



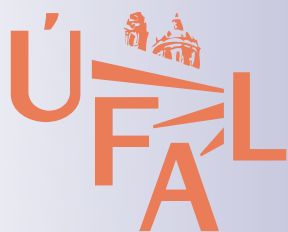
ÚSTAV FORMÁLNÍ
A APLIKOVANÉ
LINGVISTIKY



ÚSTAV FORMÁLNÍ
A APLIKOVANÉ
LINGVISTIKY

II. OBČASNÉ PREZENTAČNÍ FÓRUM

12. ŘÍJNA 2006



ÚSTAV FORMÁLNÍ
A APLIKOVANÉ
LINGVISTIKY



ÚSTAV FORMÁLNÍ
A APLIKOVANÉ
LINGVISTIKY

Eduard Bejček

- **Diplomová práce** (*do května 2006*)
 - WSD substantiv a adjektiv v PDT, strojové učení (C4.5)
 - slovník významů: 1) český WordNet, 2) PDT-VALLEX + 3) porovnání
 - a-rovina + ČWN, substantiva: baseline 90,0 % → 91,4 % ... 13,7 % zlepšení
 - t-rovina + PDT-VALLEX, subst.: basel. 84,9 % → 86,5 %
- **Opravy anotací PDT pomocí ČWN** (*říjen 2005 – říjen 2006*)
 - na začátku
 - 91 837 úspěšně anotovaných tokenů (jen 61,7 % mezianotátorská shoda)
 - 5111 cele anotovaných vět (každé slovo, které bylo lze anotovat)
 - postup
 - roztřídění neshod, data pro korektora, makra do ViMu, úpravy ČWN
 - pouze 25 nejchybovějších lemmat (= 0,5 % lemmat; získáme 7,4 % nových slov)
 - zanesení oprav zpět do dat, namapování starých synsetů na nové, kontroly konsistence dat, překlopení PDT 1.0 v CSTS do PDT 2.0 v PML („s-soubory“)
 - teď (v brzké době)
 - 102 789 úspěšně anotovaných tokenů (11 % nárůst)
 - 6941 cele anotovaných vět → přihlášeno do SemEval
- **Nástroj pro anotaci víceslovných lexémů** (*od července 2006*)
 - obecné pojmenované entity a víceslovné lexémy ze slovníku
 - úplný log předchozí anotace → persistentní undo/redo

Václava Benešová

- **Projekty:**

Centrum základního výzkumu

Informační společnost

The Valency Lexicon of Czech Verbs with Complex Syntactic-Semantic Annotation

- příprava další verze VALLEXu
- anotace slovesných tříd + jejich další rozpracování

- **Výzkumná činnost:**

- valence sloves
- alternace
- propracování tříd sloves communication tradičně označovaných jako verba dicendi a cogitandi

- **Veřejně prospěšné práce:**

- nástěnky, ufalí stránky

Ondřej Bojar



- Disertace!!!
 - Automatické generování VALLEXových hesel z korpusu a VALLEXu.
- Strojový překlad
 - Završit GAUK (budování slovníku)
 - Završit CzEng (paralelní korpus se Zdeňkem)
 - Završit workshop JHU (zpráva, publikace)
 - Euromatrix
 - Stromečkový dekodér + jeho nasazení na překlad Aj<->Čj
- Výuka
 - jen několik průměrně líných studentů na ročníkové projekty abakalářky
 - (jeden možná užije tektogramatiku v dialogovém systému)
 - v létě jsem vedl seminář na FJFI
- Budoucnost
 - pokud vše vyjde od poloviny listopadu v Melbourne (S. Bird; 8 měsíců až rok)
 - duchem stále zde (Euromatrix)

Silvie Cinková

- Pracovní náplň – projekty ap.
 - TR anotace PTB-WSJ (GAUK 489/2004, GAČR 405/06/0589)
 - PropBank -> EngValLex
- Výzkumná činnost
 - tektogramatický popis angličtiny
 - valence, určenost, gerundia a další jevy specifické pro angličtinu
 - gramatikalizace
 - dizertační práce (zejména) o gramatických a diskurzových funkcích nejčastějších švédských významových sloves a o tom, jak tato zrádná slovesa co nejlépe popsat ve švédsko-českém slovníku pro české studenty švédštiny
 - pravidlový bezslovníkový lematizátor pro švédský korpus PAROLE

Jan Hajič

- Pracovní náplň – projekty
 - Vedení
 - ÚFAL, VZ (část ÚFALu), CKL, IS/GAAV (sémantický web)
 - Kontakt (cest, fondy): Univ. of Colorado, PIRE, s Plzní Malach
 - GAČR na „PIRE“ téma (S. Cinková), 1x GAUK
 - Mez. proj.:
 - EU STREP Euromatrix, EU IP Comapnions;
 - USA: PIRE, Malach
 - Koordinace mezi projekty
- Výuka
 - PFL067/PFL068
 - Doktorandi (2006: ukončili 4)
 - Bude (snad) SW projekt (s IBM) (Enron ☺)
- Veřejně prospěšné práce
 - Organizace TLT06, částečně ACL07
 - Prog. výbory (2006: ACL, TLT, EAMTA, NAACL)
 - Recenze/panely: GAČR, MŠMT, NSF

Eva Hajičová

- Pracovní náplň – projekty
 - Výzkumný záměr
 - GAUK (Kučová, Zikánová)
 - GAČR (ÚJČ AV, Štícha; Panevová)
 - Nová mluvnice češtiny, Karolinum
- Výzkumná činnost
 - Oblast: AČV, závislostní syntax, diskurs
 - PDT: anotace a především využití anotovaných dat pro výzkum v dané oblasti

Eva Hajičová – pokrač.

- **Výuka:**
 - Letní semestr: Lingv. aspekty AI
 - Zimní semestr: Nově směry v lingvistice, seminář
 - Vedení PhD (FF, MFF: absolvovali Peterek, Frič)
 - „cotutelle“ – na FF Šmilauer (spolu s Pognanem)
 - Vedení diplomových prací na FF (anglistika)
- **Veřejně prospěšné práce:**
 - VMC
 - ACL 2007

Jiří Havelka

- Výzkum
 - Vlastnosti závislostních stromů (navazující na práci na anotaci aktuálního členění v PDT
 - Typy neprojektivních hran a závislostní analýza různých jazyků
 - Projektivizace závislostních stromů a přiřazování hloubkového slovosledu a TFA
 - Kombinatorické otázky související se závislostními stromy
 - Algoritmy pro nalezení maximální kostry a nalezení k nejlepších koster (pro neorientované i orientované grafy)
- Výuka
 - Komputační sémantika
- Veřejně prospěšné práce
 - Tajemník pro výuku za ÚFAL – servis pro učitele, ne pro studenty!
 - Správce databáze Amoeba za ÚFAL – zkontrolujte si údaje o své osobě na <http://www.mff.cuni.cz/fakulta/struktura/ufal.htm> !

Jarka Hlaváčová

Aktualizace morfologického slovníku češtiny

Morfologický servis

interní
pro ÚČNK, ÚTKL

Opravy, nová slova, kontroly

ted' konkrétně vyhledávání a opravy nekonzistentností typu

lemma + tag ----> jediná forma

Příprava nové morfologie

a/ nový program pomocí konečných automatů - studentský projekt

b/ nový systém kategorií a hodnot popisujících české slovní formy

Jarka Hlaváčová – pokrač.

- Projekty
 - Informační společnosti
 - REST (**R**esources and **T**ools for Information Systems)
 - JAZZ (Integrace **j**azykových **z**drojů za účelem extrakce informací z přirozených textů)
 - Externě ÚJČ
 - Evropský projekt LT4eL (**L**anguage **T**echnology **for** **e**Learning)
- Výuka - Perl pro lingvist(k)y

Petr Homola

- Výzkumná činnost (disertace)
 - strojový překlad
 - systém Česílko (polština, litevština, lužická srbština)
 - příbuznost jazyků na úrovni syntaxe
 - IS (1ET100300517)
- Výuka
 - C/C++, OOP, ročníkové projekty
 - bakalářská práce (čínský slovník)
- Veřejně prospěšné práce
 - rozhraní pro správu publikací

Václav Klimeš

- Pracovní náplň – projekty ap.
- grant AV ČR „Informační společnost“
 - Zeman+Semecký+Klimeš, spoluřešitel se ZČU
 - za MFF dočasné vedení místo Dana
 - překlad do znakové řeči a znakované češtiny, my: analýza až po t-rovinu
- Výzkumná činnost
- automatická tektogramatická anotace
 - rozšiřování dizertační práce o další atributy a zkusmé nasazení na angličtinu
 - pro účely IS (výše) a pro preanotaci dat
- Výuka
- bakalářské práce (nelingvistické)
- Veřejně prospěšné práce
- správce Collinsova parseru
- pravděpodobně se opět zapojím do PBML

David Klusáček

PRACOVNÍ VÝPLŇ:

- Dokončení PhD práce “*Generování formálního důkazu na základě jeho popisu v přirozeném jazyce*”.

VÝZKUM:

- Využití “deep-inference” kalkulu SKSq v neúplně řízeném proof-checkeru.
- Statistické metody pro generování hypotéz o tvaru důkazu na základě jeho neúplného textu. Korpus: knihovna systému MIZAR.
- MMI metoda pro jazkykové modelování a její zobecnění pro odhalování tříd ve strukturovaných datech (formulích).

TO LET:

- Program pro odstraňování duplicitních souborů.
- Systém bash scriptů pro automatické odevzdávání a verifikaci domácích úkolů z C/C++.

Oldřich Krůza

- Pracovní náplň
 - Zajištění přístupu ke korpusům přes Bonito
 - Maintaining systému Manatee
 - Konverze korpusů
 - Účast na JAZZu
 - Co je potřeba
- Výzkumná činnost
 - Rozpoznávání pojmenovaných entit (v rámci soutěže Semeval)
 - Statistický anotátor
 - Nástroje nutné pro chod soutěže (Evaluator, ...)
 - Kontrola konzistence dat
 - Segmentace českých vět
 - V rámci studia, nikoliv zaměstnání
 - Dělení vět na klauze
 - Syntaktická analýza

Vláďa Kuboň

- Pracovní náplň – projekty ap.
 - zodpovídám za evropský projekt (STREP) v rámci 6.rámcového programu Language Technology for e-Learning (LT4eL)
 - pracuji na tzv.“Šimově grantu“ v programu Informační společnosti...
 - podílím se na přípravě mezinárodního magisterského projektu LCT v rámci programu Erasmus Mundus
 - podílím se na vedení grantu Collegium Informaticum
 - formálně vedu GAUK Ondřeje Bojara
- Výzkumná činnost
 - segmentační analýza
 - rozklad souvětí na lingvisticky motivované úseky a následná snaha odhadnout vzáj. vztahy klauzí
 - Překlad mezi příbuznými jazyky (Česílko)
 - Strukturní míra podobnosti mezi přír.jazyky (článek na ACL workshopu s P.Homolou)
- Výuka
 - Přednáška „Úvod do poč. lingvistiky ...“ – letos 90 zapsaných studentů; dále 2 semináře; v LS cvičení Automaty a gramatiky
 - Vedu 2 softwarové projekty (FotoAS a MOAP)
 - Koordinuji výměnu Erasmus se Saarbrueckenem a Tuebingenem
 - Vedu 3 diplomky, 2 bakalářky, 4 postgraduandy
- Spolupráce s firmami
 - udržuji kontakt s firmou Skřivánek
 - testuji grammar-checker pro Microsoft

Lucie Kučová

- Rozšíření anotace koreference
 - Substantiva
 - Identická anafora
 - Bridging anafora
 - Adverbia
- GAUK – Faktory koherence textu
 - Propojení s TFA
 - Technická zpráva
 - Ukončení projektu

Markéta Lopatková

- VALLEX
 - příprava release VALLEX 2.0
 - teorie, koordinace anotací, anotace
 - alternační model slovníku
 - GAČR 2004-2006
- IS – Metody inteligentních systémů a jejich aplikace při dobývání znalostí a zpracování přirozeného jazyka (koordinace za UFAL)
 - redukční analýza, modelování FGD pomocí restartovacích automatů (s M. Plátkem)
 - segmentace
- Výuka
 - Vybrané problémy z lingvistiky
 - Matematické metody v lingvistice
 - cvičení Matematická analýza (pro informatiky)
 - LCT ~ Language and Communication Technologies
 - integrovaný magisterský program
 - Erasmus Socrates (studentská výměna)
 - koordinátor – Bolzano, Amsterdam, Utrecht
 - 4 doktorandi

Marie Mikulová

- Projekty
 - Centrum počítačnické lingvistiky
 - GAUK (JPa): Pražský závislostní korpus
 - GAUK (Hajič): Hlubková analýza mluvené řeči
- Pracovní náplň
 - Manuály pro tektogramatickou anotaci
 - velké, malé, české, anglické
 - Školení nových anotátorů pro tektogramatickou anotaci
 - Anotace české části WSJ
 - organizace, postup, pravidla (subfunktory)
 - Nová syntax češtiny
- Výuka
 - Odborné vyjadřování a styl (s Magdou Ševčíkovou)
- Veřejně prospěšné práce



Jiří Mírovský

- **Učení založené na paměti**
 - Aplikace na tagování češtiny a na parsování do analytické roviny (už ne)
- **Netgraph**
 - Vývoj vyhledávacího nástroje (Java, C)
 - Školení v- a pomoc při práci s Netgraphem
 - Správa databází stromů přístupných Netgraphem na serveru quest.ms.mff.cuni.cz
- **(Další) pracovní-vědecké zájmy**
 - tagování, tektogramatická rovina, valence, TFA, thajština
- **Veřejně prospěšné práce**
 - Nákup a půjčování software pro ÚFAL (kromě MS Windows a MS Office)
 - Nákup a správa dat vydávaných LDC; správa dat v /f/data
 - Pomocná správa databáze Amoeba (možná už ne)
 - Půjčování diaprojektoru

Petra Möllerová

- **Projekty**
 - MALACH (Multilingual Access to Large Spoken Archives)
 - Testování anotačního software
 - Pilotní studie (pro české podmínky)
 - Anotace výpovědí
 - WordNet
 - Lexikálně-sémantické anotace PDT
 - Úprava slovníku
 - Výsledky prezentovány na TSD a uveřejněny v LNAI 4188
- **Veřejně prospěšné práce**
(které mne živí)
 - knihovna

Petr Němec

- Projekt strojového překladu (z češtiny do angličtiny)
 - ve spolupráci s K.Hallem a CLSP (JHU, Baltimore)
 - generativní komponenta dokončena
 - bude vylepšena výsledky automatické časové a koreferenční analýzy
- Výzkumná činnost
 - automatická analýza časových vztahů (pro češtinu)
 - téměř uzavřeno, zbývá dokončit vyvozování na základě časových výrazů
 - automatická analýza nezájmenné koreference
 - náplň tohoto roku
- Výuka
 - dvoje cvičení z Objektově orientovaného programování (PRG032)

Václav Novák

- Informační společnost
 - návrh 4. roviny – sémantické sítě
 - příprava anotace textů z WSJ do této roviny
- Výzkumná činnost
 - Porovnání sémantických formalismů (DRT, TIL, MultiNet, KRL, OWL, MRS)
 - Upravení MultiNetu pro manuální anotaci
 - Vývoj anotačního nástroje pro MultiNet
 - Experimentování s přidáváním rysů do MST parseru
- Výuka
 - Úvod do UNIXu (SWI095)
 - V minulém akadem. roce též Internet (SWI096)
- Veřejně prospěšné práce
 - Správa McDonaldova analytického parseru (MST verze)

Petr Pajas

- Pracovní náplň – projekty ap.
 - Projekt informační společnosti – Jazz (“Integrace jazykových zdrojů za účelem extrakce informací z přirozených textů“, pod vedením Jarky Hlaváčové)
 - vývoj PML, jednotného datového formátu pro jazykové zdroje
 - vývoj a podpora programových nástrojů (TrEd, ntred, XML::LibXML, XSH, konverzní a validační nástroje,...)
- Výzkumná činnost
 - Převážně nelingvistické: teorie modelů, logika, teorie množin, nestandardní metody
- Výuka
 - Logika a teorie množin (pro učitelské obory)
 - Nestandardní seminář
- Veřejně prospěšné práce
 - Správa dat z PDT 1.0 a PDT 2.0 + podpora
 - Správa repozitáře Perlových modulů v /net/work/projects/perl_repo/
 - čištění kávovaru :-)

Jarmila Panevová

- vlastní náplň práce
 - úpravy a dotahování tektogramatické roviny (TGTS v PDT 2,5 i v teorii)
 - připravit publikaci pro nakl. Karolinum: Syntax (nové poznatky v syntaxi na základě vývoje PDT a dat v něm uložených – ve spoluautorství s mnoha dalšími autory z ÚFALu)
 - vztah přívlastkových vedlejších klauzí a participiálních adjektiv (pro GACR doc. Štíchy „Vybrané kapitoly z české gramatiky“)
- „veřejně prospěšná“ činnost
 - zástupce ředitele ÚFAL
 - předsedkyně Rady doktorského studijního oboru (RDSO) I3
 - předsedkyně rigorózní komise oboru I3
 - členka státnicové zkušební komise
 - školitelka 3 doktorandů
(1 z toho má přerušené studium)
 - členka 3 oborových komisí PGDS na FF UK
(matematická lingvistika, obecná lingvistika, současný český jazyk)
 - členka podoborové komise (POK 405) Lingvistika, literární věda GAČR
- výuka
 - Gramatická cvičení pro doktorandy
 - Úvod do obecné lingvistiky

Pavel Pecina

- disertace: **Lexikální asociační míry**
 - *extrakce kolokací, sémanticky podobných slov; word alignment*
- projekty:
 - **Malach:** *testovací kolekce pro IR, CLEF evaluace*
 - “**Český gigaword**”: *český webový korpus (nyní: čištění textů)*
- výuka: **PFL 067/068, NLP Reading Group**
 - *asistent doc. Hajiče při výuce statistických metod v NLP*
 - *týdenní setkání a čtení aktuálních článků*
- vedení: **bakalářské práce**
 - *hledání téměř identických dokumentů ve velkých kolekcích*
 - *sumarizace textu*
 - *automatický word alignment*
 - *vizualizace podobnosti dokumentů*
 - *aplikace pro ruční word alignment*
 - *MST parsing*

Nino Peterek

- Pracovní náplň
 - projekt Dialogy.Org, webové rozhraní audio-vizuálního korpusu,
 - tvorba korpusu Dialog ve spolupráci s AV ČR, ÚJČ.
- Výzkumná činnost
 - modelování prosodie
 - zkoumání prosodie aktuálního členění (Prof. Eva Hajičová),
 - modely důrazů, promluvových úseků, kadencí.
 - “rich transcription”
 - rozpoznávání replik, mluvčích, interpunkce, přechodů, neřečových událostí, scén.
- Výuka
 - PFL038 Základy rozpoznávání mluvené řeči, seminář.
- Veřejně prospěšné práce
 - správa Matlab serveru,
 - správa audio-vizuálních dat korpusu Dialog.

Jan Ptáček

- Pracovní náplň – projekty ap.
 - **TectoMT**
- Výzkumná činnost
 - **Syntéza textu**
 - **SOAP interface k morfologii**
- Výuka
 - Cvičím **PRG030** Programování I
- Veřejně prospěšné práce
 - pokus o přívětivé interní wiki stránky: <http://ufal.mff.cuni.cz/~ptacek/doku/>
 - **ÚFAL kalendář** (html / iCal)

Jan Ptáček - pokrač.

Syntéza vět

Spuštění:

/f/common/exec/generate_sentence <soubor s TGTS> [číslo
stromu]

Informace

<http://ufal.mff.cuni.cz/~ptacek/>

Kiril Ribarov

- Pracovní náplň – projekty apod.
(celková prac. náplň 75%, během pobytů v zahraničí 35%)
 - Informační společnost (M. Plátek et al.)
 - Český akademický korpus (vedoucí: B. Hladká)
 - DigiBGCult (Transfer of Knowledge project, Marie Curie Actions, Eu Commision)
 - Istituto di Linguistica Computazionale, Národní výzkumní centrum, Itálie
- Výzkumná činnost
 - analytický parsing
 - Otázky parsingu češtiny, obecně
 - Transformace závislostí v Českém akademickém korpusu na anal. rov. PDT
 - povrchový parsing Italské (chunking, funkcionální analýza)
 - staré texty
 - anotace staroslověnských textů: nástroje, metody, metodologie
- Výuka
 - společně s B. Hladkou, Úvod do strojového učení pro poč. lingv.

Jiří Semecký

Prague English Dependency Treebank

- Wallstreet Journal anotovaný na tektogramatické rovině
- Vytvoření valenčního slovníku anglických sloves (EngValLex)
- Tvorba a údržba anotačních nástrojů pro slovník a korpus
- Koordinace anotací

Disambiguace slovesných významů

- VALEVAL (podle slovníku VALLEX)
- Prague Dependency Treebank (podle slovníku PDT-VALLEX)
- Porovnání více typů příznaků popisujících výskyt slovesa a různých klasifikačních metod

Sazba tektogramatického manuálu

- Technické zprávy – český a anglický manuál
- Krátká verze manuálu

Vedení bakalářských projektů a bakalářských prací

Pavel Schlesinger

- Pracovní náplň
 - ... v rámci centra LC 536
 - agenda vedoucí k přípravě PBML
 - rozdělování čísel pro technické zprávy
 - statistický servis
 - kompletace slidů ☺
- Výzkumná činnost
 - klasifikační modely pro **koreferenci** (anaphora resolution, Semeval)
 - pokračování v rozvoji algoritmu pro Fisherovu diskriminační analýzu
 - společně s Jurjenem Duitnerem Tebbensem (AV)
 - hledání vhodných aplikací na datech z Ufalu
- Výuka
 - cvičení: Úvod do strojového učení pro poč. lingv.
 - seminář: Problémy aplikované lingvistiky (KPMS)

Otakar Smrž

Prague Arabic Dependency Treebank

- vedení projektu – publikace – správa dat
 - Information Structure with PADT (Univ. of Maryland)
 - Tips and Tricks of PADT (British Computer Society)
- morfologie, analytická syntax, TG a TFA (et alia)

Functional Arabic Morphology

Formal System and Implementation

- Functional Morphology, MorphoTrees
- Encode Arabic
- Haskell a otázky funkcionálního programování

Johanka

- Pracovní náplň – projekty ap.
 - Vývoj a údržba kombinovaných metod morfologického značkování
 - Značkování velkých dat (např. SYN*) inkriminovanými metodami
 - Diplomatický a datově výměnný interface mezi ÚFALEM, ÚTKLEM a ÚČNK
- Neváhejte se ozvat, pokud:
 - Chcete dostat do ruky aktuálně označovaný SYN200, SYN2005, SYN2006pub
 - Chcete nechat ojet vlastní data něčím progresivnějším, než je náčelníkův tagger ;-)
- Výzkumná činnost
 - Hybridní tagování uzavřeno po dosažení magické hranice (96%), nic dalšího zatím nerozjždím
- Výuka
 - momentálně žádná (dobíhající RP)
- Veřejně prospěšné práce
 - rodina je základ státu

Miroslav Spousta

- Pracovní náplň – projekty ap.
 - evropský projekt LT4eL (language technology for e-learning)
 - zajištění technické části
 - Shánění dat, anotace, konverze, programování
 - Jednoduché anotační nástroje (AJAX)
- Výzkumná činnost
 - Získávání a využití dat z webu
 - slovinský web korpus
 - čeština: chystá se velká hromada textů (.cz od L. Galamboše z KSI)
 - Problémy: extrakce textů z PDF, čištění dokumentů, ...
- Výuka
 - v LS: seminář TCP/IP (2x0/2)
 - RP a BP (1)
- Veřejně prospěšné práce
 - neprovozují

Pavel Straňák

- Lexikálně semantická anotace PDT
 - opravy starých anotací podle CWN
 - Petra Möllerová, Eduard Bejček
 - TSD 2006
 - Nové notace lexikálních významů:
 - konstrukce slovníku lex. významů
 - vlastnosti, formát, podpora v anotačním nástroji
 - anotační nástroj (vyvíjí E. Bejček)
 - s-layer PDT PML pro reprezentaci lexikálně semantických anotací
- GAUK: Možnosti extrakce morfosémantických tříd ve více jazycích
 - provedeny některé experimenty na čínštině
- koordinace našeho úkolu pro SemEval (Senseval 4)
- Veřejně prospěšné práce
 - Příprava PBML do tisku a na web (přebírá Pavel Schlesinger)
 - příležitostně další typografické práce (sborník z TLT 2006 aj.)
 - publicity for ACL 2007

Magda Ševčíková

- projekty
 - 1ET101120503: Integrace jazykových zdrojů za účelem extrakce informací z přirozených textů
 - GA-UK 352/2005: Pražský závislostní korpus: Analýza vybraných jevů z české funkční onomatologie a syntaxe
 - GD201/05/H014: Collegium Informaticum
- gramatémy
 - LREC post-conference workshop, Janov
 - SLE, Brémy
 - TLT, Praha
- pojmenované entity
 - úpravy v anotaci pojmenovaných entit v souvislosti s poskytnutím dat do SemEval
 - rozšíření anotace na číselné výrazy
 - technická zpráva o pojmenovaných entitách

Jana Šindlerová

- Pracovní náplň – projekty ap.
 - anotace PEDT
 - zvelebování EngValLexu
- Výzkumná činnost
 - srovnání české a anglické valence
 - teoretické otázky anglické syntaxe z pohledu závislostního
- Výuka
 - nic
- Veřejně prospěšné práce
 - (snad brzké) dokončení DP

Jan Štěpánek

- Projekt Informační společnost – Integrace jazykových zdrojů za účelem extrakce informací z přirozených textů
 - Datový formát PML
 - Převod různých korpusů do PML
 - Budování dalších zdrojů v PML – slovníky apod.
- Anotace českých překladů WSJ
 - Anotační nástroje, správa dat
- Výzkumná činnost
 - Udržování konzistence dat
 - XML
- Výuka
 - Komputační sémantika - s J. Havelkou
 - Pražský závislostní korpus
- Veřejně prospěšné práce
 - Databáze cest

Zdeňka Urešová

- Projekty

1. GAUK: Automatická hloubková analýza mluvené češtiny: od akustického signálu k významu
2. GAČR: Tectogrammatical Language Description for Czech-English Machine Translation and Speech Recognition
3. VZ

- Práce na projektech

- Valenční slovník sloves pro WSJ
- Školení anotátorů AR, průběžné konzultace – JÚLŠ, Bratislava
- Nová syntax češtiny

- Veřejně prospěšné práce:

- publikace ÚFAL (s P. Homolou)

- Referáty:

- SLS, Bloomington (s M. Mikulovou)
- University of Colorado (s J. Hajičem)



Barbora Vidová Hladká

- **Projekty**

- REST (IS)
 - Český akademický korpus 1.0 – cd-rom, průvodce (podrobnosti viz pondělní seminář 13. 11.)
 - Na cestě k ČAK 2.0 (česko-slovenská spolupráce)
- Cvičebnice češtiny STYX
 - Ondřej Kučera obhájil diplomovou práci
 - Soutěže (SVOČ[☺], ACM Student Research Competition[☺], Bolzanova cena?)
 - GAUK 2006 – zamítnut
 - Příprava zájmového kroužku pro ZŠ/SŠ

- **Výzkumná činnost**

- Diskurz – aktivovanost objektů
 - Anotace dalších dat - „Dvořák“
 - Publikace v recenzovaném sborníku workshopu

- **Výuka**

- Organizace tutoriálu o PDT v rámci zimního VMC (granty pro účastníky od CEI)

Jan Votrubec

- Pracovní náplň
 - editor ČAK 1.0
- Výzkumná činnost
 - Morče
 - experimenty na různých datech =>
 - přepisování cizího kódu na vlastní (vstupní parser) =>
 - cíl: PML, jiné jazyky, obecný trénovací algoritmus
 - Prozodie
 - snaha o „recyklaci“ Morčete
 - připravované experimenty s FÚ FF UK a ÚRE AV ČR
- Výuka
 - ročníkové projekty
 - cvičení Internet
 - Volný seminář pro Fonetický ústav FF UK (v jednání)

Zdeněk Žabokrtský

- **Valence sloves**
 - valenční slovník českých sloves VALLEX (2003 v.1.0, 2006 v.2.0)
 - alternační model valence (disertace 2005)
- **Treebanking**
 - PDT 2.0 (koreference, gramatémy, opravy)
 - převody složkových stromů na závislostní (PTB, Negra/Tiger)
 - závislostní parsing (čeština, slovinština, němčina, rumunština, arabština)
 - pojmenované entity
- **Strojový překlad přes tektogramatickou rovinu**
 - vývoj prototypu od podzimu 2005
- Výuka: PFL070/076 – Zdroje lingvistických dat I/II