

M A T E M A T I C K O - F Y Z I K Á L N Í F A K U L T A
P R A H A

ANOTOVÁNÍ ROZŠÍŘENÉ TEXTOVÉ KOREFERENCE A ASOCIAČNÍ ANAFORY
V PRAŽSKÉM ZÁVISLOSTNÍM KORPUSU.

ANJA NĚDOLUŽKO
20. LISTOPADU 2008

U N I V E R Z I T A K A R L O V A V P R A Z E

ÚVOD	5
A. GRAMATICKÁ KOREFERENCE	5
Dodržování koreferenčního řetězce	5
B. TEXTOVÁ KOREFERENCE	6
Textová koreference na velkou textovou vzdálenost	6
B.1. Slovnědruhá charakteristika koreferovaných párů	7
B.2. Typologie textově koreferenčních vztahů	8
B.2.1. Koreferenční vztah mezi NP se specifickou referencí, kde koreferující člen není synonymum ani hyponymum antecedenta {coref_text, typ_0}	9
Kataforický odkaz dopředu	9
Koreference osobních zájmen v dialogických textech	10
Koreference otázkového slova a odpovědi v dialogických textech	10
B.2.2. Koreference synonymních NP {coref_text, typ_SYN}	10
B.2.3. Koreference hyponymu a hyperonymu {coref_text, typ_ER}	11
B.2.4. Koreference nereferenčních a generických NP {coref_text, typ_NR}	12
B.3. Problématické případy označování textové koreference	12
B.3.1. Koreference s typem NR vs. nic	12
B.3.2. Koreference abstraktních jmen	12
B.3.3. Dvě místní určení vedle sebe (tady v Praze, u nás doma apod.)	13
B.4. Nejednoznačný výběr antecedentu	13
B.4.1. K otázce výběru antecedenta v případě apoziční skupiny	13
B.4.2. K otázce výběru antecedenta v případě koordinační skupiny	13
B.4.3. Spojení se slovy s funkcí „kontejneru“	14
B.4.4. Koreference jmen s funktorem ID	14
C. BRIDGING VZTAHY	14
C.1. Typologie bridging vztahů	14
C.1.1. bridging-vztah individual – function (FUNCT_P a P_FUNCT)	15
C.1.2. bridging-vztah část-celek (PART_WHOLE a WHOLE_PART)	15
C.1.3. bridging-vztah množina-podmnožina (SET_SUB a SUB_SET)	16
Sporný příklad 1 (vztah SET uvnitř jedné věty)	16
Sporný příklad 2 (otázka nutnosti a hloubky interpretace)	16
Sporný příklad 3 (označování některých bridging vztahů u nereferenčních NP)	16
Hraniční případy mezi bridging typy SET a PART	17
C.1.4. bridging-vztah sémantického protikladu (CONTRAST)	17
C.1.5. bridging-vztah zatím nedeterminovaný (REST)	17
Vztah rodinné příslušnosti	18
Vztah „místo – obyvatel“	18
Vztah typu „autor – kniha“	18
Vztah „věc – majitel“	18
Vztah mezi stejně vyjádřenými nebo synonymními nekoreferenčními NP	18

Vztah událost – argument.....	19
C.2. Nejednoznačný výběr antecedentů.....	19
C.2.1. Spojení se slovy s funkcí „kontejneru“	19
C.2.2. K otázce výběru antecedentu v případě apoziční skupiny.....	19
C.2.3. K otázce výběru antecedentu v případě koordinační skupiny:.....	20
D. BRIDGING NEBO TEXTOVÁ IDENTICKÁ – VÝBĚR A PREFERENCE.....	20
D.1. K otázce dvojího odkazování (textový a bridging vztahy) a preference:.....	20
D.2. „faktory – jeden z faktorů“	20
E. SPECIÁLNÍ TYPY REFERENCE (COREF_SPECIAL)	21
E.1. Exoforické odkazy.....	21
E.2. Odkazy na segmenty textu	21

Úvod

Existující anotace koreference v PDT 2.0. vychází z pojmu reference jazykových jednotek a dělí se na gramatickou a textovou koreferenci. Oblasti gramatické koreference a pronominální textové koreference jsou kompletně zpracovány na tektogramatické rovině a popsány ve „velkém manuálu“.

Další krok je rozšíření textové koreference o jiné vztahy než pronominalizace a zavádění třetí koreferenční oblasti – asociační anafory (tzv. bridging vztahy). Postup anotace je představen v následujícím textu.

Pozor! Jsme si vědomi toho, že termín koreference implikuje pouze identitu referentů objektů, přesto pro zjednodušení občas užíváme termínu koreference i pro případy bridging-vztahů a mimotextového odkazování.

A. Gramatická koreference

Antecedent je určován na základě gramatických pravidel daného jazyka. Všechny šipky, které byly, v anotaci původní koreference na PDT 2.0 zůstaly.

Dodržování koreferenčního řetězce

Gramatická koreference je propojena s textovou identickou koreferencí nepřetržitým řetězcem. To znamená, že pokud A, B a C jsou jmenné fráze následující po sobě, B je propojeno s A gramatickou koreferencí a C je textově koreferenční s A, šipka textové koreference vede od C k B. Srov.:

(13/In94204_107) Linka bezpečí 855 44 33 , kterou {coref_gram na „linka“} od zítřejšího dne provozuje v Praze nadace Naše dítě , byla zřízena především s úmyslem pomoci fyzicky i duševně týraným a pohlavně zneužívaným dětem .[...]

(17/In94204_107) Na toto telefonní číslo {coref_text na „který“} však mohou samozřejmě zavolat všichni kluci a děvčata , kteří se ocitnou ve svízelné situaci.

Dodržování koreferenčního řetězce je kontrolováno automaticky. Pokud anotátor nakreslí šipku na uzel, na který už vede šipka identické koreference, jeho šipka se automaticky překreslí na poslední uzel daného řetězce.

V případě bridging vztahu, propojujeme jmenné fráze podle významu. Srov.:

(17/In94204_107) *Na toto telefonní číslo* však mohou samozřejmě zavolat všichni kluci a děvčata , kterí {coref_gram na „a“} se ocitnou ve svízelné situaci.

(18/In94204_107) Ptali jsme se několika {bridging na „kluci a děvčata“} , jestli by takového kamaráda po telefonu považovali za dobrou věc .

B. Textová koreference

Textová koreference je vztah identity referentů antecedenta a koreferujícího člena. Je to vztah symetrický (pokud A je koreferenční s B, B je koreferenční s A) a tranzitivní (pokud A je koreferenční s B je koreferenční s C, pak A je koreferenční s C). Na dnešní etapě anotace rozlišujeme původní pronominální koreferenci a rozšířenou koreferenci.

Koreference se mohou zúčastnit NP v asertivních, otázkových i negovaných větách.

Jako koreferenční neoznačujeme:

- a) Vztah mezi jednotlivými členy apozice (z podstaty apozice vyplývá, že jednotlivá pojmenování jsou koreferenční).
- b) Vztah mezi subjektem a predikátovou částí výpovědi. Je to zřejmé zejména u přísudku jmenného se sponou (*Petr je lékař – Petr ... lékař*). Koreferenční vztah neoznačujeme ani v případě konstrukcí s „to“, i když je to do jisté míry arbitrární rozhodnutí. Srov.:

(18/In95047_061) My tady máme dost problematických dětí a ty kdyby se spělčily s Němci , to by nedělalo dobrotu .

(19/In95047_061) Slyšeli jsme , že to {coref_text, typ_0 na „Němci“} jsou děti {žádná koreferenční šipka}, které místo výkonu trestu mají být tady v Košťanech . . . " říká obyvatelka obce a signatářka petice .

V případě bridging vztahu tato konvence neplatí, odkazujeme tedy na uzel, který s koreferujícím členem bezprostředně sémanticky souvisí. Viz k tomu poznámku v C.

Textová koreference na velkou textovou vzdálenost

Textovou koreferenci anotujeme na vzdálenost nepřesahující 20 vět. Srov.

Označujeme:

(49/In94210_95) Poslanci budou muset odpočívat jinde , protože suterény jsou příliš hluboko a napojení na výše položenou kanalizaci pomocí čerpadel by provoz budov neúměrně prodražilo.

[...17 vět...]

(66/In94210_95) Existující kanalizační sítě {coref_text, typ_SYN na kanalizace} by totiž podzemní chodbu vtlačily tak hluboko do země , že by jí jistě nikdo nepoužíval.

Neoznačujeme:

(6/In95047_061) Situace začala být přirovnávána k porevolučním snahám některých západních firem " odložit " na naše území za malý peníz toxické odpady .

[...44 věty...]

(50/In95047_061) Její zástupce ing . Šedivý však veškerou odpovědnost za krizovou situaci {žádná koreferenční šipka} odmítá .

B.1. Slovnědruhá charakteristika koreferovaných párů

Jsou následujícími možnosti:

- i. koreferující člen je vyjádřen **substantivem**:
substantivum¹ – substantivum (*snímek – fotografie, Petr – ten kluk, pí Novotná – mluvčí* apod.)
zájmeno – substantivum (*on – Kolář*)
elidovaný #PersPron – substantivum (#PersPron - *Kolář*)
příslovce – substantivum (*tady – v Praze*)
sloveso – substantivum (*prodávat – prodej, plavat – tento proces*)
- ii. koreferující člen je vyjádřen **adverbiem**:
antecedent (A)² – příslovce (*Praha – (#PersPron) – (ta) – (odtud) – tam*)
- iii. jeden z členů páru je **adjektivum**. Koreferenci u adjektiv anotujeme ve dvou následujících případech:
 - a) adjektivum je přivlastňovací (*maminka – maminčin názor*). Adjektivum může mít koreferenční šipku, i když přivlastňovací význam je vyjádřen u gramaticky neutrálního adjektiva (*palácový – palác, dětský* s významem „dítěte“ např. v „*dětská mysl*“ v následujícím příkladě.

Pokud však adjektivum nemá sémantický rys přináležitosti, koreferenci neanotujeme. Tak v následujícím příkladě anotujeme koreferenci u „podnikatelovy“ nikoliv však u podnikatelský. Srov.

(19/In94208_11) Tímto faktorem je podnikatel - inovátor , který se snaží o zisk , a proto logicky nemůže existovat ve stavu statiky , která nezná ani zisk , ani ztrátu .

(22/In94208_11) Podnikatelova {coref_text, typ_0 na „košťanský“} odměna , zisk , má však svůj původ nikoliv ve fungování , ale v rozbití stacionárního systému .

(25/In94208_11) Tento druh podnikatelské {žádná šipka} odměny je vlastně monopolní rentou a je dočasné povahy .

- b) adjektivum je vytvořeno od pojmenované entity. Srov.

(96/In95047_061) Radní jednomyslně vyjádřili nesouhlas s přítomností chovanců společnosti Struktura v Košťanech .

¹ antecedent = substantivum, zájmeno apod. zde a dále vždy znamená, že tento slovní druh je kořen příp. podstromu antecedenta.

² Antecedent ∈ {substantivum, zájmeno, elidovaný člen, příslovce}

(98/In95047_061) Ředitelka košťanské {coref_text, typ_0 na „Košťany“} základní školy Jarmila Hejduková byla jednou z iniciátorek podpisové akce .

(99/In95047_061) Pikantní detail v celé záležitosti je , že třinácti až čtrnáctiletí chlapci si dům v Košťanech {coref_text, typ_0 na „košťanský“} teprve upravovali .

V ostatních případech koreferenci na adjektivum neanotujeme.

- iv. členy koreferenčního páru jsou **číslovky**. Označujeme pokud vystupují v substantivní funkci a podílejí na kohezi textu. Srov.

(16/In94207_76) Připomenu , že po vstupu vojsk SSSR 21 . srpna 1968 na naše území jsem se zdržel (vraceje se z jugoslávských prázdnin) ve Vídni .

(21/In94207_76) Vzpomínám na takzvané zelené hranice zcela bezbariérové a na dosud nevídanou blahovůli zahraničních a našich celních a policejních orgánů už na jaře 1968 {coref_text, typ_0 na „1968“}...

- v. **Sloveso** v roli koreferujícího člena vystupuje jen zřídka a v anotaci to zatím důsledně nezaznamenáváme. V některých případech je to ale vítáno. Srov.:

(68/In95047_061) Na převýchovu se pokud vím , posílali ti , kteří měli podle těchto zruďných režimů nevhodný původ .

(70/In95047_061) Naše sdružení nepřevychovává {coref_text, typ_0 na „převýchova“}, ale snaží se vychovávat .

Srov. také část řetězce generických použití (*žvýkačky – nežvýkají – žvýkání – žvýkali – žvýkání*):

(2/In94207_84) Vybrané kapitoly z dějin žvýkačky

(13/In94207_84) Lidi nežvýkají {coref_text, typ_NR na „žvýkačka“}, to jenom krávy .

(20/In94207_84) Pravda o tom , že žvýkání {coref_text, typ_NR na „žvýkat“} pro žvýkání bylo odjakživa činností veskrze lidskou – kam paměť lidského rodu sahá .

(24/In94207_84) Se stejnou radostí však zamlčí , že Řekové často a s oblibou žvýkali {coref_text, typ_NR na „žvýkání“} kousky ztuhlé mízy mastikového keře, který se pěstuje především na ostrově Chios .

(25/In94207_84) Známý antický lékař a botanik Dioscorides psal v prvním století našeho letopočtu obsáhle o léčebném a hygienickém účinku žvýkání {coref_text, typ_NR na „žvýkat“} .

Ve funkci antecedenta sloveso vystupuje zcela běžně.

B.2. Typologie textově koreferenčních vztahů

Rozlišujeme čtyři druhy vztahů mezi antecedentem a koreferujícím členem v páru uzlů propojených textovou koreferencí:

- **0** – vztah mezi NP se specifickou referencí, kde koreferující člen není ani synonymum ani hyperonymum antecedenta (viz B.2.1.)
- **SYN** (od *synonymum*) – vztah mezi synonymními NP se specifickou referencí (viz B.2.2.)
- **ER** (od *hyperonymum*) – vztah mezi NP se specifickou referencí, kde druhý člen je lexikální hyperonym ve vztahu k prvnímu (viz B.2.3.)
- **NR** (od *nespecifická reference*) – vztah mezi NP s nespecifickou nebo generickou referencí (viz B.2.4.)

Pozor! Rozdělení textové koreference na typy přidělilo všem původním koreferenčním spojením typ_0. Při anotaci rozšířené koreference se pokusíme u párů s nespecifickou referencí a již označenou textovou koreferencí typ 0 měnit na NR, příp. se to dá udělat automaticky tak, že pokud #PersPron vede na uzel, který je spojen s předchozím uzlem textovou koreferencí a má typ NR, pak i vztah #PersPron k jeho uzlu má typ NR.

B.2.1. Koreferenční vztah mezi NP se specifickou referencí, kde koreferující člen není synonymum ani hyperonymum antecedenta {coref_text, typ_0}

Tento vztah je základní a v anotaci je předvolený. Oba jeho členy mají specifickou referenci a jsou označeny stejnou nebo skoro stejnou (rozdíl v upřesňovacích adjektivech, použití ukazovacího zájmena, příp. adjektivizace u pojmenovaných entit) jmennou frází. Srov. např.

(6/In9413_006) Jeho dojetí znásobila při vyhlašování přítomnost [...] pořadatelů soutěže - Českého manažerského centra v Čelákovících .

(7/In9413_006) Na letošním ročníku soutěže {coref_text, typ_0 na „soutěž“} se spolupodílí i Profit .

Vyjádření identity pomocí textových identifikátorů. Srov:

Ten článek v dnešních novinách o otci, který utekl od ženy a dětí, aby je nemusil živit, to je strašné. Co bude teď chudák ta žena {coref_text, typ_0 na „žena“} s dětmi dělat? (Fuks, L., Spalovač mrtvol)

Propojení NP – #PersPron – NP. Srov. posloupnost vět (24)-(34):

(24/In94204_107) Sedmiletý Pěťa se půl roku neuvěřitelně trápil, že má AIDS. [...]

(29/In94204_107) #PersPron {coref_text, typ_0 na „Pěťa“} Stále na to myslel, ve škole se už nedokázal soustředit. [...]

(34/In94204_107) Pěťa {coref_text, typ_0 na #PersPron} skončil u Jany Drtilové.

Kataforický odkaz dopředu

Srov. také kataforický odkaz dopředu:

(11/cmpr9410_028) Tu nejvhodnější dobu {coref_text, typ_0 na „rok“}pan Hrabák propásl .

(12/cmpr9410_028) V osmdesátých letech se daly pořídít krásné věci za , viděno dneškem , ještě krásnější ceny.

Koreference osobních zájmen v dialogických textech.

U osobních zájmen 1. a 2. osoby koreferenci neanotujeme.

Koreference otázkového slova a odpovědi v dialogických textech.

Pro koherenci textu vztah mezi otázkovým slovem a částí odpovědi, která na tu otázku reaguje, zachycení vztahu mezi nimi je velice důležité. Je to však jiný typ koheze textu, který již přesahuje tekrogramatickou rovinu a patří spíše do roviny diskurzu. Při rozšířené anotaci tektogramatické roviny tento vztah nezachycujeme. Jde o dialogy typu:

(45/In94204_107) Kdy děti nejvíce volají [...]

(49/In94204_107) Podle zkušeností ze zahraničí se dá předpokládat, že největší frekvence telefonátů nastane vždy mezi 16 . až 18 . hodinou.

(50/In94204_107) A také při změnách počasí , které působí na citlivější organismus .

(51/In94204_107) V obdobích před a po vysvědčení .

(52/In94204_107) V době viróz .

Pozor! V dialogickém textu však běžně označujeme jiné vztahy, než koreference osobních zájmen 1. a 2. osoby a otázkové slovo – odpověď, jde-li o identickou koreferenci nebo bridging. Srov.:

(33/In94208_11) Dovožoval , že vývoj kapitalismu se historicky vyznačuje dvěma fázemi : Fází soutěžního kapitalismu a fází kapitalismu trustů .

(43/In94208_11) Schumpeter se ve svém posledním díle ptá : Který systém , kapitalismus {coref_text, typ_0 na „kapitalismus“}, či (44/In94208_11) socialismus , bude určovat budoucnost lidstva ? K údivu , úžasu či ohromení většiny svých kolegů odpovídá jednoznačně : Bude to socialismus {coref_text, typ_0 na „socialismus“}.

B.2.2. Koreference synonymních NP {coref_text, typ_SYN}

Synonymií chápeme ve velmi širokém smyslu – typ SYN mají v podstatě koreferenční páry různě lexikálně vyjádřené. Označujeme jenom u párů se specifickou referencí.

(41/In94204_107) Příjemně ji překvapilo , že se přihlásilo tolik dobrovolníků, kteří chtějí pomáhat druhým lidem.

(42/In94204_107) Nyní má linka třicet osm tzv. volontérů {coref_text, typ_SYN na „dobrovolník“}, kteří budou naslouchat volajícím.

Vztah SYN označujeme i přes #PersPron. Srov.

(57/In94207_84) Když o deset let později obrátil ke gumě pozornost louisvilleský lékárník John Colgan , existovala již řada žvýkačkových milionářů (mezi nimi {coref_text, typ_0 na „milionář“} Adams) .

(60/In94207_84) Výsledek , který dostal obchodní jméno Taffy Tolu [], se ujal okamžitě , Colgan zavřel lékárnu a během krátké doby se přidal k milionářskému klubu {coref_text, typ_SYN na #PersPron } .

Vztah SYN může vést na kořen souřadné nebo seznamové struktury. Srov.

(96/In94207_84) Jde především o zdravotní (či pseudozdravotní) námítky proti žvýkání .

(97/In94207_84) Tento boj {coref_text, typ_SYN na „čí“, sémanticky však samozřejmě na námítky } začal již počátkem století značkou Dentyne .

Za synonymickou koreferenci označujeme rovněž vztahy s pojmenovanými entity, i když to není úplně samozřejmé. Srov.

(3/In94204_87.t) Islámští fundamentalisté z Alžírsko včera zveřejnili výhrůžku na adresu Maroka , že sáhnou k násilí , budou - li Alžířané v Marockém království {coref_text, typ_SYN na „Maroko“} vystaveni špatnému zacházení .

(5/In94204_87.t) Maroko {coref_text, typ_SYN na „království“} včera poněkud zmírnilo svou vízovou politiku vůči Alžířanům , když v několika marockých městech zazněly hrozby pumových atentátů .

Typ SYN označujeme také mezi úplným pojmenováním a zkratkou (ČR - Česká republika, ODS - Občanská demokratická strana).

B.2.3. Koreference hyponymu a hyperonymu {coref_text, typ_ER}

Koreferující člen je vyjádřen hyperonymním výrazem ve vztahu k antecedentu. Pojmy hyperonym a hyponym chápeme v širokém smyslu se zaměřením k referenci. Tento vztah označujeme jenom u párů se specifickou referencí.

Protože tenhle Adolf Hitler nebyl vůdce velkoněmecké říše, ale pták druhu tučňák královský. A to jméno dostal vlastně dodatečně. (Zábrana, J., Vražda v zastoupení)

[...] a kolem trůnu duha jako smaragdová . Mari odstrčí židli , na které sedí, a postaví se." Duha?" Kněz přiloží prst k tomu slovu, aby nezapomněl, kde skončil. (Ludva, R., Jezdci pod slunečníkem)

Často s časovými údaji:

A tak jsme ten rok vyjeli o něco dříve - koncem června . V tu dobu tam je sice ještě moc velká zima na koupání, ale ryby berou výborně . (Haňka, L., S puškou a udicí Severní Amerikou)

Koreferenci typu ER označujeme v párech sloveso (situace) – obecný název té situace. Srov.

(11/cmpr9410_001.t) Jistotu v tomto směru dávají nejnovější kroky vlády SR , která se rozhodla zavést již před časem avizovanou desetiprocentní dovozní příirážku na zboží zahraniční [provincie]provenience .

(12/cmpr9410_001.t) Byť má na tento krok {coref_text, typ_ER na „rozmluva“} určité právo (jako člen GATT) , v daném okamžiku však vyznívá jako tvrdé politické rozhodnutí vlády , která se snaží velice rezolutními administrativními kroky zredukovat mnohamilionové pasívum v obchodní výměně s ČR .

B.2.4. Koreference nereferenčních a generických NP {coref_text, typ_NR}

Koreferenční vztah hlavně generických jmenných frází.

(19/cmpr9410_028.t) Zkratka doby na investování do obrazů bývaly lepší , ale obliba některého umění a vkus lidí jsou věci nevyzpytatelné.

(22/cmpr9410_028.t) Při výběru obrazu {coref_text, typ_ER na „obraz“} bude hrát určitě velkou roli autor .

Pozor! Nespecifickou referenci zaznamenáváme u uzlů závislých na „konteinerech“ (sklenice mléka apod.)

B.3. Problématické případy označování textové koreference

B.3.1. Koreference s typem NR vs. nic

Koreference NP s nespecifickou referencí je poněkud jiného druhu, než koreference NP se specifickou referencí. V některých případech není úplně jasné, kde se taková koreference končí. V následujícím případě koreferenci už nezaznamenáváme. Srov.:

(76/In94204_107) Chceme , aby ze sebe problémy dostaly ven .

(77/In94204_107) Jakmile se začnou svěřovat , už se s tím dá něco dělat .

(78/In94204_107) Jde o to , aby se z problémku nestal problém .

B.3.2. Koreference abstraktních jmen

V anotaci se snažíme rozlišovat specifickou a nespecifickou referenci i u abstrakt. Děláme to tak, že pokud daná NP má jasně specifickou referenci, dáme jí t coref_text, typ 0. Pokud tomu tak není, označíme ji jako coref_text, typ NR. Srov. např.

(32/In94204_107) Přiznal, z čeho má strach. [...]

(35/In94204_107) Všechno nakonec dobře dopadlo, ale tohle dítě zbytečně prožilo půl roku strachu {coref_text, typ_0 na „strach“ nebo žádná šipka} a děsivých představ.

Srov. také

(33/In94208_11) Dovožoval , že vývoj kapitalismu se historicky vyznačuje dvěma fázemi : Fází soutěžního kapitalismu a fází kapitalismu trustů .

(43/In94208_11) Schumpeter se ve svém posledním díle ptá : Který systém , kapitalismus {coref_text, typ_0 na „kapitalismus“}, či (socialismus , bude určovat budoucnost lidstva ?

44/In94208_11) K údivu , úžasu či ohromení většiny svých kolegů odpovídá jednoznačně : Bude to socialismus {coref_text, typ_0 na „socialismus“} .

Srov. s nespecifickou referencí:

(19/In94208_11)Tímto faktorem je podnikatel - inovátor , který se snaží o zisk , a proto logicky nemůže existovat ve stavu statiky , která nezná ani zisk {coref_text, typ_NR na „zisk“}, ani ztrátu .

(27/In94208_11) Na konci tohoto difusního procesu se systém vrátí ke statické rovnováze , v níž nebudou opět ani zisky {coref_text, typ_NR na „zisk“}, ani ztráty .

B.3.3. Dvě místní určení vedle sebe (tady v Praze, u nas doma apod.)

Pokud v tektogramatickém stromě jsou dvě místní určení jako sestry a nejsou přitom ve vztahu apozice, propojíme je textovou koreferencí postupně po sobě. Srov.

(42/In94207_76) Na stůl přinášel kuchyni studenou , chlebičky , uzeniny , šunku a západoněmecké sýry mnoha druhů a zde jsem žasl nad jejich kvalitou , kterou jsem z domova nepředpokládal .

(46/In94207_76) Při tlumeném světle přicházela na přetřes politická situace u nás (coref_text, typ=0 na „domov“) doma (coref_text, typ=0 na „#PersPron“) .

B.4. Nejednoznačný výběr antecedentu

B.4.1. K otázce výběru antecedenta v případě apoziční skupiny:

Pokud koreferující uzel odkazuje na apoziční spojení, koreferenční šipka vede na spojku. Srov.

(35/In95047_061) Zastihli jsme tady pouze jediného vychovatele - pana Fuchse .

(37/In95047_061) Podle pana Fuchse {coref_text, typ_0 na „#Dash“} nejde o žádné kriminálníky ani delikventní mládež .

B.4.2. K otázce výběru antecedenta v případě koordinační skupiny:

Pokud jde o identickou koreferenci, šipku vedeme na spojku nebo od ní (srov. jinak u bridging C.3.3.).

(34/cmpr9410_028) Vznik moderního umění se spojuje s rokem 1907 , kdy byla založena populární Osma , a v kubistickém a fóbistickém duchu malují Filla , Kubišta , Špála a další .

(35/cmpr9410_028) Ceny jejich {coref_text, typ_0 na spojku „a“} obrazů šplhají do statisíců a dobře se prodávají v cizině .

B.4.3. Spojení se slovy s funkcí „kontejneru“

Ve spojení se slovy s funkcí „kontejneru“ (spousta, řada, milion apod.) šipka vede na kontejner. Srov. ne zcela intuitivní příklad:

(43/In94207_76) Ale přitom hostitel otevíral láhve alkoholu .

(44/In94207_76) Byla to {coref_text, typ_NR na „láhev“} vína a německé alkoholy tvrdé .

B.4.4. Koreference jmen s funktorem ID

Jména s funktorem ID nereferují, šipka vede vždy na řídicí uzel. Srov.

(28/In95047_061) Posledním místem , kam byli chlapci firmou Struktura umístěni , byl bývalý dům dětí a mládeže v Duchcově .

(48/In95047_061) Tajemná Struktura {coref_text, typ_0 na „firma“}

(67/In95047_061) Pavel Vondráček : " Termín {PAT} převýchova {ID} znám pouze z nacistického a komunistického slovníku .

(68/In95047_061) Na převýchovu {coref_text, typ_0 na „termín“} se pokud vím , posílali ti , kteří měli podle těchto zruďných režimů nevhodný původ .

C. Bridging vztahy

Bridging vztahy označujeme u NP, které nejsou koreferenční, ale mezi nimi existuje určitý sémantický vztah. Označujeme několik typů bridging anafory.

Pozor! Oproti identické koreferenci (gramatické a textové – viz A), pomocí bridging-vztahů propojujeme pokud možno jenom lexikálně vyjádřené autosémantické uzly (žádné „který“ a rekonstruovaný „PersPron“). V některých případech to nejde, potom můžeme odkazovat i na lexikálně nevyjádřený uzel. Srov. např.:

(24/In9413_006) Třeba při rozvozu #PersPron jsem denně přenesl pěkných pár tun na zádech {bridging, typ=PART, na „#PersPron“}.

C.1. Typologie bridging vztahů

V rámci anotace bridging-anafory vyčleňujeme čtyři následující vztahy a jednu sběrnou pro to, co zatím nemůžeme nikam zařadit:

- **FUNCT (FUNCT_P, P_FUNCT)** – vztah individual – function (oba směry), viz C.1.1.
- **PART (PART_WHOLE, WHOLE_PART)** – vztah část-celek (oba směry), viz C.1.2.
- **SET (SUB_SET, SET_SUB)** – vztah množina-podmnožina (element množiny) (oba směry), viz C.1.3.
- **CONTRAST** – vztah sémantického protikladu, viz C.1.4.
- **REST** – blíže neupřesněná kategorie: je vyznačený bridging vztah, ale není specifikovaný jeho druh, viz C.1.5.

Pozor! Bridging vztah neoznačujeme pokud uzly jsou spojené jednou závislostní šipkou, přičemž závislý uzel má funktor aktantu APP nebo AUTH. Takto spojené uzly mají mezi sebou velice často sémantické vztahy, z kterých mnoho se dá zařadit mezi vztahy typu bridging. Avšak tyto uzly už jsou dostatečně spojené závislostní šipkou a zatěžovat strom dalšími šipkami v tomto případě se nám zdá zbytečné. Srov. páry typu *obyvatelka obce* (vztah PART, funktor PAT), *opat kláštera* (vztah FUNCT, funktor APP), *starosta obce* (vztah FUNCT, funktor PAT), *člen výboru parlamentu* (vztah PART, funktor APP), dílo Wagnera (funktor AUTH), *jejich obrazy* (funktor AUTH) apod.

??? V některých případech však toto pravidlo nejde dodržovat pravidelně. (Pokud původní nepronominální antecedent je v jiné větě?) Srov.

(81/In94207_76) Kdykoliv jsem navštívil Mnichov nebo jím projížděl , neopomněl jsem zajít do pověstné mnichovské pivnice Bierbräukeller .

(82/In94207_76) Její obrovská hala {bridging, typ=PART, na „pivnice“}, kde se sedělo u dubových stolů a bavorské pivo se nalévalo do litrových korbělů s příklopkami .

C.1.1. bridging-vztah individual – function (FUNCT_P a P_FUNCT)

Vztah FUNCT funguje v obou směrech (funkce před objektem nebo po něm). V anotaci používáme FUNCT_P – pro vztah funkce – objekt, P_FUNCT – pro vztah objekt – funkce. Srov.

(52/In95047_061) Mezitím starosta obce Košťany Jindřich Abrhám požádal o pomoc člena branně – bezpečnostního výboru parlamentu poslance Čapka (Levý blok), který se obrátil na ministra Rumla s žádostí o prošetření činnosti firmy Struktura.

(53/In95047_061) Ministerstvo {bridging, typ=FUNCT_P, na „ministr“}vnitř však odpovědělo , že nejsou známy žádné negativní informace o její činnosti .

(25/In95047_061) Někteří se vrátili do Německa, další přešli na faru v nedalekém Jeníkově a spojili se se skupinou , která tady byla již ubytována.

(26/In95047_061) Podobné zkušenosti jako v Oseku potvrdil i farář {bridging, typ=P_FUNCT, na „fara“} v Jeníkově pan Matfiak.

C.1.2. bridging-vztah část-celek (PART_WHOLE a WHOLE_PART)

Vztah části a celku funguje v obou směrech. V anotaci používáme PART_WHOLE – pro vztah celého k části, WHOLE_PART – pro vztah části k celku. Srov.

(44/In9413_006) Dělal jsem bez přestávky celé týdny , často v nocí {bridging, typ=WHOLE_PART, na „týden“} .

Vztah bridging_PART označujeme také v případě neodlučitelných částí, např. u míst a geografických názvů. Srov. páry *Maroko - marocká města*, *Maroko – Marrákeš*, *Německo – Bavorsko – Mnichov* apod.

Pozor! Bridging vztah typu PART neoznačujeme pokud jsou uzly propojeny vztahem ACMP. Srov.

(74 /In94207_76) Na břehu Starnberského jezera u místa utonutí byla postavena kaplička s královými daty {žádná koreferenční šipka na „kaplička“} narození , vlády a smrti , s křížem {žádná koreferenční šipka na „kaplička“} a mramorovou pamětní deskou {žádná koreferenční šipka na „kaplička“} .

Pozor! (vztah město - muzeum)

Vztah mezi např. městem a tím, co v tom městě je, NENÍ bridging_PART, tedy neanotujeme. S podobnými vztahy se v textech setkáváme dost často. Přesně řečeno, muzeum není část města a obraz není část galerie. Srov. např.

(49 /In94207_76) V Mnichově jsou muzea {bridging, typ=PART, na „Mnichov“???) a galerie {žádná koreferenční šipka} se vzácnými obrazy {žádná koreferenční šipka}, částečně (podle času) jsem je navštívil a zhlédl překrásný královský zámek {žádná koreferenční šipka} Nymphenburg .

C.1.3. bridging-vztah množina-podmnožina (SET_SUB a SUB_SET)

Vztah mezi množinou a podmnožinou nebo elementem množiny. Tento vztah je oboustranný. V anotaci používáme SET_SUB – pro vztah množina – podmnožina (element), SUB_SET – pro vztah podmnožina - množina. Srov. např.:

(14/In94210_95) Na rozdíl od dobře vybaveného FS dnes nikdo z téměř dvou stovek poslanců kromě předsedy {bridging, typ=SET_SUB, na „poslanec“} a místopředsedů {bridging, typ=SET_SUB, na „poslanec“} sněmovny a šéfů {bridging, typ=SET_SUB, na „poslanec“} jejich výborů nemá svou kancelář , pracovní stůl , židli a telefon .

Sporný příklad 1 (vztah SET uvnitř jedné věty).

Vztah typu SET mezi subjektem a predikátovou částí výpovědi. Hromadí se šipky, ale nese sémantickou informaci. Zatím označujeme. Srov.:

(43/In94204_107) Jsou mezi nimi například studenti vysokých škol {bridging, typ=SET, na „#PersPron“}, herečka {bridging, typ=SET, na „#PersPron“}, kunsthistorik {bridging, typ=SET, na „#PersPron“}, učitelka {bridging, typ=SET, na „#PersPron“}, psycholožka {bridging, typ=SET, na „#PersPron“}.

Sporný příklad 2 (otázka nutnosti a hloubky interpretace).

V následujícím příkladu označíme vztah bridging SET:

(44/ In94204_107) [volontéři] Absolvovali školení v první pomoci pro člověka v nouzi . [...]

(56/ In94204_107) Když dítě zavolá , dostane buď radu hned , nebo si s ním volontér {bridging, typ=SET na „volontér“} domluví další hovor .

Sporný příklad 3 (označování některých bridging vztahů u nereferenčních NP)

(51/In94210_95) Ani pro parlamentní knihovnu nezbude na Malé Straně místo .

(53/In94210_95) Umístit knížky {žádná koreferenční šipka} na speciálně upravených půdách , jako je tomu v Národní knihovně v Klementinu , prý z bezpečnostních důvodů nelze.

V následujícím příkladě bridging vztah SET neoznačujeme. NP „knihovna“ v (51) má specifickou referenci, stejně jako v (53), zatímco NP „knížka“ v (53) má generickou referenci, nemůže tedy být elementem „knihovny“ v (51) ani „knihovny“ v (53).

Hraniční případy mezi bridging typy SET a PART.

Rozdíl mezi typy SET a PART je dost vágní a řešení často závisí na počitatelnosti daného substantiva. Zatím však označujeme oba vztahy. Srov.

(47/In94210_95) Jeho hlavní výhodou by mělo být lepší napojení na televizní přenosovou techniku : zatímco dnes přenosové vozy {bridging, typ=SET nabo PART, na „technika“} blokují parkovací prostor před starou sněmovnou , v budoucnu zajedou do Thunovské a kabely {bridging, typ=SET} se snadno spojí s tiskovým centrem .

(73/In94208_11) Když ho smrt překvapila u psacího stolu , revidoval právě text Prezidentské adresy , kterou pronesl několik dní před tím v Americké ekonomické asociaci .

(74/In94208_11) Poslední věta {bridging, typ = PART, na „text“}, kterou v životě napsal , zněla : Stagnacionisté se mýlí v diagnóze důvodu , proč by kapitalistický proces měl stagnovat .

C.1.4. bridging-vztah sémantického protikladu (CONTRAST)

Vztah protikladu v textu. Srov.

(53/In9413_006) A přesvědčen jsem ještě o jednom - je třeba mít vysoké cíle a s malými [cíly] {bridging, typ=CONTRAST, na „cíl“} se nespokojit.

(13) Lidi nežvýkají , to jenom krávy {bridging, typ=CONTRAST, na „člověk“}.

Pozor! Bridging vztah typu CONTRAST neoznačujeme v případě když dané NP jsou potomky uzlu s funktoem ADVS, většinou spojky nebo čárky. Tento funktoer vztah kontrastu už vyjadřuje. Srov.

(17/In94208_11) Dočasný podnikatelův zisk bude anulován , ale trvalý zisk {žádná koreferenční šipka} z jeho inovace zůstane zachován společnosti ve formě nižších cen nebo technicky dokonalejších výrobků .

C.1.5. bridging-vztah zatím nedeterminovaný (REST)

Do dané skupiny patří příklady, které by se teoreticky šlo zařadit do nové skupiny, ale nechceme zatím množit typy. Nedáváme tam všechno, co s daným uzlem sémanticky souvisí, ale jenom to, co bychom mohli potom příp. klasifikovat. Srov. některé možné skupinky:

Vztah rodinné příslušnosti

(10/In9413_006) Úzce navazuje na tradici podnikání svého rodu, především dědy.

(12/In9413_006) Od něj získal vnuk {bridging, typ_REST na „děda“} výtečné základy , ač sám vystudoval školu zaměřenou na dopravu.

Vztah „místo – obyvatel“.

(35/In94207_84) Poté , co byl v roce 1845 jako prezident svržen a na deset let vypovězen na Kubu , vydal se do New Yorku s jedinou myšlenkou - získat zpět vládu nad Mexikem .

(38/In94207_84) Tak jako každý Mexičan {bridging, typ_REST na „Mexiko“}, i Santa Anna znal a občas žvýkal mízu sapidilly zvanou chicle

Vztah typu „autor – kniha“

(Pokud není označen funktorem AUTH)

(22/cmpr9410_028) Při výběru obrazu bude hrát určitě velkou roli autor {bridging, typ_REST na „obraz“}.

Vztah „věc – majitel“

Neoznačujeme, pokud je zachycen funktorem.

(54/cmpr9410_028) Obraz výrazně stoupne na ceně , má - li majitel {bridging, typ_REST na „obraz“} doklad o tom , že byl vystaven na výstavě , či je publikován v knize či katalogu .

Vztah mezi stejně vyjádřenými nebo synonymními nekoreferenčními NP

Vztah mezi uzly, které jsou stejně lexikálně vyjádřeny nebo synonymní, nejsou koreferenční (třeba jeden člen má generickou, druhý však specifickou referenci) ale které přitom podílejí na koherenci textu.

a) stejné NP:

(77/In95047_061) " Sever Čech má za sebou svízelnou minulost , má před sebou po skončení vlády komunistů novou naději , " domnívá se Raimond Strathman , člověk zodpovědný za akci Evangelické diakonie v České republice .

(78/In95047_061) " Staráme se o děti , které se ne vlastním přičiněním dostaly do těžké situace .

(79/In95047_061) Vždyť i ony mají před sebou novou naději {bridging, typ_REST na „naděje“} .

(39/In94207_84) Právě v té době přihrála náhoda Santa Annovi do cesty Thomase Adamse , fotografa a především vynálezce všeho druhu .

(45/In94207_84) Psal se rok 1869 a do hry vstoupila další náhoda {bridging, typ_REST na „náhoda“}.

b) synonymní NP:

(23/In95047_061) Měli časté incidenty s městskou policií : " Díky tomu , že měli dostatek finančních prostředků , byli často opilí , nabalovala se na ně místní mládež , a také si brali do kláštera holky , " vzpomíná obecní strážník .

(27/In95047_061) I tady si prý chlapci , kteří měli být vychováváni na faře , užívali děvčat {bridging, typ_REST na „holky“} a svobody .

(20/In94207_76) Stačilo jen razítko na hranicích , celní kontroly jejich orgány nedělaly , náš pas - tedy to , že jsme z Československa , byla sama o sobě průkazná vizitka a vstup na jejich území (a stejně tak i na další) byl hladký .

(23/In94207_76) Naši celníci už nás čekali , viděli nás , jak vyjíždíme od Němců , zastavili jsme , dali štempl {bridging, typ_REST na „razítko“} a měli jsme jet dál .

Vztah událost – argument

(11/cmpr9410_031.t) Relativně tak stát vynakládá na tržně konformní podporu malého a středního podnikání přibližně [1,6]1.6 - [1,8]1.8 % hrubého domácího produktu .

(13/cmpr9410_031.t) V rámci rozpočtové podpory poskytují ministerstva malým a středním podnikatelům {bridging, typ_REST na „podnikání“} zvýhodněné informační služby a poradenskou činnost buď přímo nebo prostřednictvím specializovaných institucí .

(14/cmpr9410_031.t) Potěšitelné ovšem je , že podpora malého a středního podnikání {bridging, typ_REST na „podnikatel“, coref_text, typ_0 na „podnikání“} má výrazný regionální aspekt .

Srov. také páry: *spor – účastník konfliktu, léčebna – léčba* apod. Tyto vztahy zatím neoznačujeme, ale nejsem si tím úplně jistá.

C.2. Nejednoznačný výběr antecedentů

C.2.1. Spojení se slovy s funkcí „kontejneru“

Ve spojení se slovy s funkcí „kontejneru“ (spousta, řada, milion apod.) bridging šipka vede na kontejner, stejně jako v případě textové koreference (viz B.4.3.). Srov.

(57/In94207_84) Když o deset let později obrátil ke gumě pozornost louisvilleský lékárník John Colgan , existovala již řada žvýkačkových milionářů (mezi nimi Adams {bridging, typ_SET na „řada“, nikoliv na „milionář“}) .

C.2.2. K otázce výběru antecedentu v případě apoziční skupiny

V případě apoziční skupiny (u koreferujícího člena nebo antecedenta), podobně jako u textové koreference (viz B.2.4.1.), šipku vedeme na konektor (resp. od něj). Srov.

(21) Vzpomínám na takzvané zelené hranice zcela bezbariérové a na dosud nevídanou blahovůli zahraničních a našich celních a policejních orgánů už na jaře 1968 , tedy ještě chvíli před zábořem , kdy jsme jeli autem na výlet do Západního Německa s Hankou Bělohradskou a Dušanem Hamšíkem .

(22) V nějakém městečku za Schirndingem , #Comma/APPS {bridging, typ_PART na „Německo“} snad v Marktrewitzu , jsem si koupil vynikající umělé květiny , jaké se u nás neviděly , také pětaticetimetrovou opici , huňatou , milou a pár podobných i komických drobností jiných .

C.2.3. K otázce výběru antecedentu v případě koordinační skupiny:

Pokud jde o bridging vztah, šipkou spojujeme autosémantické uzly, nikoliv spojku (v případě identické koreference je to jinak – viz B.2.4.2.). Srov. např.

(24/cmpr9410_028) Minulé století je bohaté na slavná jména .

(25/cmpr9410_028) Snad vůbec nejvýznamnější jsou obrazy Karla Purkyněho {bridging, typ_SET na „jméno“} a Jaroslava Čermáka {bridging, typ_SET na „jméno“} .

D. Bridging nebo textová identická – výběr a preference

D.1. K otázce dvojího odkazování (textový a bridging vztahy) a preference:

Pokud NP odkazuje k bližšímu antecedentu bridging vztahem, přičemž je to důležité pro koherenci textu, a zároveň má (třeba i ve vzdálenějším kontextu) textově koreferenční uzel, preferujeme textovou koreferenci. Srov.

(11/In95047_061) V obci Košťany na Teplicku ještě chlapci ani nebyli , ale místní již dali dohromady petici : " My rodiče dětí základní školy Košťany protestujeme proti umístění ubytovny pro potrestané německé chlapce .

(12/In95047_061) Víme , že na našem území páchali dál trestnou činnost a nevhodně pokřikovali na kolemjdoucí .

(13/In95047_061) Pokud zástupci města neprovedou opatření proti umístění těchto německých chlapců v našem městě {coref_text, typ=SYN na „obec“}, nebudeme posílat svoje děti do školy z důvodu strachu o jejich bezpečnost .

D.2. „faktory – jeden z faktorů“

Situaci, když v antecedentu je NP „x“, druhý člen páru je konstrukce „jeden (některé) z x“ řešíme jako v následujícím příkladu:

(16/In94208_11) V praxi se tato rovnováha realizuje tím , že se každý faktor³ chová rutinním a adaptivním způsobem .

(17/In94208_11) Dynamika kapitalistického systému vzniká teprve tím , že se jeden {bridging, typ_SET na „faktor“ v (16)} z faktorů {coref_text, typ=0 na „faktor“} začne chovat netradičně .

³V daném případě sice jde o distributivní referenci, ale v daném kontextu je to totéž co „všechny faktory“, takže považujeme to za množinu.

Podobně se chovají generické NP ve spojeních s „kontejnery“. Propojujeme oba řetězce – jak kontejnery tak i to co obsahují. Srov.:

(73/In94207_84) V rámci reklamní kampaně předal osobně každému členu washingtonského Kongresu jednu krabici {specifická reference, distributivní} své gumy {generická reference} značky Yucatan .

(76/In94207_84) I král dostal svou krabici {specifická reference, žádná šipka} gumy {generická reference, coref_text, typ=NR na „guma“ v (73)} a nádavkem i doslova trhoveckou prezentaci .

(77/In94207_84) Samozřejmě , že novinové zprávy o králi s krabicí {specifická reference, coref_text, typ=0 na „krabice“ v (76)} gumy {generická reference, coref_text, typ=NR na „guma“ v (76)} byly reklamou k nezaplacení .

(78/In94207_84) I když konzervativní Anglie jeho čin odsoudila , guma {generická reference, coref_text, typ=NR na „guma“ v (77)} se zde chytla a Británie se pro žvýkačku stala bránou do Evropy .

E. Speciální typy reference (coref_special)

E.1. Exoforické odkazy

Exoforické odkazy jsou již označeny, pokud odkazuje zájmeno. Když však exoforický odkaz obsahuje jmenná fráze, anotujeme to na etapě rozšířené koreference. K takovým případům patří především časová (v *tomto roce*) ale i prostorová deixe. Srov.:

(28/In94210_95) Dokončeny by měly být do 31 . prosince 1995 , a to i přes jisté zdržení způsobené opožděným stěhováním nájemníků z domů čp. 8 a 518 do náhradních bytů na sídlišti Barrandov v těchto dnech {coref_special, typ exoph} .

(11/cmpr9410_001.t) Byť má na tento krok určité právo (jako člen GATT) , v daném okamžiku {coref_special, typ exoph} však vyznívá jako tvrdé politické rozhodnutí vlády , která se snaží velice rezolutními administrativními kroky zredukovat mnohamilionové pasívum v obchodní výměně s ČR .

V posledních letech {coref_special, typ exoph} lidé znovu začínají stále více využívat vodu čerpanou ze studní.

Srov. také exoph odkazy u adverbii:

(104/In94207_84) A tu {coref_special, typ exoph} se dostáváme zpět k počátku tohoto textu.

Pozor! Exoforickou šipku neanotujeme v případě když deiktický výraz je součástí lexikální sémantiky daného slova (výrazy typu *dnes, zítřka, letos* apod.)

E.2. Odkazy na segmenty textu

Odkazy na segmenty textu dodáváme v případech, kdy jmenná fráze (většinou s identifikátorem) odkazuje na více než jednu větu v předchozím kontextu. Srov. např.

(10/cmpr9410_001.t) Celní unie bude sice existovat na papíře ještě dalších dvanáct měsíců (a třeba i déle) , ale v praxi dostanou vzájemné vztahy punc tvrdosti mezinárodního obchodu .

(11/cmpr9410_001.t) Poroste administrativa

(10/cmpr9410_001.t) Jistotu v tomto směru {coref_special, typ segm} dávají nejnovější kroky vlády SR , která se rozhodla zavést již před časem avizovanou desetiprocentní dovozní přírážku na zboží zahraniční [provinencie]provenience .

(38/In94207_84) Tak jako každý Mexičan , i Santa Anna znal a občas žvýkal mízu sapodilly zvanou chicle (prý z mayského slova tsictle) , a tak se zrodil nápad pokusit se z chicle udělat náhražku kaučuku .

(39/In94207_84) Právě v té době {coref_special, typ segm} přihrála náhoda Santa Annovi do cesty Thomase Adamse , fotografa a především vynálezce všeho druhu .