si ukrojil chleba a [Petr] snědl ho[=chleba].

Odkazy v diskurzu: základní algoritmus AR

Odkazy v diskurzu: základní algoritmus AR

1. objekty promluvy (PO): promluvový zásobník (*history list*)
2. při každé zmínce objektu se PO posune na vrchol zásobníku
3. každý odkaz se nahradí PO, který je nejbliž vrcholu zásobníku a obsahuje gram. shodu (číslo, příp. rod)

Odkazy v diskurzu: základní algoritmus AR

1. objekty promluvy (PO): promluvový zásobník (*history list*)
2. při každé zmínce objektu se PO posune na vrchol zásobníku
3. každý odkaz se nahradí PO, který je nejbliž vrcholu zásobníku a obsahuje gram. shodu (číslo, příp. rod)

Honza ztratil peněženku. [on] Hledal ji v zadním voze tramvaje nejmíň půl hodiny. Nakonec ho odtam [oni] vyhodili, přestože ji nenašel. [on] Byl z toho pěkně naštvaný.

peněženka

Honza

Odkazy v diskurzu: základní algoritmus AR

1. objekty promluvy (PO): promluvový zásobník (*history list*)
2. při každé zmínce objektu se PO posune na vrchol zásobníku
3. každý odkaz se nahradí PO, který je nejbliž vrcholu zásobníku a obsahuje gram. shodu (číslo, příp. rod)

Honza ztratil peněženku. [on] Hledal ji v zadním voze tramvaje nejmíň půl hodiny. Nakonec ho odtam [oni] vyhodili, přestože ji nenašel. [on] Byl z toho pěkně naštvaný.

Honza
peněženka

Odkazy v diskurzu: základní algoritmus AR

1. objekty promluvy (PO): promluvový zásobník (*history list*)
2. při každé zmínce objektu se PO posune na vrchol zásobníku
3. každý odkaz se nahradí PO, který je nejbliž vrcholu zásobníku a obsahuje gram. shodu (číslo, příp. rod)

Honza ztratil peněženku. [on] Hledal **ji** v zadním voze tramvaje nejmíň půl hodiny. Nakonec ho odtam [oni] vyhodili, přestože **ji** nenašel. [on] Byl z toho pěkně naštvaný.

peněženka

Honza

Odkazy v diskurzu: základní algoritmus AR

1. objekty promluvy (PO): promluvový zásobník (*history list*)
2. při každé zmínce objektu se PO posune na vrchol zásobníku
3. každý odkaz se nahradí PO, který je nejbliž vrcholu zásobníku a obsahuje gram. shodu (číslo, příp. rod)

Honza ztratil peněženku. [on] Hledal ji v zadním voze tramvaje nejmíň půl hodiny. Nakonec ho odtam [oni] vyhodili, přestože ji nenašel. [on] Byl z toho pěkně naštvaný.

zadní vůz tramvaje

peněženka

Honza

Odkazy v diskurzu: základní algoritmus AR

1. objekty promluvy (PO): promluvový zásobník (*history list*)
2. při každé zmínce objektu se PO posune na vrchol zásobníku
3. každý odkaz se nahradí PO, který je nejbliž vrcholu zásobníku a obsahuje gram. shodu (číslo, příp. rod)

Honza ztratil peněženku. [on] Hledal ji v zadním voze tramvaje nejmíň půl hodiny. Nakonec **ho** odtam [oni] vyhodili, přestože ji nenašel. [on] Byl z toho pěkně naštvaný.

zadní vůz tramvaje

peněženka

Honza

Odkazy v diskurzu: základní algoritmus AR

1. objekty promluvy (PO): promluvový zásobník (*history list*)
2. při každé zmínce objektu se PO posune na vrchol zásobníku
3. každý odkaz se nahradí PO, který je nejbliž vrcholu zásobníku a obsahuje gram. shodu (číslo, příp. rod)
4. v jedné klauzi se PO vyskytuje jen jednou

Honza ztratil peněženku. [on] Hledal ji v zadním voze tramvaje nejmíň půl hodiny. Nakonec ho **odtam** [oni] vyhodili, přestože ji nenašel. [on] Byl z toho pěkně naštvaný.

Honza?

zadní vůz tramvaje

peněženka

Odkazy v diskurzu: základní algoritmus AR

1. objekty promluvy (PO): promluvový zásobník (*history list*)
2. při každé zmínce objektu se PO posune na vrchol zásobníku
3. každý odkaz se nahradí PO, který je nejbliž vrcholu zásobníku a obsahuje gram. shodu (číslo, příp. rod)
4. v jedné klauzi se PO vyskytuje jen jednou

Honza ztratil peněženku. [on] Hledal ji v zadním voze tramvaje nejmíň půl hodiny. Nakonec ho odtam [oni] vyhodili, přestože ji nenašel. [on] Byl z toho pěkně naštvaný.

oni

Honza?

zadní vůz tramvaje

peněženka

Odkazy v diskurzu: základní algoritmus AR

1. objekty promluvy (PO): promluvový zásobník (*history list*)
2. při každé zmínce objektu se PO posune na vrchol zásobníku
3. každý odkaz se nahradí PO, který je nejbliž vrcholu zásobníku a obsahuje gram. shodu (číslo, příp. rod)
4. v jedné klauzi se PO vyskytuje jen jednou

Honza ztratil peněženku. [on] Hledal ji v zadním voze tramvaje nejmíň půl hodiny. Nakonec ho odtam [oni] vyhodili, přestože **ji** nenašel. [on] Byl z toho pěkně naštvaný.

peněženka

oni

Honza?

zadní vůz tramvaje

Odkazy v diskurzu: základní algoritmus AR

1. objekty promluvy (PO): promluvový zásobník (*history list*)
2. při každé zmínce objektu se PO posune na vrchol zásobníku
3. každý odkaz se nahradí PO, který je nejbliž vrcholu zásobníku a obsahuje gram. shodu (číslo, příp. rod)
4. v jedné klauzi se PO vyskytuje jen jednou

Honza ztratil peněženku. [on] Hledal ji v zadním voze tramvaje nejmíň půl hodiny. Nakonec ho odtam [oni] vyhodili, přestože ji nenašel. [on] Byl z toho pěkně naštvaný.

Honza

peněženka

oni

zadní vůz tramvaje

Odkazy v diskurzu: základní algoritmus AR

1. objekty promluvy (PO): promluvový zásobník (*history list*)
2. při každé zmínce objektu se PO posune na vrchol zásobníku
3. každý odkaz se nahradí PO, který je nejbliž vrcholu zásobníku a obsahuje gram. shodu (číslo, příp. rod)
4. v jedné klauzi se PO vyskytuje jen jednou

Honza ztratil peněženku. [on] Hledal ji v zadním voze tramvaje nejmíň půl hodiny. Nakonec ho odtam [oni] vyhodili, přestože ji nenašel. [on] Byl z toho pěkně naštvaný.

Honza

peněženka

oni

zadní vůz tramvaje