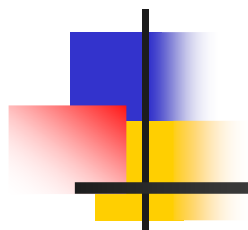
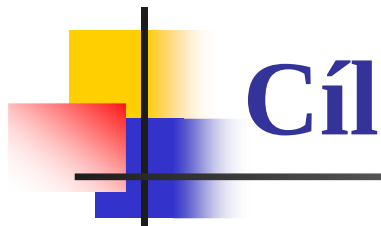


Automatická analýza časových vztahů v diskurzu

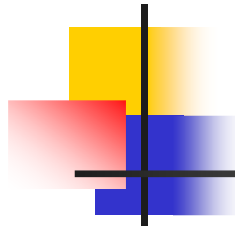


Petr Němec
ÚFAL, 26.2.2007



Cíl

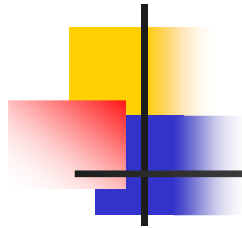
- určení časového uspořádání dějů vyjádřených v diskurzu
 - manuální anotace $\Rightarrow$ korpus
 - automatická analýza
- (navíc) zachycení významu časových výrazů a jejich použití pro automatické vyvozování relací



Příklad

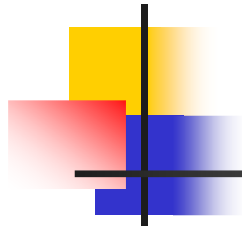
Konsorcium soukromých investorů fungující jako BPH Funding Co. včera sdělilo, že by mohlo dát nabídku ve výši 2 mil. dolarů. Dnes prohlásilo, že tuto možnost již nezvažuje.

- *sdělilo* platí v minulosti (včera)
- *(by mohlo) dát* platí v budoucnosti (pokud vůbec)
- *prohlásilo* platí v minulosti (dnes) po *sdělilo*
- *nezvažuje* platí v okamžiku promluvy
- *fungující* platí pravděpodobně v okamžiku promluvy
- ...a další



Motivace

- krok ke zpracování přirozeného jazyka obecně
- strojový překlad
 - např. bohatší časový jazykový systém angličtiny vyžaduje znalost řazení dějů (jako součást vstupu z časově jednoduššího jazyka, např. češtiny) pro použití některých časových konstrukcí v generování anglické věty: „Čekala tam už dvě hodiny, když přišel.“ $\Rightarrow$ „She had been waiting there for two hours when he came.“
- information retrieval
- sumarizace textu
- zodpovídání otázek
- a další...



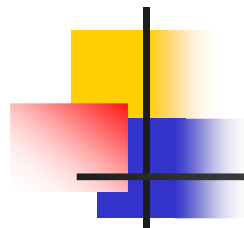
Dnešní přednáška

- anotační schéma
- anotovaný korpus
- Panevové algoritmus jako základ systému automatické analýzy
- protipříklady a rozšíření tohoto algoritmu



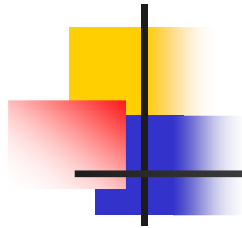
Část I

Anotační schéma



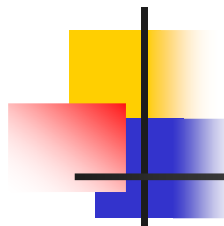
Základní principy

- bod začátku a konce každého (tvrzeného, negovaného, žádaného, ...) děje vyjádřeného v diskurzu
 - odpovídá pojetí Vaška Nováka [Novák, 2004] tak, jak je zachycuje ve své diplomové práci
 - děje jsou vyjádřeny primárně slovesy, někdy též adjektivy a substantivy
- okamžik promluvy pro každou promluvu v diskurzu
- relace ($<$, $>$, $\leq$, $\geq$, $=$) určují uspořádání na takto vzniklém *časovém prostoru*



Metodika

- anotujeme posloupnost dějů tak, jak je vyjádřena mluvčím v dané promluvě
- určujeme relace tak, jak odpovídají našemu „přirozenému“ chápání diskurzu (článku), nikoli pouze nutně platné relace
- mezivětné relace pouze mezi sousedními souvětími
 - počet potenciálních relací roste kvadraticky s počtem dějů, je téměř nemožné (přinejmenším časově) spolehlivě určit relace v celém diskurzu
- kompletní anotace je dána tranzitivním uzávěrem vložených relací



Příklad

- 1) *Konsorcium soukromých investorů fungující jako BPH Funding Co. včera sdělilo, že by mohlo dát nabídku ve výši 2 mil. dolarů.*
- 2) *Dnes prohlásilo, že tuto možnost již nezvažuje.*

Časový prostor: *1.p, 2.p* (okamžiky promluvy), *fg.s, fg.e* (fungující), *sd.s, sd.e* (sdělilo), *dt.s, dt.e* (dát), *pr.s, pr.e* (prohlásilo), *nzv.s, nzv.e* (nezvažuje)

$$1.p < 2.p$$

$$sd.s = sd.e < 1.p$$

$$sd.s < dt.s = dt.e$$

$$pr.s = pr.e < 2.p$$

$$sd.s < pr.s$$

$$nzv.s < pr.s < nzv.e$$

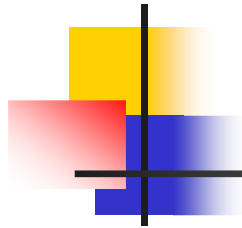
$$nzv.s < 2.p < nzv.e$$

$$fg.s < 1.p < fg.e$$

$$fg.s < 2.p < fg.e$$

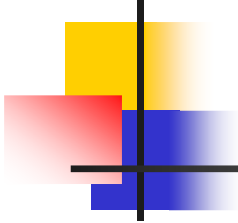
$$fg.s < sd.s$$

$$fg.s < nzv.s$$



Časové výrazy

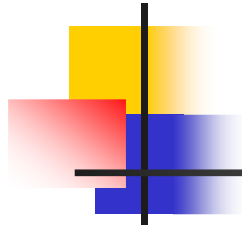
- cílem je zachytit význam výrazů jako *před 5. květnem, během následujícího roku, začátkem léta, v polovině 80.let, apod.*, aby na jejich základě bylo možné automaticky vyvozovat další relace
- funkční přístup - význam je zachycen jako kompozice elementárních funkcí a operátorů, které mohou být efektivně zpracovány a porovnány



Funkce a operátory (zkráceno)

- *yy|mm|dd|hh|mi|ss* vrací *časový_interval* dané entity
- *shift(A:časový_bod,R:časové_rozpětí)* vrací *časový_bod* vzdálený *R* od *A*
- *entity_range(T:typ_entity, N:počet)* vrací *časové_rozpětí* odpovídající *N* entitám typu *T*
- *find(A:časový_bod, E:entita, N:int)* nalezne *N*-tý výskyt *E* od *A*
- *part(I:časový_interval, E:entita, N:int)* nalezne *N*-tý výskyt *E* uvnitř *I*
- *series(A: časový_bod, E:entita, N:počet)* vrací množinu *N* entit *E* od *A*

- konstrukce a manipulace s intervaly
- aritmetika na časových rozpětích
- alternativy : []



Příklady (I)

1) *Vrátí se do pěti minut.*

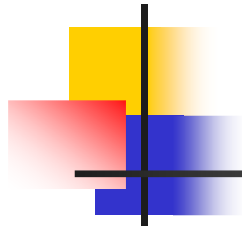
$vrátí.s (= vrátí.e) \leq shift(p, entityRange(minute, 5))$

2) *Výbor se zprávou zabýval pouze tento čtvrtek.*

$zabýval.s \text{ in } find(p, Thursday, -1)$

$zabýval.e \text{ in } find(p, Thursday, -1)$

$(zabýval.s < zabýval.e)$



Příklady (II)

3) *Zásilka dorazí nejdříve v květnu příštího roku.*

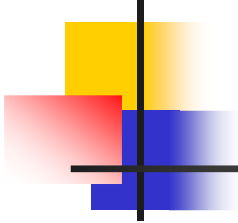
dorazí.s (= dorazí .e) $\geq$ start(part(find(p,year,1),May))

4) *Vedení společnosti bude zvažovat novou strategii ráno na poradě 4.5.2006.*

zvažovat.s in part(2006|5|4,morning)

zvažovat.e in part(2006|5|4,morning)

(zvažovat.s < zvažovat.e)



Příklady (III)

5) *Letadlo, které si America West objednala, bude dodáno příští pondělí nebo úterý:*

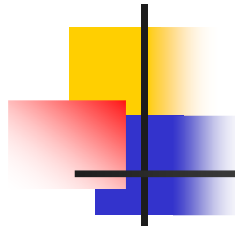
dodáno.s (= dodáno.e) in [

find(p,Monday,1),

find(p,Tuesday,1)

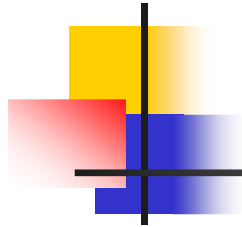
]

- není třeba znát přesnou hodnotu okamžiku promluvy, abychom mohli výraz anotovat a vyvozovat relace uspořádání
 - znalost přesné hodnoty obecně umožňuje více inferencí
- tento mechanismus může být použit obecněji (např. pro absolutní specifikaci trvání děje)



Plán

- zachycení iterativních dějů
- posloupnost dějů v „plánu“ představuje jeden průběh iterativní sekvence dějů
- jsou specifikovány *hranice plánu*
- plán může obsahovat bližší specifikaci opakování (např. *každý den*, *často*, apod.)



Příklad

Minulý měsíc jsem vždy po probuzení 20 minut cvičil.

$P.s = \text{start}(\text{find}(p, \text{month}, -1))$

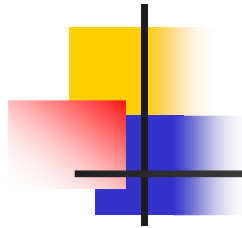
$P.e = \text{end}(\text{find}(p, \text{month}, -1))$

$P.\text{specifier} = \text{entityRange}(\text{day}, 1)$

$P.\text{probuzení}.s = P.\text{probuzení}.e$

$P.\text{probuzení}.e < P.\text{cvičil}.s < P.\text{cvičil}.e$

$\text{distance}(P.\text{cvičil}.s, P.\text{cvičil}.e) = \text{entityRange}(\text{minute}, 20)$



„Distributivní značka“

- je třeba zachytit situaci, kdy se daný děj odehrává pro každého aktora zvlášť

Každá ze společností si vybudovala své sídlo v Bostonu.

... $D.vybudovala.s = D.vybudovala.e < p$...

- plány a distributivní značky se mohou kombinovat

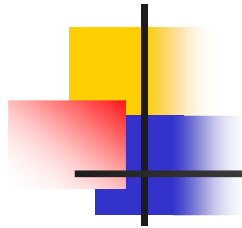
Mnoho lidí každé ráno vstane a cvičí.

... $D.P.s < p < D.P.e$...



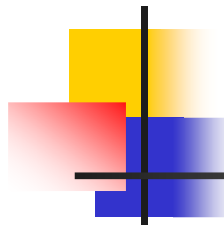
Část II

Korpus



Datové zdroje

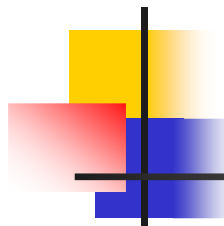
- Prague Czech-English Dependency Treebank:
 - paralelní korpus, Wall Street Journal
 - český překlad testovacích dat (části sekcí 22, 23 a 24)
 - manuálně vytvořené tektogramatické stromy pro tyto věty
 - DTest: **233** vět ve **12** člancích, ETest: **231** vět v **10** člancích
 - sémanticky omezená doména - vhodný základ pro převod do logické formy
 - stejná data použita v projektu strojového překladu z češtiny do angličtiny, možnost ověřit význam časové informace pro překlad
 - **nevýhoda:** PDT 1.0 styl, nestandardní stromy, chyby (gramatémy času a vidu byly manuálně opraveny)



Anotované entity

typ entity	počet	poměr
slovesa	660	61.57%
body promluvy	233	21.74%
substantiva	97	9.05%
adjektiva	51	4.76%
hranice plánů	31	2.89%
celkem	1072	100%

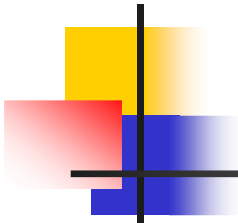
celkem bodů: **1687**



Anotované relace

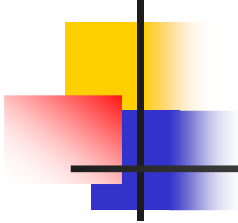
	bod promluvy	sloveso	substantivum	adjektivum	hranice plánu
bod promluvy	3124 (2.88%)	X	X	X	X
sloveso	29008 (26.74%)	45428 (41.87%)	X	X	X
substantivum	4068 (3.75%)	11671 (10.76%)	907 (0.84%)	X	X
adjektivum	2180 (2.01%)	5329 (4.91%)	439 (0.40%)	403 (0.37%)	X
hranice plánu	1431 (1.32%)	3649 (3.36%)	482 (0.44%)	234 (0.22%)	148 (0.14%)

celkem relací: **108501**



Srovnání s jinými přístupy (I)

- Temporal Discourse Models [Mani, Pustejovsky, 2004] a anotační schéma TimeML [Pustejovsky et al., 2004]
 - informativnější (hierarchie dějů, abstraktní děje)
 - podstatně větší korpus
 - anotace nepředpokládá existenci podkladové struktury ⇒ relevantní informace (syntax, morfologie) je přímo součástí anotace
 - reprezentace časových výrazů se zdá být příliš mělká pro přímé vyvozování relací uspořádání
- též další formálně sémantické přístupy (např. SDRT, RST) většinou přistupují k temporalitě obecněji



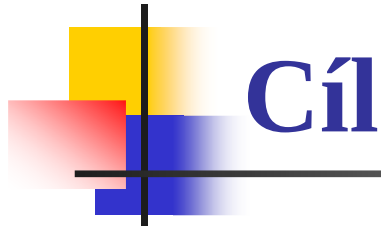
Srovnání s jinými přístupy (II)

- naše anotační schéma je možné vnímat jako „jádro“ (implicitně) sdílené různými formálně sémantickými přístupy
 - možnost využití výsledků dosažených automatickou analýzou pro naše anotační schéma
- funkční přístup je ve srovnání s TimeML vhodnější pro přímé porovnávání extenze časových výrazů

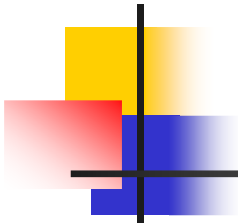


Část III

Automatická Analýza
Panevové algoritmus

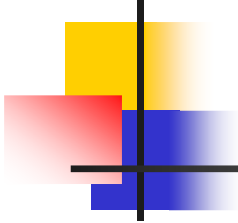


- jednoduchý: vyvinout systém, který na základě tektogramatických stromů pro daný diskurz určí uspořádání dějů automaticky
- strategie:
 - Panevové algoritmus [Panevová 1971], určující vztahy mezi finitními slovesy v rámci jednoho souvětí, zvolíme za páteř systému, jeho úspěšnost je naší baseline
 - pokusíme se opravit vzniklé chyby
 - pokusíme se algoritmus rozšířit

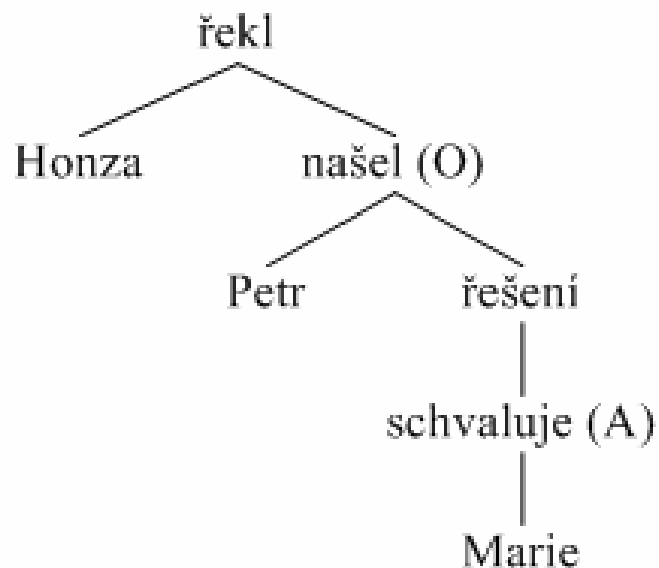


Panevové algoritmus (PA)

- Vyjadřuje rekurzivní vlastnost času v češtině.
- Morfologický čas finitního slovesa (**A**) určuje jeho časový vztah – předčasnost, současnost, následnost – vůči referenčnímu bodu (**R**).
- Je-li **A** řídícím slovesem hlavní věty, **R** je okamžik promluvy.
- Je-li **A** řídícím slovesem obsahové vedlejší věty, **R** je řídící sloveso nadřazené věty.
- Je-li **A** řídícím slovesem jiného typu vedlejší věty, necht' **B** je řídící sloveso první (i nepřímou) nadřazené obsahové vedlejší věty.
Referenčním bodem **A** je referenční bod **B**. (Neexistuje-li takové **B**, je referenčním bodem okamžik promluvy.)



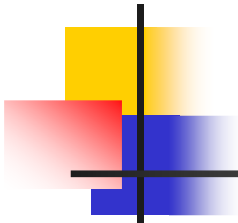
Příklad



$\text{řekl.s} = \text{řekl.e} < p$ (**řekl** $\rightarrow$ **okamžik promluvy**)

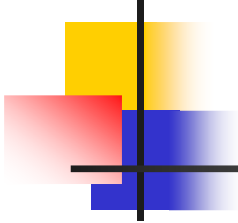
$\text{našel.s} = \text{našel.e} < \text{řekl.s}$ (**našel** $\rightarrow$ **řekl**)

$\text{schvaluje.s} < \text{řekl.s} < \text{schvaluje.e}$ (**schvaluje** $\rightarrow$ **řekl**)



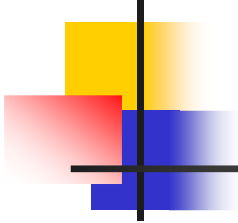
Evaluační metodika (I)

- Pro každý diskurz tranzitivně uzavřeme matici časového prostoru reference (manuální anotace) a hypotézy (výstup systému automatické analýzy).
- Vypočteme **precision P** jako podíl počtu správně určených relací a všech relací v hypotéze.
- Vypočteme **recall R** jako podíl počtu správně určených relací a všech relací v referenci.
- Vypočteme míru **F** jako harmonický průměr P a R.
- Relace je považována za správně určenou, shoduje-li se přesně v hypotéze a referenci.
 - Není-li jeden z bodů správně určen jako součást plánu či označen jako součást „distributivního“ děje, není relace považována za správnou.



Evaluační metodika (II)

- Relace je považována za tzv. slabě správně určenou, je-li namísto správné silné relace ($<, >, =$) určena odpovídající slabá relace ($\leq, \geq$).
- Vypočteme též $\mathbf{P}_w$, $\mathbf{R}_w$ a $\mathbf{F}_w$ jako varianty P, R a F, které
 - slabě správné relace považují za správné
 - neberou ohled na plány a distributivní značky
- Výsledky pro každé pravidlo uvádíme jako rozdíl stavu, kdy je toto pravidlo vypnuto, oproti stavu, kdy jsou všechna pravidla zapnuta (celkový výkon systému).
 - Pravidla totiž nejsou nezávislá, některá vyžadují pro správný chod předchozí aplikaci jiných.



Výsledky pro PA

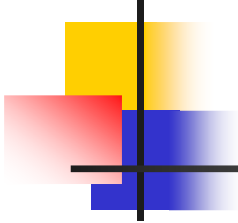
- **P:** 69.51
- **R:** 37.45
- **F:** **48.67**

- **P_w:** 78.36
- **R_w:** 42.22
- **F_w:** **54.87**



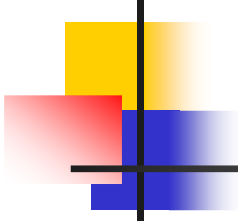
Část IV

Chyby a protipříklady



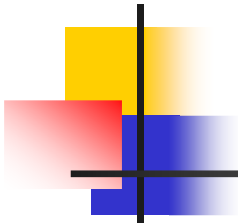
Přímá řeč

- víceméně technická záležitost: tektogramatický strom neposkytuje informaci o tom, zda je dané souvětí součástí segmentu přímé řeči
- “nejvyšším” referenčním bodem má být sloveso promluvy, které tuto přímou řeč uvozuje, nikoli okamžik promluvy-zápisu souvětí s přímou řečí
- vyžaduje detekci segmentů přímé řeči a uvozujících sloves promluvy



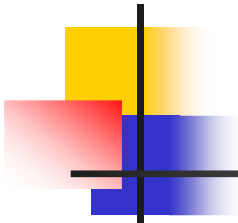
Obsahové věty

- v kontextu časových vztahů a Panevové algoritmu (PA) se některé obsahové věty chovají jako neobsahové a naopak
- “neobsahové” chování:
 - *Udělá (A), co bude považovat (B) za nejlepší. $(B.s \leq A.s \leq B.e)$*
 - prof. Panevová tvrdí, že B je ve skutečnosti přívlastkovou vedlejší větou ($\sim$ to, co)
- “obsahové” chování:
 - *Řekl (A), že firma propustila (B) zaměstnance, aby ušetřila (C). $(B.s \leq C.s)$, žádná relace mezi A a C*



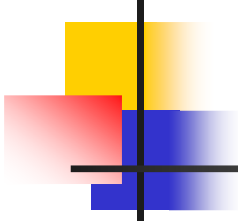
„Výjimečné“ vedlejší věty

- vedlejší věty synonymní v přítomnosti i budoucnosti (v přítomnosti dává v kontextu PA chybné výsledky):
 - *Chce vybudovat podnik, který prosperuje* (= bude prosperovat).
- „komentáře“ (popsané již prof. Panevovou):
 - *Pan Shidler řekl, že společnost chce odkoupit pozemky v Ohiu, které nyní vláda plánuje rozdělit.*
 - Poslední přívlastková věta je komentářem autora článku, není součástí nepřímé řeči.
- žádné zřejmé řešení



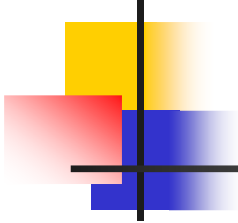
Kondicionály (I)

- v PCEDT je gramatém času téměř vždy nastaven na minulost (odpovídá činnému příčestí), což vede k mnoha chybám
- bylo by třeba určit skutečný „obsahový“ čas příslušného děje
 - *Ty projekty jsou sice velké, ale nešli bychom do nich sami.* (budoucnost)
 - *Nezdá se, že by stávka měla nějaký efekt.* (přítomnost)
 - *Tehdy by si to (býval) vzal, ale ona nechtěla.* (minulost)



Kondicionály (II)

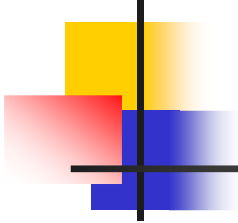
- opět žádné zřejmé řešení
- nevyvozujeme vztahy kondicionálu vůči referenčnímu bodu
- je možné zkusit nastavit nejčastější možnost - budoucnost nebo přítomnost pro nedokonavá slovesa apod., ale obojí vede k mnoha chybám
 - může zlepšit F-measure za cenu poklesu P



Zajímavý kondicionál

- 1) Stavba muzea, které by shromažďovalo moderní umění dvacátého století, byla někdejším francouzským prezidentem Georgesem Pompidouem zadána k řešení zhruba sedmi stovkám architektů v roce 1969.

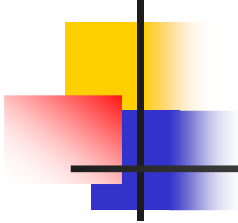
- 2) Stavba muzea, které bude shromažďovat moderní umění dvacátého století, byla někdejším francouzským prezidentem Georgesem Pompidouem zadána k řešení zhruba sedmi stovkám architektů v roce 1969.



Vid slovesa a trvání děje

V některých případech vid slovesa nekoresponduje s trváním příslušného děje:

- “historický přítent“
 - *Pan Otradovec v rozhovoru říká (= řekl, říká.s = říká.e), ...*
 - *Ale Zpráva říká, že otec přijede zítra. (říká.s < říká.e)*
 - V doméně WSJ článků detekujeme životnost aktorů, algoritmus využívá informací z ontologie (EuroWordNet). Obecně však může jít o gnómický děj-stav jako např. *Maminka říká, že nemám chodit do lesa.*
- negované dokonové sloveso:
 - *Boeing zásilku zatím neobdržel. (neobdržel.s < neobdržel.e)*
 - *But Boeing včera v 16:00 zásilku neobdržel (neobdržel.s = neobdržel.e)*
 - *Boeing v dohodnutém termínu zásilku neobdržel. (??)*



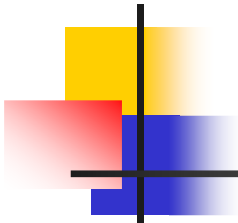
Iterativní děje

- mnoho problémů s časem obecně - typicky jsou všechny relace vůči okamžiku promluvy špatně:
 - *BPH Funding Co. běžně nejprve složí zálohu a poté koupí pozemek.* ($P.s < p < P.e$, $P.složí.s < P.koupí.s$), ale žádný vztah mezi p a $složí$ či $koupí$
- je-li děj iterativní a vyjádřený nedokonavým videm, může (ale nemusí) být bodový v rámci jednoho běhu iterativní sekvence.
 - hranice je neostrá, často jsou obě varianty přijatelné
 - *Společnost běžně nejprve skládá zálohu a až poté kupuje pozemek.* ($D.P.skládá.s = D.P.skládá.e < D.P.kupuje.s = D.P.kupuje.e$)
 - *Pavel denně běhá 20 minut.* ($P.běhá.s < P.běhá.e$)



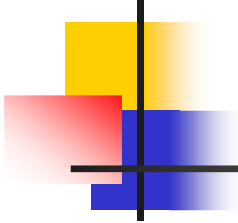
Část V

Rozšíření



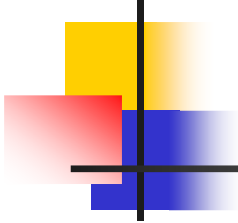
Společná komplexní kotva (I)

- Je-li uvažované sloveso (**A**) nedokonavé a v přítomném čase, je dle PA současné (tj. má společný bod) se svým referenčním bodem (**R**). Je-li ovšem R též progresivního charakteru, nelze vyvodit žádné relace mezi koncovými body obou dějů.
- Je-li R též v přítomném čase, zvolíme za referenční bod A referenční bod samotného R (**R'**). Je-li i R' progresivní přítomný, postupujeme k jeho referenčnímu bodu atd.



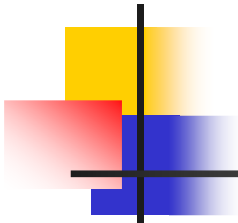
Společná komplexní kotva (II)

- Např. *Pavel řekl, že si myslí, že pracuje dobře.*
 - $(pracuje.s < řekl.s < pracuje.e)$
- Potenciální protipříklad by spočíval v možné netranzitivitě relace současnosti v kaskádě nedokonavých sloves.
- Takový protipříklad nebyl nalezen.



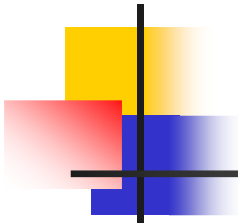
Infinitivy (I)

- Uvažme situaci, kdy infinitiv (I) přímo závisí na autosémantickém finitním slovesu (F).
- Hypotéza: je-li I nebo F dokonavé a
 - F nevyjadřuje děj ukončení (například *přestat*), potom $F.s \leq I.s$
 - F vyjadřuje děj ukončení, potom $I.e \leq F.s$
- Nebyl nalezen žádný protipříklad.



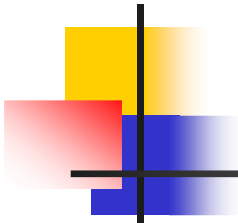
Infinitivy (II)

- je-li F i I nedokonavé, je vzájemná relace zřejmě značně závislá na sémantice obou částí:
 - *To dělníkům pomáhá vyvíjet tlak na zaměstnavatele.*
(možná $F.e \leq I.e$, ale rozhodně ne $F.s \leq I.s$)



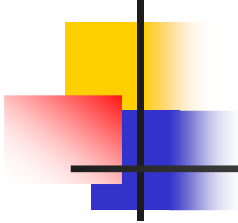
Deverbativní adjektiva (I)

- “progresivní” adjektiva (-ící) mohou vyjadřovat různé časy (v příslušné slovesné parafrázi):
 - *Řekl, že (namísto současné trosky) předá synovi prosperující podnik. (podnik, který bude prosperovat)*
 - *Slíbil jí, že její prosperující podnik neprodá. (podnik, který prosperuje/prosperoval)*
- plná synonymie se slovesným vyjádřením není zřejmě vždy ani možná:
 - *Předá synovi prosperující podnik odpovídá Předá synovi podnik, který bude prosperovat pouze v jedné interpretaci druhé věty.*



Deverbativní adjektiva (II)

- “rezultativní” adjektiva (-tý, -ný) mohou též vyjadřovat minulost i budoucnost:
 - *Před loupeží slíbil, že ukradené peníze budou rozděleny rovným dílem.* (peníze, které budou ukradeny)
 - *Policie prohlásila, že ukradené zlato bude navraceno bance.* (zlato, které bylo ukradeno)
- V datech PCEDT je i samotná identifikace “rezultativních” deverbativ obtížná, protože tyto nejsou zachyceny morfologickou analýzou.
- Spolehlivá pravidlová analýza obou typů deverbativních adjektiv je zřejmě nemožná.



Vyvozování z časových výrazů (I)

- tři komponenty
 - **parser** pro relevantní podstromy v tektogramatickém stromě
 - **interpretační modul** vytvářející funkční kompozice na základě výsledku parsingu
 - **porovnávající modul** porovnávající časovou extenzi funkčních kompozic
- interpretační modul je spojen s parserem, jednotlivé části kompozice vznikají při úspěšné analýze jednotlivých částí podstromů
- parsuje se podle speciální stromové gramatiky
 - obsahuje pravidla přepisující neterminál (zastupující podstrom) na terminální řídící uzel podstromu a jeho děti (terminály či neterminály)

Vyvozování z časových výrazů (II)

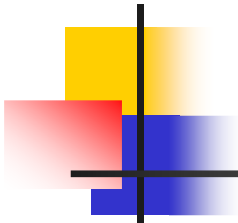
S: $\begin{array}{c} \text{Část} \\ \swarrow \quad \searrow \\ (\text{Předložka}) \text{VÝRAZ} \end{array} \mid \text{VÝRAZ} \left(\begin{array}{c} \text{začátku} \\ \swarrow \quad \searrow \\ \text{na} \quad \text{VÝRAZ} \end{array} \right)$

VÝRAZ: $\begin{array}{c} \text{Jmenovatel} \\ \swarrow \quad | \quad \searrow \\ (\text{Předložka}) (\text{Čítatel}) \text{JÁDRO} \end{array} \mid \text{JÁDRO} \left(\begin{array}{c} \text{třetiny} \\ \swarrow \quad \searrow \\ \text{druhé} \quad \text{JÁDRO} \end{array} \right)$

JÁDRO: DATUM | ABS_SPECIFIKACE | INDEXICKÝ_VÝRAZ | TYP_ENTITA

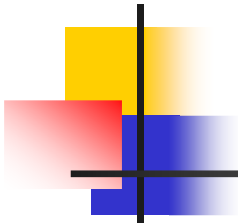
TYP_ENTITA:

$\begin{array}{c} \text{Entita} \\ \swarrow \quad \swarrow \quad \swarrow \quad \swarrow \quad \searrow \\ (\text{Přívlastek}) \quad (\text{Kvantifikátor}) \quad (\text{Ordinál}) \quad ([\text{SPECIALIZACE}]) \quad (\text{Počet}) \end{array}$



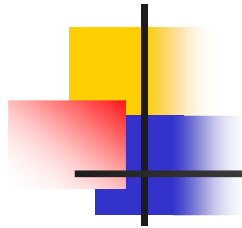
Vyvozování z časových výrazů (III)

- interpretační funkce jsou asociovány s jednotlivými pravidly, na základě parsingu a interpretace dětí je sestavena interpretace příslušného podstromu
- jednotlivá pravidla mohou obsahovat různá omezení
 - logické podmínky na souvýskyt volitelných dětí
 - povrchový slovosled
 - hodnoty gramatémů či funktorů
- ne vždy lze interpretovat spodní část podstromu bez znalosti parsingu horní části $\Rightarrow$ je možné předávat parametry pro parsing neterminálu reprezentující dítě



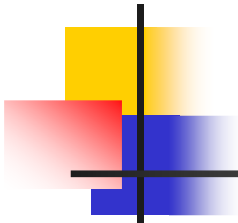
Vyvozování z časových výrazů (IV)

- porovnávající modul porovná časovou extenzi funkčních kompozic
 - maximálně jedna proměnná v nejvnitřnějších funkcích
- prochází kompozicí zevnitř ven a udržuje si tzv. paprsek:
 - dolní a horní odhad nejlevějšího a nejpravějšího bodu (obecně) množiny intervalů, která je výsledkem procházené funkce
 - dolní a horní odhad velikosti intervalu, který je výsledkem procházené funkce
 - absolutní hodnota výše uvedených položek, je-li známa
 - informaci o tom, zda nejlevější a nejpravější bod je na hranici nějaké entity (den, měsíc, apod.)
 - informaci o tom, zda výsledek procházené funkce je nějaká speciální hodnota (pondělí, duben, apod.)



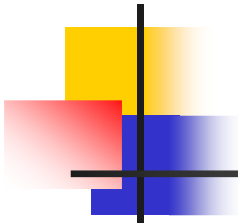
Vyvozování z časových výrazů (V)

- pro projití obou porovnávaných kompozic se porovnají výsledné paprsky
- je tedy možné porovnávat výrazy bez znalosti hodnoty bodu promluvy, pro každý typ funkce se upraví odhady v paprsku vstupních argumentů
- je-li však tato hodnota známa (či jde-li o absolutní výraz), je možné počítat přesně
 - používáme perlový CPAN modul DateTime (Dave Rolsky)



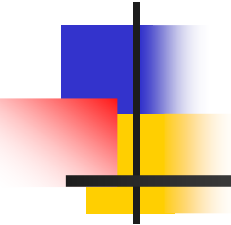
Vyvozování z časových výrazů (VI)

- v našem korpusu lze testovat pouze relace vyvozené z výrazů v jedné větě či v sousedních větách
- takových výrazů je v korpusu velmi málo
- výsledky bohužel slabé (téměř 100% přesnost, ale velmi malý nárůst pokrytí)
- bez znalosti absolutní hodnoty bodu promluvy navíc téměř nejde naším systémem srovnávat
 - *příští měsíc x koncem tohoto roku*
 - *zítra x v lednu příštího roku*



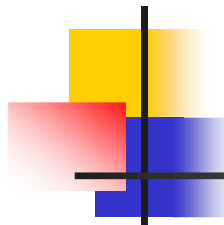
Další rozšíření

- přítomnost některých adverbii (*potom, předtím, apod.*) lze použít pro další inference
- vedlejší časové věty určují nejen vztah vůči svému referenčnímu bodu podle PA, ale též vůči řídící větě
 - využíváme vedlejší věty s TSIN, TTILL, částečně TWHEN
- faktory jako TFRWH či TOWH by též mohly být využity pro detekci substantiv nesoucích časovou informaci, ale v PCEDT se téměř nevyskytují



Část VI

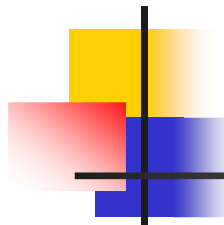
Vyhodnocení, srovnání a závěr



Vyhodnocení (DTest)

	P	R	F	Pw	Rw	Fw
Baseline (PA)	69.51	37.45	48.67	78.36	42.22	54.87
Všechna pravidla	75.29	41.64	53.62	84.70	46.84	60.32

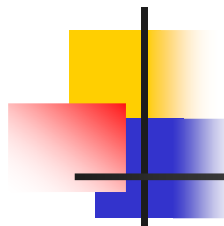
	dP	dR	dF	dPw	dRw	dFw
Historický přítom	-3.19	-0.99	-1.63	-3.79	-1.22	-1.98
Společná kompl. kotva	-0.33	-1.63	-1.45	-0.35	-1.82	-1.61
Kondicionály	-2.46	-0.51	-1.05	-2.56	-0.45	-1.03
Přímá řeč	-1.59	-0.13	-0.52	-1.57	-0.02	-0.42
Relace s infinitivy	0.61	-0.47	-0.24	0.57	-0.59	-0.35
Negace kompl. dějů	-0.16	-0.21	-0.21	-0.13	-0.21	-0.21
Časové výrazy	-0.07	-0.22	-0.20	-0.04	-0.22	-0.19
Obsahové věty	-0.49	-0.08	-0.19	-0.39	0.00	-0.10
Časové vedlejší věty	-0.12	-0.08	-0.10	0.00	-0.02	-0.02



Vyhodnocení (ETest)

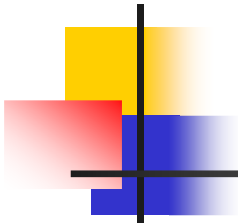
	P	R	F	Pw	Rw	Fw
Baseline (PA)	70.77	44.17	54.39	79.24	49.46	60.90
Všechna pravidla	72.32	48.09	57.77	81.74	54.35	65.29

POUZE SLOVESA	P	R	F	Pw	Rw	Fw
Baseline (PA)	70.77	61.32	65.71	79.24	68.66	73.57
Všechna pravidla	72.32	66.75	69.42	81.74	75.44	78.46



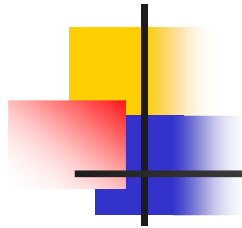
Chyby

- iterativní děje (zdaleka nejvíce)
- popsané výjimečné případy
- neopravený historický přítom (např. kvůli koreferenci)
- chyby v tektogramatické anotaci
- chyby či neúplnosti v anotaci temporality



Srovnání s jinými přístupy

- několik prací pro různé jazyky
 - [Mani a Wilson, 2000], [Schilder and Habel, 2005], [Li et al., 2000], [Boguraev and Ando, 2005] a další
 - výsledky nelze přímo srovnávat
- pravidlové i strojově-učící přístupy
- ukazuje se, že strojové učení může značně vylepšit pravidlový systém (Mani et al., 2005)
 - intuitivně patrné i v našem případě (kondicionály, deverbativa)



Závěr

- PA se naprostou většinou zasazuje o celkový výkon systému, popsané korekce a rozšíření mají celkově malý vliv (cca 4% nárůst míry F)
- z obecně lingvistického hlediska však práce umožnila ověřit (do jisté míry) některé hypotézy týkající se temporality v češtině a doplnit tak existující teorii