Název projektu :	*Centrum komputační lingvistiky*
Řešitel:	Prof. PhDr. Eva Hajičová, DrSc.
Příjemce dotace:	Univerzita Karlova v Praze, Ovocný trh 3/5, Praha 1, 110 00, IČO: 00216208

Průběžná zpráva o realizaci projektu[1] v roce 2002

1. Stručný přehled dílčích cílů projektu splněných v uplynulém období

Uvádíme přehled dílčích cílů, které byly stanoveny v šesti bodech (A) až (F) v původním návrhu programu a které byly splněny v r. 2002. Cíle jsou konkretizovány v rámci tří nosných výzkumných projektů, které vykrystalizovaly během dosavadní existence Centra, a to rozvíjení Pražského závislostního korpusu (bod B původního návrhu), projekt strojového překladu (bod F původního návrhu) a v rámci výzkumu v oblasti zpracování mluvené řeči pak participace na mimořádně rozsáhlém mezinárodním projektu MALACH (bod E původního návrhu). Souběžně s těmito projekty a v návaznosti na ně pokračoval výzkum v oblasti teoretických aspektů komputační lingvistiky, tedy jejích matematických i lingvistických základů (body A, C a D) a rovněž vyvíjení některých aplikačních systémů (bod F původního návrhu). Nově byla loňského roku zařazena další významná organizační aktivita Centra, totiž příprava 17. Světového kongresu lingvistů.

V následujícím přehledu jsou jednotlivé body upřesněného programu pro rok 2002 označeny (T1) až (T6), popř. dalším členěním podle dílčích cílů plánovaných na rok 2002 (podle upřesnění uvedeného v průběžné zprávě o realizaci projektu v roce 2001).

T1: Rozvíjení Pražského závislostního korpusu (bod B původního návrhu)

Pražský závislostní korpus (Prague Dependency TreeBank, PDT) byl stěžejním projektem CKL i v roce 2002. PDT byl rozšiřován na tektogramatické rovině o anotaci struktury, aktuálního členění a koreference. Do anotace byl z důvodů konzistence integrován valenční slovník, vedle toho byl připravován komplexně anotovaný valenční slovník.

Na tektogramatické rovině bylo anotováno 20 tisíc vět (tj. velký soubor), asi 1000 větám bylo přiřazeno též TFA značkování. Dodávání hodnot koreference je v počátcích. Přibližně 300 vět bylo anotováno kompletně (tj. i s koreferencí a gramatémy).

Dále probíhala manuální anotace lexikálně-sémantická (rozlišení polysémie).

Bylo rozvíjeno softwarové vybavení pro anotace, následnou kontrolu a zpracování dat a dokumentace obecně.

Dílčí cíle Centra řešené v r. 2002 se týkaly následujících úkolů:

[1] Zpráva podepsaná řešitelem, která byla schválena oponentním řízením, se současně se zápisem o oponentním řízení, (pokud bylo pořádáno), vyúčtováním za uplynulé období, upřesněním dílčích cílů a rozpočtu pro následující období zasílá v jednom vyhotovení zadavateli (závěrečná zpráva se zasílá ve dvou vyhotoveních).

Název projektu :	*Centrum komputační lingvistiky*
Řešitel:	Prof. PhDr. Eva Hajičová, DrSc.
Příjemce dotace:	Univerzita Karlova v Praze, Ovocný trh 3/5, Praha 1, 110 00, IČO: 00216208

T1-1 Při specifikaci valenčních rámců neslovesných slovních druhů byl kladen důraz především na valenční rámce deverbativních substantiv, s cílem v budoucnosti vybudovat valenční slovník substantiv. Byla tedy nejen hlouběji propracována teorie valence substantiv (zejména z hlediska redukce valenčních doplnění daného substantiva vzhledem k počtu valenčních doplnění základového slovesa a z hlediska kombinatoriky možností povrchových vyjádření daných valenčních doplnění), ale byly učiněny i první kroky k algoritmizaci převodu valenčních rámců sloves na valenční rámce příslušných substantiv.

T1-2 Verifikace stanovených kritérií pro přiřazování valenčních rámců deverbativním substantivům na základě anotací PDT probíhala především při rozhodování, zda přiřadit doplněním daného substantiva slovesné funktory, nebo funktory substantivní (u deverbativních substantiv bylo třeba v konkrétních textech z PDT odlišit dějové užití těchto substantiv (*naše.ACT výplaty mezd.*PAT *zaměstnancům.*ADDR) od jejich užití substantivního (*to je moje.*APP *první.*RSTR *výplata*), dále bylo třeba posoudit dějovost názvů artefaktů (*Jágrův.*ACT *gól*) a rozhodnout o přiřazení slovesných, příp. substantivních funktorů u substantiv s již slabou spojitostí se základovým slovesem (*příklad / případ bezohledného chování*.PAT)).

T1-3 Byla ověřena a následně zpřesněna kritéria pro přiřazování hodnoty pro kontrastivní topic (C) v atributu pro aktuální členění věty (TFA).

T1-4 Podle pokynů pro anotátory se implementoval částečný přechod z analytické na tektogramatickou úroveň tam, kde pokyny byly algoritmizovatelné. Při tomto částečném přechodu se vyplňují hodnoty atributů, doplňují se uzly na povrchu elidované a ruší se uzly synsémantických slov a interpunkce. Během implementace se pokyny zpětně zpřesnily. V dalším kroku přechodu se implementovalo využití vznikajícího valenčního slovníku k doplnění uzlů elidovaných na povrchu věty a k vyplnění některých atributů jejich i dalších uzlů. Obě tyto implementace již byly použity (Workshop CLSP, Johns Hopkins University, Baltimore, 2002, viz další aktivity Centra).

T1-5 Pokračovalo budování komplexního valenčního slovníku sloves. Byla prohlubována koncepce slovníku, který zachycuje všechny relevantní syntaktické a syntakticko-sémantické informace. Zpřesňovala se kritéria pro určování jednotlivých valenčních doplnění (např. rozlišování funktorů adresát a benefaktor), zachycovaly se další jevy – zejména reciprocita (např. ve větě *Jan a Marie se milovali.* zaplňuje koordinace *Jan a Marie* aktor i patient slovesa *milovat*) a kontrola (vyznačení "controlleru" pro infinitivní doplnění sloves kontroly, např. ACT pro sloveso *začít*: *Jan*$_i$ (ACT) *začal pracovat* (Sb$_i$)). Dále se zpracovávaly vidové dvojice a sjednocovala se jejich anotace. Tímto způsobem

bylo do konce roku zpracováno přes 1000 nejčastěji užívaných sloves (k tomuto tématu byla předložena Technická zpráva).

Pracovní verze valenčních rámců anotátorů (zpracování jednotlivých významů sloves) zatím zůstává oddělena od komplexně zpracovávaného slovníku, ve kterém jsou pro každé sloveso zachycovány všechny jeho významy včetně všech relevantních vlastností.

T1-6,7 Na konverzi formátu dat Ústavu pro jazyk český, vytvořeného v 70. letech, se pracovalo do března 2002, a dále od listopadu 2002. V mezidobí (březen až říjen 2002) bylo nutno provést co nejrychleji morfologickou anotaci části korpusu (přes 150.000 slov), který se v současné době anotuje na tektogramatické rovině. Tuto anotaci bylo nutno přidělit témuž zkušenému pracovníku, který pracuje na konverzi, a dokončení práce na konverzi korpusu ÚJČ se tedy posunulo do 1. pololetí příštího roku (2003).

T2: Strojový překlad (bod F původního návrhu)

Projekt strojového překladu byl fakticky zahájen v r. 2002. Byly vybudovány základy systému strojového překladu, a to na následujících zásadách:

1. Systém je založen na "klasickém" modelu ANALÝZA – TRANSFER – SYNTÉZA, přičemž jednotkou překladu je rovněž tradičně jedna věta. Pro srovnávací účely byl vytvořen systém překladu založený na "přímé metodě" (ale na základě částečné analýzy až do úrovně morfologické) pomocí dekodéru GIZA++ a ISI dekodérů.
2. Reprezentace struktury a významu věty na úrovni transferu je založena na teoreticky propracované tektogramatické (podkladové) rovině jazykového popisu, z níž se vychází i při anotování Pražského závislostního korpusu.
3. Projekt v první fázi počítá s překladem mezi angličtinou a češtinou, a to
   (a) obousměrně v případě použití statistického překladu pomocí přímého systému GIZA++ a ISI dekodéru;
   (b) ve směru čeština → angličtina pro systém založený na strukturním překladu.

Projekt rovněž těží z předchozích projektů v oblasti strojového překladu řešených nebo spoluřešených v ÚFALu a jeho předchůdcích (zejména ze systému anglicko-českého překladu APAČ a česko-ruského překladu RUSLAN), a snaží se využívat nástrojů a lingvistických dat vytvořených v rámci bývalé Laboratoře pro zpracování jazykových dat ÚFAL a CKL. Zejména pak velmi těsně navazuje na projekt Pražského závislostního korpusu a využívá tak ve velké míře jeho náročně vytvořená data, neboť projekt strojového překladu v sobě zahrnuje téměř všechny problémy zpracování přirozeného jazyka, od morfologie přes syntax a sémantiku až po generování (syntézu).

Na projektu jsme pracovali ve spolupráci s Center for Language and Speech Processing na Johh Hopkins University v Baltimore (USA), kde se někteří pracovníci CKL zúčastnili letního workshopu – viz zpráva o dalších aktivitách CKL. Některé komponenty systému již byly dokončeny, a systém je připraven pro provedení prvních srovnávacích testů podle nové standardní metodologie vyvinuté v IBM Research.

T3: Zpracování mluvené řeči (bod E původního návrhu)

T3-1 Hlavní aktivity byly zaměřeny zejména do oblasti rozpoznávání souvislé mluvené řeči v úlohách s rozsáhlými slovníky. Proběhly experimenty s jazykovými modely, které kombinují slovní *n*-gramy s *n*-gramy založenými na lemmatech a slovních druzích (POS tag). Vzhledem k tomu, že experimenty byly provedeny pomocí AT&T dekodéru, byl navržen způsob, jak reprezentovat kombinaci těchto modelů ve formátu konečných automatů, který zmíněný dekodér využívá. Ukázalo se, že z implementačního hlediska je výhodnější použít násobení pravděpodobností jednotlivých modelů místo standardní lineární interpolace.

Všechny experimenty byly provedeny technikou „lattice rescoring" – konečný automat s nejpravděpodobnějšími posloupnostmi slov (lattice) generovaný dekodérem byl reskórován pomocí různých bigramových a trigramových modelů založených na lemmatech a slovních druzích. Skóre pocházející ze slovního jazykového modelu použitého při prvním průchodu dekodérem bylo v lattice zachováno. Přehled výsledků je uveden v následující tabulce:

Typ jazykového modelu	Přesnost rozpoznávání
Slovní bigram	69.93%
Slovní trigram	70.68%
Bigram s lemmaty	70.93%
Trigram s lemmaty	71.53%
Bigram s POS tagy	71.61%
Trigram s POS tagy	72.25%

Z výše uvedené tabulky plyne, že největší zlepšení přineslo reskórování trigramovým modelem založeným na lemmatech, kdy se výsledky rozpoznávání zlepšily o 2.32% oproti slovnímu bigramovému modelu. Uveďme, že popisované experimenty se týkaly rozpoznávání čtené řeči se slovníkem 60 tisíc slov.

Pokračovaly také práce s jazykovými modely založenými na morfémech (kmenech a koncovkách). Podařilo se nám vyvinout techniku, která umožňuje vytvářet konečné

automaty pracující s těmito modely korektním způsobem – to znamená, že na výstupu rozpoznávače se nemohou objevit například dvě koncovky za sebou či kombinace kmen-koncovka, které tvoří slovo v češtině neexistující. Navržené metody se v současné době testují a výsledky budou publikovány v následujícím roce v připravované disertační práci.
V souladu se schváleným plánem na rok 2002 probíhaly i velmi intenzivní práce spojené s řešením dílčích úloh na projektu MALACH (Multilingual Access to Large Spoken Archives). Tento vysoce prestižní projekt byl přijat na období 5 let a jeho cílem je vývoj systémů pro automatický předpis svědeckých výpovědí lidí, kteří přežili holocaust. Svědecké výpovědi byly pořízeny ve více než 30 různých jazycích a česká strana je spoluzodpovědná za zpracování jazyků střední a východní Evropy. Na projektu participují Visual History Foundation v Hollywoodu, Johns Hopkins University v Baltimore, University of Maryland, IBM, MFF UK v Praze a ZČU v Plzni. V roce 2002 byly zpracovávány svědecké výpovědi vedené v českém jazyku. Anotační práce na zpracování těchto svědeckých výpovědí jsou podporovány National Science Foundation (USA), Project #0122466 (www.clsp.jhu.edu/research/malach), práce na konstrukci systému rozpoznávání řeči byly pak částečně prováděny na půdě CKL. Mimořádná náročnost řešení této úlohy vyplývá z toho, že výpovědi jsou spontánní a vysoce zatíženy emocionálním stavem řečníka. Průměrný věk lidí, kteří poskytli svá svědectví, byl 75 let. Problémy s rozpoznáváním spontánní promluvy ve svědeckých výpovědích vedených v českém jazyku byl dále násoben

- velkým množstvím nespisovných slov - nespisovná slova nejsou obsažena v dostupných zdrojích používaných pro jazykové modelování (těmito zdroji jsou obvykle vhodné texty vybírané z novin a knih, tedy obecně z psaného a nikoliv mluveného textu)
- odlišnou stavbou vět v hovorovém a psaném jazyce - to opět způsobuje problémy při jazykovém modelování, tentokrát z toho důvodu, že jazykový model natrénovaný na psaném textu obsahuje jiné rozložení pravděpodobností n-gramů.

Výše uvedené charakteristiky znesnadňují rozpoznávání spontánní češtiny obecně, v projektu MALACH se k těmto faktorům přidává ještě velké množství osobních a geografických jmen a cizích slov. Hlavním úkolem při konstrukci jazykového modelu bylo proto najít taková data, která by co nejvíce respektovala vlastnosti spontánní češtiny. Pro akustické modelování byly využity promluvy 336 řečníků. Anotované promluvy měly rozsah celkem 84 hodin. Ke konstrukci jazykových modelů byly využity ruční přepisy výpovědí určené pro akustické trénování ve spojení se speciálním výběrem vět z Lidových novin. Rozpoznávaný slovník měl až 80 tisíc slov. Podrobnější rozbor této problematiky spolu s

výsledky experimentů s rozpoznáváním spontánně vyslovených výpovědí je uveden v [3]. Na přelomu listopadu a prosince 2002 byla zkušební verze tohoto systému instalována pracovníky CKL na JHU v Baltimore k dalším rozsáhlým testům.

T3-2 Pro usnadnění studia prozodie a aktuálního členění jsme vyvinuli a naprogramovali nový modul anotačního nástroje Transcriber tak, aby na dlouhých promluvách zobrazoval kromě přepisu i průběhy prozodických parametrů. Vlastní modely bude možné natrénovat až pomocí minimálně šesti hodin zkontrolovaných trénovacích dat.

T3-3 V rámci tvorby prozodické databáze jsme pokračovali v nahrávání a přepisu MapTask dialogů. Dosud jsme tak získali čtyři hodiny prozodicky anotovaných dialogů. Začali jsme také nahrávat prozodicky zajímavé čtené věty, které nám umožňují srovnání se spontánními promluvami i detailnější studium vybraných prozodických jevů.

T4: Teoretické aspekty komputační lingvistiky, její matematické i lingvistické základy (body A, C a D původního návrhu):

Jak uvádíme v cílech programu CKL i v předchozí výroční zprávě, teoretický výzkum Centra je neoddělitelně spjat s výše zmíněnými projekty, a slouží jednak jako předpoklad pro jejich formulaci a teoretický základ jejich řešení, tak i k ověření teoretických hypotéz a jejich zpřesňování, a přináší nové důležité podněty pro další teoretické bádání.

V roce 2002 jsme se soustředili především na následující body:

T4-1 Podněty vzniklé v průběhu tektogramatického anotování PDT (srov. zde , bod T 1) byly vyhodnocovány z hlediska teorie valence vypracované v rámci funkčního generativního popisu. Ukazuje se, že pro valenční rámce adjektiv a adverbií je možné vystačit se zavedeným souborem funktorů, substantiva mají sice své specifické funktory, ale sdílejí i soubor funktorů typických pro slovesa. Použití „dialogového" testu užívaného pro zjištění sémantické obligatornosti doplnění naznačuje, že bude třeba pracovat s jemnějším tříděním „hloubkových" protějšků povrchově nepřítomných valenčních členů: Zatím byla vedle tektogramatické jednotky Gen (general) zavedena jednotka nová, Unspc (unspecified) pro valenční člen omezenější co do rozsahu, než je jeho hodnota Gen (jeho všeobecnost), ale bez možnosti jasně vymezené lexikální specifikace (např. *V rozhovoru*.Unspc.ACT Unspc.ADDR Gen.PAT *pro rádio ECHO Milan Uhde uvedl, že…,* není zcela jasné, kdo byl původce o kdo adresát rozhovoru, víme jen, že byl veden mezi Uhdem a zástupci rádia ECHO, patientem byla patrně všechna témata (Gen), o kterých v rozhovoru byla řeč).

Přistoupilo se k vytváření skupin sémanticky příbuzných sloves, jejichž valenční rámce jsou konfrontovány a upravovány na základě materiálu z PDT. Do slovníkového hesla se kromě

valence budou postupně zapojovat i údaje o schopnostech jednotlivých lexií vstupovat do různých diatezí s příslušnými důsledky pro jinou hierarchizaci aktantů v odlišných diatezích (typy pasíva, reciproční konstrukce, popř. další diateze jako rezultativní posesívní *mám uvařeno, dostal jsem zaplaceno*).

T4-2 Na materiálu z Pražského závislostního korpusu byla zvláštní pozornost věnována tzv. fokalizátorům, mezi něž nově zařazujeme i operátor negace. Ukazuje se, že je výhodné zkoumat sémantické vlastnosti těchto prvků (jako jsou např. v češtině částice *jenom, také, pouze, i, …*) v souvislosti se členěním věty na základ (topic) a ohnisko (fokus). V primárním případě stojí fokalizátor na hranici mezi těmito dvěma částmi věty a jeho dosah zasahuje celé ohnisko, v sekundárních případech může být fokalizátor součástí základu; je třeba ještě dalším zkoumáním ověřit, zda v takovém případě dosah fokalizátoru končí na hranici mezi základem a ohniskem. Podobně zůstává otevřenou otázkou, co je v dosahu fokalizátoru, je-li fokalizátor jediným prvkem ohniska.

T4-3 Významného posunu jsme dosáhli v empirickém zkoumání a teoretickém popisu tzv. kontrastivního základu (topiku). Vypracovali jsme operativní test pro identifikaci kontrastního základu (na základě parafrází daných vět s užitím dlouhého tvaru osobních zájmen) a v několika studiích (viz seznam publikací Centra) jsme se pokusili stanovit jistou škálu kontrastu od nejméně výrazného k nejvýraznějšímu a ukázat, jakými kontexty jsou tyto stupně podmíněny a jakými výrazovými prostředky jsou podepřeny.

T4-4 V průběhu anotování PDT na tektogramatické rovině se objevují konstrukce, v nichž se sice vyskytuje koreference elementů, ale není úplně jasné, zda jde o jevy gramatické, nebo textové koreference. Zpracovávají se teoreticky konstrukce s infinitivem jako *nechal se zabít, nechal děti spát, nechal si opravit boty*, ale i konstrukce jako *je vidět Sněžku/Sněžka*, *bylo/a vidět jeho rozpačitost, je znát rozpaky*, nebo také *je možné odjet k moři*, *lze odjet k moři* (kde členem kontrolujícím subjekt infinitivu, který není explicitně přítomen, není žádný z aktantů řídícího predikátu).

Na základě předchozích studií o typech koreference byl vypracován scénář pro přiřazování lexikálních hodnot k uzlům vypuštěným v povrchové podobě věty a zrekonstruovaným v její podobě hloubkové (tektogramatické), a také instrukce pro anotátory umožňující onsistentní přiřazování hodnot v atributech pro koreferenci.

T4-5 V oblasti možné logicko-pragmatické interpretace a anotace proběhla řada seminářů a diskusí na toto téma, a bylo pozváno několik zahraničních odborníků. V současné době se připravuje konkrétní seznam jazykových a kognitivních jevů, které v současných formalismech nejsou zachytitelné vůbec nebo jen nedostatečně. Takové podklady budou později sloužit k návrhu nového formalismu.

T4-6 V oblasti matematických metod soustavně probíhaly přípravné práce na nástrojích použitelných pro experimenty s metodami maximální entropie, zejména v oblasti selekce rysů (features). Tyto práce jsou specifikačně i programově velmi náročné a v souvislosti s plánovaným a probíhajícím rozšiřováním množiny rysů vyžadují testování na existujících korpusech, proto vlastní trénovací a vyhodnocovací experimenty budou probíhat až v roce 2003.

Pro účely zkvalitnění morfologicky anotovaných korpusů se dále vyvíjí metoda tzv. negativních trigramů, která má na základě manuálně připravených omezení (pravidel) odhalit místa, kde anotace je lingvisticky nepřípustná (s ohledem na kontext). Byla vyzkoušena na menším německém korpusu Negra, který je ve vývoji, a byla schopna odhalit řadu nedostatků. Tuto metodu lze rovněž využít pro předzpracování morfologicky analyzovaných dat před automatickou statistickou disambiguací, kde ovšem zatím dává jen minimální zlepšení výsledné chybovosti.

T5: Vyvíjení některých menších aplikačních systémů (bod F původního návrhu)

Vedle soustředěné práce na projektu strojového překladu uvedeného v bodu T2 výše pokračovaly práce na dalších menších aplikačních systémech:

T5-1 Výsledky projektu Česílko (překlad čeština-slovenština) byly aplikovány na anotovaný korpus PDT (Pražský závislostní korpus), a to po takových modifikacích, které umožnily zachovat stávající anotace na morfologické i syntaktické rovině ve velmi dobrém stavu. Tento projekt je příkladem možností přenosu anotace na další jazyky, jehož výsledkem jsou jednak nové znalosti o vztahu blízkých jazyků, jako je čeština a slovenština, tak i velké úspory při jinak velmi drahé manuální anotaci. V současné době se jedná o použití této metody na budovaný slovenský národní korpus, ve spolupráci s Pedagogickou fakultou UK Bratislava a Jazykovedným ústavem Ľ. Štúra.

T5-2 Úspěšně také proběhly nové experimenty s překladem do polštiny v rámci projektu Česílko a byla vyhodnocena jejich úspěšnost. V souladu s předpoklady se ukázalo, že úspěšnost je sice nižší než při překladu do slovenštiny, přesto je ale výsledek překladu použitelný jako podklad pro posteditaci. Byla také provedena první studie možností rozšíření systému Česílko o neslovanské jazyky, v této fázi jmenovitě o litevštinu.

T5-3 V rámci výzkumu lingvisticky podporovaného vyhledávání informací byly vypracovány studie o signifikantních kolokacích a o lexikální disambiguaci, obě s ohledem na specifika českých textů a obě s praktickými komputačními výsledky. Pracuje se na ruční anotaci trénovacího korpusu pro lexikální disambiguaci (cca 1000 vět). Byly provedeny úvodní experimenty týkající se automatického odhadu specificity a automatického shlukování

významových slov. Je hotova specifikace textového systému MATES2, který tyto a mnohé další možnosti lingvistické podpory vyhledávání integruje. Systém MATES2 bude funkční do poloviny příštího roku (viz i Technická zpráva).

T6: Organizační přípravy 17. Světového kongresu lingvistů

Plynule pokračovaly přípravy 17.Světového kongresu lingvistů, který se bude konat 24.-29. července 2003 v Praze a jehož je Centrum hlavním spolupořadatelem a prof. Hajičová předsedkyní organizačního výboru. Ve spolupráci s předsedou programového výboru prof. Kieferem byl stanoven rámcový program Kongresu, zajištěny prostory Kongresového centra v Praze pro jeho pořádání a připraven i společenský program. Na organizaci Kongresu se velmi aktivně a podstatnou měrou podílí řada pracovníků Centra a do přípravných prací jsou zapojeni i doktorandi a studenti.

Další aktivity CKL plánované na rok 2002

- CKL **pokračovalo** ve vydávání **technických zpráv** (ve spolupráci s ÚFALem MFF UK) o dílčích výsledcích výzkumu; v roce 2002 byly vydány čtyři výzkumné zprávy, které budou též k dispozici na webových stránkách CKL:
  - Holub, M., Pecina, P.: Sémanticky signifikantní kolokace. TR-2002-13
  - Hana, J., Hanová, H.: Manual for Morphological Annotation. TR-2002-14
  - Lopatková, M., Žabokrtský, Z., Skwarská, K., Benešová, V.: Tektogramaticky anotovaný valenční slovník českých sloves. TR-2002-15
  - Gramatovici, R., Plátek, M.: D-trivial Dependency Grammars with Global Word-Order Restrictions. TR-2002-16

- V rámci **Dne otevřených dveří** na MFF UK, který se konal 4.12.2002, byla připravena prezentace, na níž byla představena činnost CKL. Vzhledem k návštěvníkům, především zájemcům ze středních škol, byla mladými pracovníky Centra živou formou přiblížena hlavní témata naší práce i dosažené výsledky, a také možnosti, které pracoviště může poskytnout studentům a mladým výzkumným pracovníkům (připravili J. Cuřín, V. Honetschläger a Z. Žabokrtský).

- Byly uspořádány **přednáškové pobyty** předních zahraničních profesorů (prof. E. Bach a prof. C. Townsend, oba USA). V letním semestru proběhl dvoutýdenní intenzívní kurz z oblasti formální sémantiky (prof. B. Partee, University of Massachusetts, Amherst, USA).

V zimním semestru se Praze uskutečnil semestrální pobyt prof. M. Kuramitsu (Soka University, Japan). Dále CKL navštívili a přednesli přednášky prof. H. Schnelle (Bochum, Německo), doc. B. Byrne (John Hopkins University, Baltimore, USA) a doc. Doug Oard (University of Maryland, USA). V zimním semestru na MFF UK vedl dva celosemestrální kurzy host CKL Dr. Radu Gramatovici (University of Bucharest, Bukurešť, Rumunsko; kurz *Tree Adjoining Grammars* a kurz *Marcus contextual grammars*)

V rámci 17. Cyklu jarní školy (viz následující odstavec) byl uspořádán kurs z oblasti rozpoznávání mluvené řeči (prof. F. Jelinek, John Hopkins University, Baltimore, USA), který bude pokračovat v jarních měsících následujícího roku; otázkami mluvené řeči se bude zabývat i kurs prof. Julie Hirschbergové (Columbia University, New York, USA) v 18. Cyklu této jarní školy v březnu 2003.

- **17. CYKLUS JARNÍ ŠKOLY VILÉMA MATHESIA, 11. - 22. března 2002, Praha**

  V době od 11. do 22. března 2002 se uskutečnil 17. cyklus jarní školy Viléma Mathesia (http://ufal.mff.cuni.cz/vmc). Kromě českých vyučujících (viz níže) pozvání přijali odborníci z USA (Robert Frank, Frederick Jelinek, Martin Kay, Barbara Partee), Německa (Hans Kamp), Itálie (Nicoletta Calzolari, Paolo Ramat), Velké Británie (Greville Corbett, Geoffrey Leech), Ruska (Vladimir Borschev) a Maďarska (Ferenc Kiefer). Celkem se zúčastnilo 58 studentů a výzkumných pracovníků (včetně místních účastníků). Většině cizích účastníků z Bulharska, Číny, Chorvatska, Estonska, Francie, Gruzie, Německa, Maďarska, Kyrgyzstánu, Litvy, Makedonie, Polska, Rumunska, Ruska, Slovenska, Španělska, a Ukrajiny byl udělen grant na ubytování, stravu a kapesné ve výši 2000,- Kč. Platící účastníci přijeli z Chorvatska a Maďarska. Akce byla zajištěna z prostředků poskytovaných European Commission, Research DG, Human Potential Programme, High Level Scientific Conferences (contract number: HPCF-2000-00336), Open Society Institute – HESP a Matematicko-fyzikální fakultou Univerzity Karlovy.

  J. Hajič, E. Hajičová, J. Panevová a P. Sgall (všichni CKL) vedli kursy s následujícími názvy (po řadě): *From text corpus to structural annotation , Topic-Focus Articulation: Theoretical Framework and Corpus Annotation , Valency and Functional Generative Description, The Core and the Periphery of the Language System.*

  V rámci jarní školy se na půdě Pražského lingvistického kroužku konala každoroční speciální Jakobsonovská přednáška s názvem *Language Universals and Translation*, kterou přednesl Prof. Paolo Ramat z Pavijské university v Itálii.

Název projektu :	*Centrum komputační lingvistiky*
Řešitel:	Prof. PhDr. Eva Hajičová, DrSc.
Příjemce dotace:	Univerzita Karlova v Praze, Ovocný trh 3/5, Praha 1, 110 00, IČO: 00216208

Další aktivity CKL uspořádané v roce 2002

- CKL uspořádalo v roce 2002 dva **výjezdní semináře**. Hlavním cílem těchto seminářů bylo společné soustavné projednávání jednotlivých bodů z pracovní náplně Centra za účasti zástupců všech zúčastněných pracovišť.

  21.-26.1.2002, Rokytnice nad Jizerou

  Semináře se zúčastnila většina pracovníků CKL, včetně doktorandů; přítomni byli i studenti FF UK, kteří se podílejí na tektogramatickém značkování PDT.

  Seminář se věnoval zejména konkrétním problémům spojeným s anotováním PDT a upřesnění instrukcí pro anotátory:

  - Zavedení funktoru OPER pro zachycení matematických operací
  - Zpracování modálních konstrukcí s negací
  - Zachycení speciálního typu apozičního vztahu mezi větným členem a klauzí
  - Zacházení se zájmenem *ten*
  - Rozebírání podstromů v parentezi
  - Vymezení podmínek pro doplnění „prázdného slovesa“
  - Problematické (chybné) konstrukce s více řídícími uzly věty.

  3.10.-6.10.2002, Nové Hutě

  Druhý seminář byl koncipován jako fórum, na kterém mladí pracovníci centra prezentovali výsledky své práce. Semináře se zúčastnila většina pracovníků Centra, dále v CKL hostující profesor, bohemista M. Kuramitsu (Soka University, Japonsko), PhDr. M. Šimková (Jazykovedný ústav Ľ. Štúra SAV, Slovensko) a zejména studenti, kteří se začínají podílet na projektech Centra.

  Hlavní témata semináře:

  - seznámení s výsledky workshopu Natural Language Generation in Machine Translation (viz další bod, J. Hajič, M. Čmejrek);
  - projekt s pracovním názvem Prague Arabic Dependency Treebank (O. Smrž);
  - současný stav PDT, harmonogram jeho budování, možnost využití (J. Hajič);
  - jevy zachycované ve valenčního slovníku (M. Lopatková);
  - koncepce vyvíjených parserů (K. Ribarov, Z. Žabokrtský);
  - stav projektů souvisejících se zpracováním řeči (N. Peterek, P. Machek);
  - mezinárodní projekt MALACH

  Byly představeny nové verze nástrojů vyvinutých v Centru – TrEd (editor užívaný pro anotaci PDT, P. Pajas), NetGraph (prohlížeč pro vyhledávání v PDT přes internet, J. Mírovský) a XSH (nástroj pro práci s XML, P. Pajas).

Proběhly diskuse o dalších možnostech vylepšení informačních toků v rámci CKL a o dalším zpřístupněním výsledků práce Centra.

- V roce 2002 proběhl workshop projektu **Natural Language Generation in Machine Translation** v Center for Language and Speech Processing na Johns Hopkins University v Baltimore, MD, USA. Z pracovníků CKL se ho zúčastnili Jan Hajič (vedoucí týmu) a Martin Čmejrek, doktorand v interním studiu na ÚFAL. V týmu dále pracovali výzkumníci a profesoři z University of Michigan, Johns Hopkins University, University of Pennsylvannia, University of Toronto, a studenti z MIT a Stanford University. Na projektu pracovali také další pracovníci CKL, kteří jsou zároveň doktorandy ÚFAL, a to Jan Cuřín, Ivona Kučerová, Václav Honetschlaeger, Zdeněk Žabokrtský, Petr Pajas a nepřímo i řada dalších, především týmy anotátorů jazykových korpusů. Projekt se zabýval problémem generování povrchové podoby anglických vět jednak samostatně, jednak jako výsledku automatického překladu z češtiny. Výsledkem projektu bylo několik experimentů s generováním angličtiny, jejichž výsledky byly vyhodnoceny novým jazykově nezávislým evaluačním nástrojem pro hodnocení automaticky generované povrchové podoby textu pocházejícím z IBM Research. V rámci těchto experimentů byl vypracován zárodek software založeného na strojovém učení a pravděpodobnostních modelech, který by měl umožnit obecné lokální transformace stromových struktur (např. pro generování jazyka). O projektu bude též vydána technická zpráva na CLSP Johns Hopkins University.

- Na pracovišti ÚJČ se významně využívá PDT pro zjišťování frekvence základních slovosledných schémat (SVO a varianty) a pro zjišťování přiřazených syntaktických značek (atribut, předmět, doplněk) vzhledem k dotazům a odpovědím uloženým v poradenské databázi ÚJČ.

- Pracovníci CKL se podíleli na přípravě skript pro studenty MFF UK a FF UK – Úvod do teoretické a počítačové lingvistiky. První díl - **Teoretická lingvistika**, autoři Hajičová, Panevová, Sgall - vyšel v r. 2002, na přípravě druhého dílu **Počítačové zpracování přirozeného jazyka** (autoři Kuboň a kol.) se pracovalo.

- Barbora Hladká a Kiril Ribarov (oba CKL) připravili koncepci workshopu Adaptation of Automatic Learning Algorithms for Analytical and Flective Languages, který byl zařazen do programu ESSLLI 2003 (15th European Summer School in Logic Language and Information, konaná v Budapešti).

- V květnu 2002 (20.5.) proběhlo v sekci CKL na MFF UK natáčení části krátkého dokumentárního filmu, v němž byla představena činnost CKL.

Přínos zahraničních cest

Přehled zahraničních cest pracovníků CKL (níže) i oddělený soupis zahraničních cest mladých pracovníků a doktorandů (viz bod 6.) jasně ukazuje, že o výsledky výzkumu CKL je v zahraničí velký zájem: většina cest totiž pokrývá buď účast na mezinárodní konferenci s pozvaným nebo přijatým referátem, nebo aktivní účast na mezinárodním vysoce ceněném workshopu, nebo přednáškový pobyt na některé zahraniční univerzitě. Řada pobytů byla částečně hrazena zahraničním grantem nebo přímo zvoucí organizací, takže z prostředků Centra byl hrazen jen nutný doplněk výdajů. Se souhlasem poskytovatele grantu jsme rovněž do položky zahraniční cesty přesunuly část nákladů na pobyt zahraničních hostů ušetřených nikoli krácením těchto pobytů, ale díky získání prostředků z Fondu Mobility UK a z projektu vědecké spolupráce ČR – USA. To umožnilo téměř všem pracovníkům vystoupit na mezinárodním odborném fóru, mít srovnání s výzkumem na zahraničních pracovištích a získat tak cenné podněty pro další vědeckou práci v rámci Centra.
V průběhu roku jsme rovněž konzultovali možnosti zapojení Centra do 6. rámcového programu EU; ke spoluúčasti na tzv. "integrated project" nebo "network of excelence" centrum pozvali někteří perspektivní koordinátoři projektů (např. kolegové ze Sárské univerzity, z univerzity v Pise, v Szegedu a z Ústavu východních jazyků v Paříži). Pokračovali jsme v dosavadní mezinárodní spolupráci, jak jsme o ní podávali zprávy v minulých letech, a navázali jsme některé kontakty nové, viz o nich v bodu 4. níže. O dobrém mezinárodním zvuku výzkumné práce centra svědčí rovněž jmenování vedoucí Centra prof. Hajičové členkou vědecké rady Institutu komputační lingvistiky (jeden z ústavů italské Akademie věd) v Pise.

Přehled nesplněných úkolů CKL

Centrum splnilo všechny cíle projektu pro rok 2002, které byly stanoveny v zakládající smlouvě CKL a podrobněji specifikovány v přehledu dílčích cílů ve zprávě za rok 2001. Aktivity CKL, které nabyly plánovány, jsou uvedeny ve zvláštním odstavci bodu 1.

Publikační činnost pracovníků CKL

Seznam konkrétních výsledků a výstupů získaných v rámci realizace projektu jsou uvedeny v příloze **Publikace**.

Název projektu :	*Centrum komputační lingvistiky*
Řešitel:	Prof. PhDr. Eva Hajičová, DrSc.
Příjemce dotace:	Univerzita Karlova v Praze, Ovocný trh 3/5, Praha 1, 110 00, IČO: 00216208

2. Personální a organizační zabezpečení činnosti centra

Na řešení projektu se podílí celkem 41 pracovníků, dohromady zastávají 29,45 plných úvazků (rok 2000: 28 prac.; r. 2001: 42 prac.). Z toho 29 pracovníků má úvazek rovný nebo vyšší než 0,7 (r. 2000: 18 prac.; r. 2001: 25 prac.) , dalších 12 pracovníků má úvazek nižší než 0,7 (r. 2000: 10 prac., r. 2001: 17 prac.). Složení pracovního týmu je z hlediska kvalifikace ve vztahu k náplni v Centru vyvážené. Pracovní tým CKL se skládá ze čtyř profesorů, deseti vědeckých pracovníků a 24 odborných pracovníků. Jeho činnost zajišťují tři techničtí pracovníci.
S CKL dlouhodobě spolupracuje řada studentů magisterského studia (15-20studentů, podle časových možností) a další odborní pracovníci (3).
Věkovým zastoupením svých pracovníků CKL splňuje podmínku zaměstnávat především mladé vědecké pracovníky – 25 pracovníků CKL je mladší než 35 let, dohromady zastávají 22,8 plných úvazků (tj. více než 75%), další objem práce odvádějí studenti magisterského studia (všichni do 35 let).

Kvalifikace, prac. zař.	počet	prům. věk	vážený věk	úvazek
Profesoři	4	64	67	2,3
vědečtí pracovníci	10	47	43	5,35
odborní pracovníci	24	27	27	19,5
technická podpora	3	38	35	2,3
Celkem	**41**			**29,45**

V personálním obsazení plzeňské sekce Centra došlo během roku 2002 k drobné změně. Jeden z dříve spolupracujících studentů byl po nástupu do doktorského studia přijat do CKL na částečný úvazek 0,5.
V současné době je tedy personální obsazení plzeňské sekce CKL následující: prof.ing. Josef Psutka, CSc. (0,3), ing. Pavel Ircing (0,8) a ing. Josef V. Psutka (0,5). Plzeňská sekce sídlí v prostorách katedry kybernetiky FAV, ZČU s tím, že může dle potřeby využívat veškeré technické a personální zabezpečení katedry a univerzity.

3. Spolupráce Centra

a) rozvoj odborné spolupráce v rámci ČR, s ostatními zakládajícími a spolupracujícími organizacemi ve sledovaném období

Centrum spolupracuje velmi těsně zejména se Západočeskou univerzitou v Plzni (zakládajícím členem Centra) v oblasti rozpoznávání mluvené řeči. Kontakty se stále udržují s pracovištěm v Brně (Laboratoř přirozeného jazyka FI MU Brno), a to zejména při konzultacích o tvaroslovném značkování a problematice korpusů a jejich zpřístupnění. Nadále spolupracujeme s Ústavem Českého národního korpusu, který slouží jako zdroj dat, a s Ústavem teoretické a komputační lingvistiky na propojení statisticky založených systémů a systémů pravidlových.

b) nová zapojení do mezinárodních struktur ve sledovaném období

- V r. 2002 jsme zahájili po delším období neformálních kontaktů velmi těsnou spolupráci s Akademií věd Slovenské republiky, konkrétně s Jazykovědným ústavem Ľudovíta Štúra, na projektu využití blízkosti češtiny a slovenštiny pro tvorbu anotovaných dat pro sloveštinu. Nadále probíhaly kontaktu v rámci společných projektů s pracovišti v USA, Kanadě, Německu, Slovinsku, Maďarsku, Rakousku, Francii, Itálii a severských zemích.
- Dvě sekce CKL, a to plzeňská i univerzitní pražská, se zapojily do mezinárodního projektu MALACH a spolupracují s Johns Hopkins University v Baltimore, University of Maryland, IBM (Human Language Technologies Group, Yorktown) a Visual History Foundation v Hollywoodu.
- Centrum též bylo vyzváno zahraničními pracovišti ke spoluúčasti na projektech 6. rámcového programu EU (Sárská univerzita, univerzity v Pise, v Szegedu a ústav východních jazyků v Paříži, viz bod 1. výše) a na konkrétní náplni návrhů se intenzivně pracuje.
- Nově se Centrum zapojilo do evropského projektu ENABLER koordinovaného Univerzitou v Kodani a v Pise (od května 2002), viz též bod 7.

c) rozvoj spolupráce s aplikační sférou a v rámci regionu ve sledovaném období

Nadále probíhá spolupráce s aplikační sférou v oblasti lematizace češtiny a jejím užitím v aplikacích pro vyhledávání dokumentů. Byly obnoveny kontakty i s velkými zahraničními softwarovými firmami a jednání a využití technologie dostupné na MFF UK.

d) způsob využívání výsledků a výstupů projektu aplikační sférou a v rámci regionu

V rámci spolupráce uvedené v bodě (c) se výsledků využívá v komerčních informačních systémech, např. v produktu ASPI. Řada dalších softwarových produktů využívá vyvinuté technologie prostřednictvím dříve udělených licencí.

Název projektu :	*Centrum komputační lingvistiky*
Řešitel:	Prof. PhDr. Eva Hajičová, DrSc.
Příjemce dotace:	Univerzita Karlova v Praze, Ovocný trh 3/5, Praha 1, 110 00, IČO: 00216208

4. *Podpora a výchova mladých výzkumných pracovníků*

Obě sekce CKL, jejichž mateřským pracovištěm je univerzita (MFF UK a ZČU), jsou významným způsobem zapojeny do doktorských programů. Podílejí se na výchově mladých výzkumných pracovníků, jimž jednak poskytují příležitost k vlastnímu bádání a k jeho prezentaci, jednak umožňují jejich zapojení do větších výzkumných úkolů. V roce 2002 se silně rozvinula i spolupráce s nadanými studenty magisterského studia (a to jak z matematicko fyzikální fakulty a z filozofické fakulty UK, tak z katedry kybernetiky ZČU), u nichž po dokončení magisterského studia připadá v úvahu doktorandské studium; za svou práci jsou odměňováni stipendiem.
Přístup k odborným časopisům a publikacím umožňuje všem výzkumným pracovníkům sledovat aktuální vývoj na předních zahraničních pracovištích oboru. Dobré přístrojové vybavení dovoluje efektivní práci na vlastních tématech.
CKL umožňuje v co nejširší míře svým pracovníkům prezentaci výsledků na mezinárodních konferencích a jejich konfrontaci s přístupy k podobným problémům ve světě. Podporuje účast mladých pracovníků na mezinárodních letních školách i jejich pracovní pobyty na zahraničních.

a) doktorské studijní programy

CKL se významným způsobem podílí na doktorském programu I3 – Informatika – Počítačová lingvistika na MFF UK a na doktorském programu v oboru Kybernetika na Fakultě aplikovaných věd, ZČU.
Pracovníci CKL z MFF vedli 12 přednášek a seminářů v letním semestru a 14 přednášek a seminářů v zimním semestru, které navštěvovali studenti i doktorandi příslušných fakult, stali se školiteli nebo konzultanty 6 nově přijatých doktorandů. Celkem tedy pracovníci CKL na MFF UK školí 31 doktorandů.
Prof. Psutka školí celkem 11 doktorandů, kteří studují obor Kybernetika akreditovaný na Katedře kybernetiky ZČU.
Další pracovník CKL, František Štícha (ÚJČ), učil v letním i zimním semestru na Institutu translatologie FF UK.

b) podíl mladých výzkumníků

Jak již bylo uvedeno v bodu 2., 25 pracovníků CKL je mladší než 35 let, navíc se na práci CKL podílí řada studentů magisterského studia. CKL umožňuje doktorandům a mladým pracovníkům prezentovat výsledky vědecké činnosti na mezinárodních konferencích, zapojuje je do společných projektů a podporuje jejich účast na dalších akcích CKL.

Název projektu :	*Centrum komputační lingvistiky*
Řešitel:	Prof. PhDr. Eva Hajičová, DrSc.
Příjemce dotace:	Univerzita Karlova v Praze, Ovocný trh 3/5, Praha 1, 110 00, IČO: 00216208

5. Způsoby zpřístupnění výsledků a výstupů centra veřejnosti

- Pracovníci CKL se zúčastnili řady mezinárodních konferencí a jiných odborných setkání, na kterých přednesli zvané přednášky (5 zvaných přednášek) a recenzované příspěvky o výsledcích dosažených v projektech Centra (13 referátů), případně prezentovali své výsledky na posterech (6 posterů). Seznam názvů referátů je uveden v bodu 1.d, zahraniční cesty a v bodu 6.a, účast studentů a doktorandů na mezinárodních konferencích.
- Publikace v domácích i zahraničních časopisech a ve sbornících mezinárodních konferencí zpřístupňují výsledky Centra široké odborné veřejnosti. Celkem vyšlo 67 publikací (seznam viz Příloha).
- CKL vydalo v roce 2002 čtyři technické zprávy, které podrobně dokumentují úroveň a metody zpracování jednotlivých dílčích cílů. Technické zprávy jsou k dispozici v tištěné formě na pracovišti na MFF UK a elektronicky na adrese http://ckl.mff.cuni.cz/ufal/?a=techrep.
- Byly nově strukturovány www stránky CKL, http://ckl.mff.cuni.cz/ s cílem podat mezinárodní i české odborné veřejnosti komplexní informaci o činnosti Centra. Na základě doporučení MŠMT zde lze nalézt kromě základních informací o hlavních výzkumných tématech a jejich metodologickém zázemí, o struktuře Centra a mezinárodní spolupráci také odkazy na www stránky jednotlivých projektů, dále např. seznam publikací členů CKL, technické zprávy, informace pro studenty a informace o aktuálních událostech.
- CKL jako člen mezinárodní sítě ENABLER Network je zainteresováno na vypracování přehledu dostupných jazykových zdrojů. Poskytlo též základní informace o Pražském závislostním korpusu, čímž zásadně přispělo k informovanosti odborné veřejnosti o svém stěžejním projektu.
- V rámci Dnu otevřených dveří na MFF UK byla představena činnost CKL (viz bod 1.)
  Na MFF UK proběhlo natáčení krátkého filmu přibližujícího výzkum a výsledky CKL (viz bod 1.)

Název projektu :	*Centrum komputační lingvistiky*
Řešitel:	Prof. PhDr. Eva Hajičová, DrSc.
Příjemce dotace:	Univerzita Karlova v Praze, Ovocný trh 3/5, Praha 1, 110 00, IČO: 00216208

Přehled a upřesnění dílčích cílů projektu a postupu při jejich naplňování pro následující období, tj. pro r. 2003[2]

Uvádíme zde upřesnění cílů pro r. 2003, jak byly stanoveny v šesti bodech (A) až (F) v původním návrhu programu. Cíle jsou konkretizovány v rámci tří nosných výzkumných projektů, které vykrystalizovaly v prvních 18 měsících existence Centra a jak byly též prezentovány ve výroční zprávě za rok 2002., a to rozvíjení Pražského závislostního korpusu (bod B původního návrhu), projekt strojového překladu (bod F původního návrhu) a v rámci výzkumu v oblasti zpracování mluvené řeči pak participace na mimořádně rozsáhlém mezinárodním projektu MALACH (bod E původního návrhu). Souběžně s těmito projekty a v návaznosti na ně bude pokračovat výzkum v oblasti teoretických aspektů komputační lingvistiky, tedy jejích matematických i lingvistických základů (body A, C a D) a rovněž vyvíjení některých aplikačních systémů (bod F původního návrhu).

V následujícím přehledu jsou jednotlivé body konkretizovaného programu pro rok 2003 označeny (T1) až (T6) , popř. dalším členěním, a to v souladu s časovým harmonogramem uvedeným pod přehledem.

T1: Rozvíjení Pražského závislostního korpusu

Pražský závislostní korpus (PZK) je stěžejním projektem CKL. Výzkum v roce 2003 bude plynule navazovat na výsledky let předcházejících, bude vytvářet další předpoklady pro jeho využití uživateli z nejrůznějších oblastí výzkumné i aplikační činnosti a soustředí se především na tyto body:

T1-1 Jazyková databáze: formulace a implementace nových algoritmů
Budou provedeny experimenty s novou definicí slovníku na úrovni 1 (morfologie) za použití technologie XML. Nově se bude zpracovávat verze algoritmu pro práci s tímto slovníkem.

T1-2 Zpřesňování podmínek pro automatickou specifikaci valenčních rámců neslovesných slovních druhů, formulace algoritmu.

T1-3 Formulace a implementace úplnějšího algoritmu pro anotaci AČV v PZK, především z hlediska zachycení kontrastivního topiku.

T1-4 Na základě analýzy struktury diskurzu v souborech PZK značkování koreferenčních vztahů v části tzv. velkého souboru a evaluace výsledků po zpracování ověřovacího podsouboru; na základě výsledků této evaluace zpřesnění instrukcí pro anotátory.

T1-5 Bude dokončena první fáze anotace tzv. velkého souboru (55 tis. vět).

[2] Uvádí se bližší specifikace cílů stanovených smlouvou a jejich rozpis na dílčích cíle pro daný kalendářní rok, vč. časového harmonogramu

T1-6 Bude zahájena kontrola anotace na první podúrovni (tzv. velký soubor). Budou formulovány a implementovány metody pro automatickou a poloautomatickou kontrolu anotovaných dat.

T2: Strojový překlad

V oblasti strojového překladu se jedná o první ucelený rok práce na systému.

T2-1 Statistický model pro transformace stromových struktur, vývoj (implementace) jednotlivých modulů, dokončení modulů započatých na WS02 v Baltimore.

T2-2 Generování v systému strojového překladu (angličtina). Vypracování systému klasifikátorů využívajících globálního kontextu, vypracování a implementace tzv. baseline systému pro porovnání.

T2-3 Tvorba a obohacování slovníků vhodných pro strojový překlad, angličtina-čeština a čeština angličtina.

T2-4 Vyhodnocování kompletního systému pro strojový překlad a vyhodnocování jednotlivých modulů podle metodiky IBM (BLEU).

T3: Zpracování mluvené řeči

T3-1 viz „Upřesnění dílčích cílů pro rok 2003 – ZČU“

T3-2 Pokračování ve studiu prozodických opozic mezi kontrastivním základem, průvodními členy základu a vlastním ohniskem na základě získaných dat a vizualizace prozodických parametrů.

T3-3 Rozšíříme prozodickou databázi o dialogy MapTask a čtené prozodicky charakteristické věty.

T4: Teoretické aspekty komputační lingvistiky, její matematické i lingvistické základy (body A, C a D původního návrhu):

Teoretický výzkum v rámci Centra je neoddělitelně spjat s výše zmíněnými projekty, a to jednak jako předpoklad pro jejich formulaci a teoretický základ pro jejich řešení, jednak tyto projekty přinášejí vedle ověřování platnosti navržených hypotéz i důležité další podněty pro teoretické bádání a pro obohacení daného pojmového rámce. V roce 2003 bude výzkum pokračovat v následujících bodech:

T4-1 Budou pokračovat práce na projektu kombinace strukturních a statistických metod pro české morfologické značkování. V jejich rámci budou dále zkoumány tzv. metody negativních n-gramů.

T4-2 Budou pokračovat práce na syntaktické analýze češtiny, a to metodami statistickými i pomocí pravidel. Budou studovány možnosti kombinace těchto metod, a výsledky budou vyhodnoceny a porovnávány pomocí standardní metodiky.

T4-3 Bude pokračovat práce na rozvíjení slovníku pro strojový překlad za pomoci automatického získávání terminologických korespondencí na základě paralelního korpusu a dalších dostupných

lexikálních zdrojů. V jejich rámci budou zkoumány možnosti automatické identifikace nominálních frází.

T4-4 Pokračování ve studiu hloubkové (tektogramatické) struktury věty na základě dat získaných anotací PZK, a to především těchto aspektů:

- podrobné studium valence polysémních sloves; získání podkladů pro sémantické třídění sloves
- studium gramatických podmínek pro povrchové vypouštění aktantu
- studium konkurence shodného a neshodného substantivního atributu
- v oblasti aktuálního členění věty pokračování ve studiu kontrastu, a to z hlediska možnosti či nutnosti zavedení rozlišení kontrastivní části ohniska věty od jiných členů ohniska
- v oblasti diskurzu prohloubení formulovaného algoritmu přiřazování stupňů aktivovanosti jednotlivým členům věty a jeho rozšíření na jiné než substantivní členy vět

T4-5 Pokračování ve studiu a specifikaci reprezentace těch sémantických (kognitivních) aspektů, které přesahují jazykový význam, pro případné doplnění anotačního scénáře PZK o další úroveň.

T5: Vyvíjení některých menších aplikačních systémů (bod F).

T5-1 Pokračování na rozšiřování kompletního systému automatického překladu čeština-slovenština. Zaměření na rozšíření překladového slovníku a na přenos anotace ze zdrojového jazyka.

T5-3 V rámci výzkumného projektu 'Sémantické modely textu' budou integrovány automatické procedury pro analýzu morfologie a syntaxe češtiny, detekci signifikantních kolokací, lexikální disambiguaci, konstrukci sémanticky orientovaného slovníku, segmentaci textu na tématicky koherentní pasáže, shlukování podobných dokumentů a indexaci dokumentů na základě abstraktních konceptů. Projekt je zaměřen na plně automatickou konstrukci modelů, které umožní testovat sémantický obsah dokumentu. Na projektu spolupracuje též šest studentů a doktorandů.

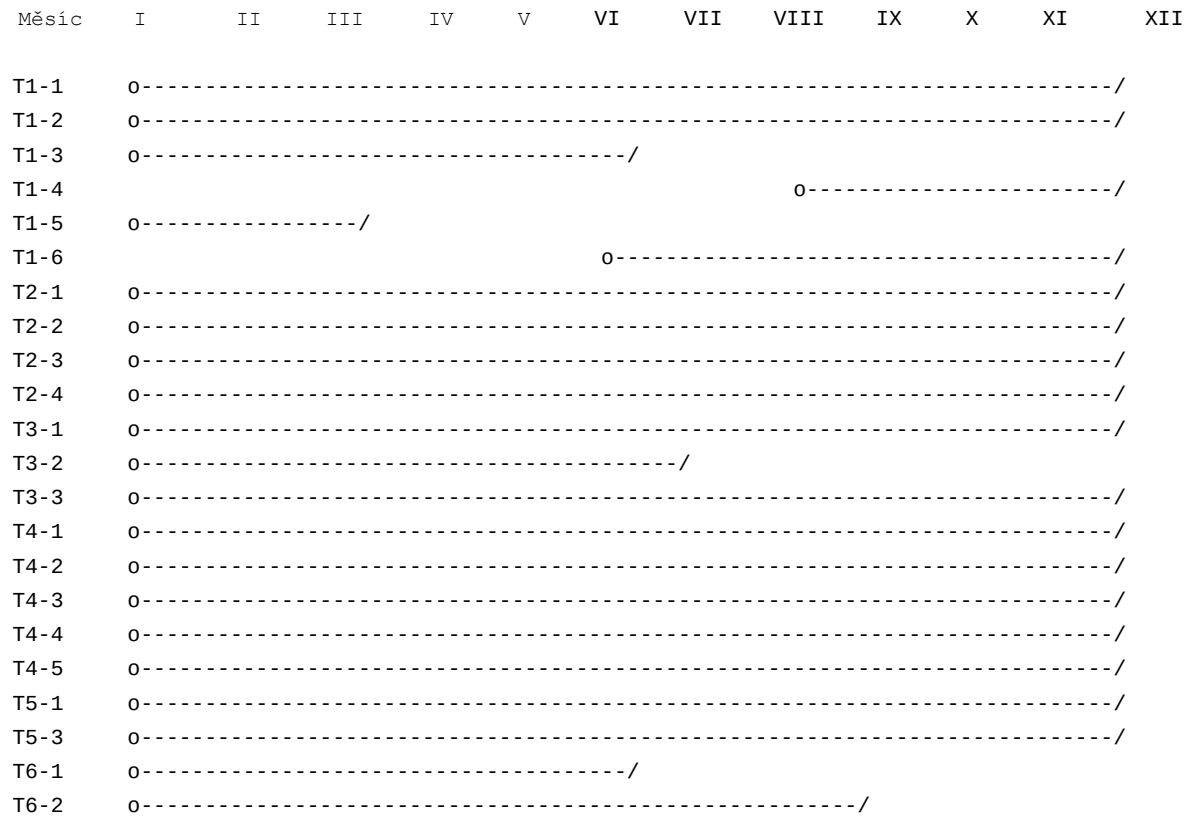
T6: Přípravy velkých mezinárodních akcí

T6-1 Dokončení organizační přípravy 18. běhu mezinárodních cyklů Centra Viléma Mathesia a zajištění jejich plynulého průběhu. Vyhodnocení cyklu a jeho vyúčtování.

T6-2 Pokračování v organizačních přípravách 17. Světového kongresu lingvistů, který se bude konat 24.-29. července v Praze a jehož je Centrum hlavním spolupořadatelem. Zajištění vlastního průběhu kongresu, a to jak po stránce organizační, tak i po stránce programové a společenské. Jde o velice prestižní světové setkání lingvistů (konané každých pět let, pokaždé v jiné zemi, v r. 2003 bude poprvé konáno v Praze), a proto je jeho příprava nesmírně důležitou aktivitou Centra.

Název projektu :	*Centrum komputační lingvistiky*
Řešitel:	Prof. PhDr. Eva Hajičová, DrSc.
Příjemce dotace:	Univerzita Karlova v Praze, Ovocný trh 3/5, Praha 1, 110 00, IČO: 00216208

Časový harmonogram:



Upřesnění dílčích cílů pro rok 2003 – ÚJČ

Pro cíle jazykového poradenství a připravovaného projektu „Možnosti a mezi gramatiky češtiny ve světle Českého národního korpusu“ budeme svou účast na Centru více orientovat na možnosti, jak vyhledávat syntaktické struktury zapsané v podobě stromů v Pražském závislostním korpusu. Půjde zejména o slovosledné a syntakticko-slovosledné struktury, jejichž vyhledávání v ČNK není možné nebo je možné jen omezeně a s neúměrným nárokem na manuální práci:

- vývojová dynamičnost v nominální skupině: Jak se mění a) poměr (přibývající) prepozice a (ubývající) postpozice adjektivních atributů; b) poměr prepozice a postpozice se zřetelem na rozměr rozvíjejících členů c) konkurence posesivního adjektiva a genitivu substantiva; d) užívání neslovních / slovních nesklonných výrazů v prepozici (typ fotbalová Gambrinus liga); e) vzájemná poloha genitivního a předložkového atributu k témuž řídícímu jménu (typ překlad do ruštiny kolektivní monografie);
- slovosled věty a aktuální členění větné: poloha verba finita vůbec, poloha rematického verba finita, poloha verba finita ve vedlejších větách;
- statistika začátku a konce věty: zakotvenost věty v předcházejícím větném kontextu.

Název projektu :	*Centrum komputační lingvistiky*
Řešitel:	Prof. PhDr. Eva Hajičová, DrSc.
Příjemce dotace:	Univerzita Karlova v Praze, Ovocný trh 3/5, Praha 1, 110 00, IČO: 00216208

Upřesnění dílčích cílů pro rok 2003 – ZČU

V příštím roce předpokládáme pokračující práce na systému rozpoznávání spontánní češtiny. Především bychom chtěli testovat možnosti kombinace jazykových modelů vzniklých z několika zdrojů trénovacích dat a dále také vyzkoušet využití slovních kategorií při jazykovém modelování. Cílem je zvýšit robustnost systému ASR při nedostatku trénovacích dat pro jazykové modelování a současně určitým „na znalostech založeným" přístupem výběru slov pro rozpoznávací slovník postupně snižovat stále velmi vysoké procento slov mimo slovník – OOV rate.

Současně budou také pokračovat práce na zpracování dalšího slovanského jazyka v rámci projektu MALACH – ruštině. Práce spojené s anotacemi ruských výpovědí budou opět hrazeny z projektu MALACH, práce spojené s vývojem a testováním systému ASR pro spontánní ruštinu budou probíhat ve spolupráci s JHU v Baltimore na půdě CKL. Počítáme s tím, že budou využity všechny zkušenosti získané řešením ASR pro spontánní češtinu. Podle dostupných informací půjde vůbec o první na světě konstruovaný systém pro rozpoznávání spontánní souvislé ruštiny s velmi rozsáhlým slovníkem.

Další aktivity Centra plánované pro rok 2003

- Budeme **pokračovat** ve vydávání **technických zpráv** (ve spolupráci s ÚFALem MFF UK) o dílčích výsledcích výzkumu; v roce 2003 předpokládáme vydání tří výzkumných zpráv. Tyto zprávy budou též k dispozici na webových stránkách CKL.
- Alespoň na jednom z pracovišť Centra **uspořádáme Den otevřených dveří**, na němž seznámíme živou formou širší odbornou veřejnost a především zájemce ze středních škol s tématy, na nichž pracujeme, a s našimi výsledky.
- Předpokládáme krátkodobé příležitostné **přednáškové pobyty několika předních zahraničních profesorů** a pořádání alespoň dvou intenzivních přednáškových kursů zahraničních profesorů v průběhu kalendářního roku 2003, které se budou týkat aktuální problematiky řešené v rámci programu CKL.
- Centrum se podstatnou měrou podílí a bude podílet na organizaci 18. běhu mezinárodních **cyklů přednášek Centra Viléma Mathesia** v Praze 9.-22. března 2003. Kurzy budou vedeny 10 významnými zahraničními odborníky a nejméně čtyřmi pracovníky CKL. Veškeré organizační zajištění je dílem Centra. Bezplatně se přednášek zúčastní kolem 20 českých doktorandů a mladých vědeckých pracovníků.
- CKL uskuteční alespoň jeden **výjezdní seminář**, na kterém se budou soustavně projednávat jednotlivé úkoly Centra a mladí pracovníci Centra budou prezentovat své výsledky.

Název projektu :	*Centrum komputační lingvistiky*
Řešitel:	Prof. PhDr. Eva Hajičová, DrSc.
Příjemce dotace:	Univerzita Karlova v Praze, Ovocný trh 3/5, Praha 1, 110 00, IČO: 00216208

PŘÍLOHA: *Publikace*

1. Böhmová, Alena; Sgall, Petr (2002): The simple core and the complex periphery of natural language - A formal and a computational view. Proceedings of The 19th International Conference COLING 2002 (ed. Shu-Chuan Tseng), pp. 925-931.
2. Buráňová, Eva; Hajičová, Eva; Sgall, Petr (in print): Topic-Focus Articulation and degrees of salience in the Prague Dependency Treebank. In the festschrift for Eloise Jelinek, Arizona.
3. Byrne, William J.; Demner-Fushman, Dina; Dorr, Bonnie; Gustman, Samuel; Hajič, Jan; Oard, Douglas W.; Picheny, Michael; Ramabhadran, Bhuvana; Resnik, Philip; Soergel, Dagobert (2002): Cross-Language Access to Recorded Speech in the MALACH Project. TSD02, pp. 57-64.
4. Byrne, William; Gustman, Samuel; Hajič, Jan; Ircing, Pavel; Mírovský, Jiří; Psutka, Josef; Psutka, Josef V.; Radová, Vlasta; Ramabhadran, Bhuvana (submitted): Language Model Data Selection for Czech ASR in the MALACH Project. ICASSP 2003 in Hong Kong.
5. Byrne, William J.; Gustman, Samuel; Hajič, Jan; Ircing, Pavel; Psutka, Josef; Psutka, Josef V.; Radová, Vlasta; Ramabhadran, Bhuvana (2002): Automatic Transcription of Czech Language Oral History in the MALACH Project: Resources and Initial Experiments. TSD02, pp. 253-260.
6. Cuřín, Jan; Havelka, Jiří; Čmejrek, Martin (in print): Czech-English Dependency-based Machine Translation. PBML 78. MFF UK, Prague.
7. Ding, Yuan; Eisner, Jason; Hajič, Jan; Koo, Terry; Parton, Kristen; Penn, Gerald; Radev, Drago; Rambow, Owen; Čmejrek, Martin (2002): Natural Language Generation in the Context of Machine Translation. Workshop'02 Final Report, CLSP Technical Reports.
8. Džeroski, Sašo; Sgall, Petr; Žabokrtský, Zdeněk (2002): A Machine Learning Approach to Automatic Functor Assignment in the Prague Dependency Treebank. Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002), Las Palmas de Gran Canaria, Spain. Volume 5, pp. 1513-1520.
9. Hajič, Jan (2002): Tectogrammatical Representation: Towards a Minimal Transfer in Machine Translation. Proceedings of the 6th International Workshop on Tree Adjoining Grammars and Related Frameworks (TAG+6), 20-23 May 2002, pp. 216-226.
10. Hajičová, Eva (in print): Aspects of discourse structure. Natural language processing between linguistic inquiry and system engineering, Hamburg, April 29, 2002. In: Festschrift fuer Walther von Hahn.
11. Hajičová, Eva (in print): Possibilities and Limits of Optimality in Topic-Focus Articulation. Current issues in formal Slavic linguistics (eds. G. Zybatow, U. Junghanns, G. Mehlhorn and L. Szucsich), pp. 385-394. Frankfurt/M.: Peter Lang.
12. Hajičová, Eva (2002): řada hesel publikace. Encyklopedický slovník češtiny (eds. P. Karlík, M. Nekula, J. Pleskalová).
13. Hajičová, Eva (2002): Theoretical description of language as a basis of corpus annotation: The case of Prague Dependency Treebank. Prague Linguistic Circle Papers 4, pp. 111-127.
14. Hajičová, Eva; Kučerová, Ivona (2002): Argument/Valency Structure in PropBank, LCS Database and Prague Dependency Treebank: A Comparative Pilot Study. LREC 2002 Proceedings.
15. Hajičová, Eva; Pajas, Petr (2002): Corpus annotation on the tectogrammatical layer: Summarizing of first stages of evaluation. PBML 77, pp. 5-18.

16. Hajičová, Eva; Panevová, Jarmila; Sgall, Petr (2002): K nové úrovni bohemistické práce: Využití anotovaného korpusu (Towards a new level of the research of Czech: The use of an annotated corpus). Část 1 (Part 1). Slovo a slovesnost 63, pp. 161-177.
17. Hajičová, Eva; Panevová, Jarmila; Sgall, Petr (2002): K nové úrovni bohemistické práce: Využití anotovaného korpusu (Towards a new level of the research of Czech: The use of an annotated corpus). Část 2 (Part 2). Slovo a slovesnost 63, pp. 241-262.
18. Hajičová, Eva; Panevová, Jarmila; Sgall, Petr (2002): Úvod do teoretické a počítačové lingvistiky (Introduction to Theoretical and Computational Linguistics). Teoretická lingvistika, I. svazek (Theoretical Linguistics, Vol. I).
19. Hajičová, Eva; Sgall, Petr (2002): Are linguistic frameworks comparable? Computational Linguistics for the New Millenium: Divergence or Synergy? (eds. Manfred Klenner and Henriette Visser), pp. 113 - 122.
20. Hajičová, Eva; Sgall, Petr (in print): Dependency syntax in Functional Generative Desription. In the festschrift for P. Hellwig, Heidelberg.
21. Hajičová, Eva; Sgall, Petr; Veselá, Kateřina (in print): Information structure and contrastive topic. Presented at FASL 2002 Amherst.
22. Hana, Jiří; Hanová, Hana (in print): Manual for Morphological Annotation. Technical Report. MFF UK, Prague.
23. Holub, Martin; Pecina, Pavel (2002): Sémanticky signifikantní kolokace. UFAL/CKL Technical Report TR-2002-13.
24. Honetschläger, Václav (2002): Analytical and Tectogrammatical Syntactical Parsing. WDS 2002.
25. Ircing, Pavel; Psutka, Josef (2002): Lattice Rescoring in Czech LVCSR System Using Linguistic Knowledge. International Workshop Speech and Computer SPECOM 2002, St. Petersburgh, Russia, pp. 23-26.
26. Kučerová, Ivona (in print): Subjekt-predikátová shoda v češtine: univerzální, nebo specifická jazyková forma? Čeština. Univerzalia a specifika 4. Lidové noviny, Praha.
27. Kučerová, Ivona; Žabokrtský, Zdeněk (in print): Transforming Penn Treebank Phrase Trees into (Praguian) Tectogrammatical Dependency Trees. PBML 78. MFF UK, Prague.
28. Ljubopytnov, Vladimír; Němec, Petr; Pilátová, Michaela; Reschke, Jakub; Stuchl, Jan (2002): Oraculum, a System for Complex Linquistic Queries. Proceedings of Sofsem 2002 (29th Annual Conference on Current Trends in Theory and Practice of Informatics).
29. Lopatková, Markéta; Řezníčková, Veronika; Žabokrtský, Zdeněk (2002): Valency Lexicon for Czech: from Verbs to Nouns. TSD2002, Proceedings (eds. P. Sojka, I. Kopeček, K. Pala), Lecture Notes in Artificial Intelligence, vol.2448.
30. Lopatková, Markéta; Žabokrtský, Zdeněk; Skwarská, Karolina; Benešová, Václava (2002): Tektogramaticky anotovaný valenční slovník českých sloves. TR-2002-15.
31. Mírovský, Jiří; Ondruška, Roman (2002): Netgraph System-Searching through Prague Dependency Treebank. Prague Bulletin of Mathematical Linguistics 77.
32. Mírovský, Jiří; Ondruška, Roman; Průša, Daniel (2002): Searching through Prague Dependency Treebank-Conception and Architecture. Proceedings of "The First Workshop on Treebanks and Linguistic Theories", 20th and 21st September 2002, Sozopol, Bulgaria, pp. 114-122.
33. Mokrý, Karel; Smrž, Otakar (2002): External Tools Not Only for ArabTeX Documents. Proceedings of the International Symposium on the Processing of Arabic, pp. 161-165.

34. Panevová, Jarmila (2002): Corpus-based Grammar or Corpus Grammar-based? Referát přednesený na zasedání Komise pro gramatickou stavbu slovanských jazyků (Virrat, Finsko, 30.8. až 4.9.).
35. Panevová, Jarmila (2002): řada hesel publikace. Encyklopedický slovník češtiny (eds. P. Karlík, M. Nekula, J. Pleskalová).
36. Panevová, Jarmila (in print): Sloveso: centrum věty; valence: centrální pojem syntaxe. referát na konferenci Aktuálne otázky slovenskej syntaxe, Budmerice, listopad 2002.
37. Panevová, Jarmila; Ribarov, Kiril (in print): Za poleznosta na elektronskite jazični korpusi (vrz primerot na eden tip na imenskata fraza vo češkiot jazik). Slavistički studii. Skopje, Makedonie.
38. Panevová, Jarmila; Urešová, Zdeňka; Řezníčková, Veronika (2002): The Theory of Control Applied to the Prague Dependency Treebank (PDT). Proceedings of the 6th International Workshop on Tree Adjoining Grammars and Related Frameworks, 20-23 May 2002, pp. 175-180.
39. Pravdová, Markéta (2002): K povaze reklamního diskurzu. Naše řeč 85, pp. 177-189.
40. Pravdová, Markéta (2002): McSvět a místo člověka v něm. Studentská vědecká konference v Praze 26. a 27.4.2002, pp. 418-431.
41. Pravdová, Markéta (in print): Reklama jako zvláštní typ sdělování. Sborník ze 3. mezinárodního setkání mladých lingvistů, 14.-15.5.2002, Olomouc.
42. Ribarov, Kiril (2002): Old Sources and Modern Procedures: Computer Processing of Old-Church Slavonic. Proceedings of the Third International conference on Language Resources and Evaluation (LREC) 2002, Las Palmas de Gran Canaria, Spain. Volume 5, pp. 1622-1626.
43. Ribarov, Kiril (2002): On Rule-Based Parsing of Czech. PBML 77.
44. Ribarov, Kiril; Smrž, Otakar (2002): Searching for non-linearities in natural language. Poster at 7th Experimental Chaos Conference, August 25th-29th.
45. Řezníčková, Veronika (in print): PDT: Two Steps in Tectogrammatical Annotation ... with respect to the issues of deletions. PBML 78. Charles University, Prague.
46. Sgall, Petr (in print): Dynamics in the meaning of the sentence and of discourse. Meaning: The dynamic turn (ed. J. Peregrin). Kluwer.
47. Sgall, Petr (2002): Spoken Czech revisited. In press in the festschrift for Ch. Townsend, USA.
48. Sgall, Petr (2002): The freedom of language. Prague Linguistic Circle Papers 4, pp. 309-329.
49. Sgall, Petr (2002): Topic-Focus Articulation in Corpus Annotation. In press in the festschrift for Walther von Hahn, Hamburg.
50. Sgall, Petr (in print): Underlying structures in annotating Czech National Corpus. Current issues in formal Slavic linguistics (eds. G. Zybatow, U. Junghanns, G. Mehlhorn and L. Szucsich), pp. 499-505. Frankfurt/M.: Peter Lang.
51. Smrž, Otakar; Zemánek, Petr (in print): Sherds from an Arabic Treebanking Mosaic. PBML 78. MFF UK, Prague.
52. Smrž, Otakar; Zemánek, Petr; Šnaidauf, Jan (2002): Prague Dependency Treebank for Arabic: Multi-Level Annotation of Arabic Corpus. Proceedings of the International Symposium on the Processing of Arabic, pp. 147-155.
53. Straňáková-Lopatková, Markéta; Žabokrtský, Zdeněk (2002): Valency Dictionary of Czech Verbs: Complex Tectogrammatical Annotation. LREC2002, Proceedings, vol.III. (eds. M. González Rodríguez, C. Paz Suárez Araujo).

54. Straňáková-Lopatková, Markéta; Žabokrtský, Zdeněk (2002t): Valenční slovník stokrát jinak: co je pod povrchem? (Abstrakt) In: Čeština - univerzália a specifika 4. Sborník konference ve Šlapanicích u Brna, (ed.. Z. Hladká, P. Karlík), pp. 361-363.
55. Štícha, František (2002): Čas slovesný. Diateze. Gramatičnost. Hierarchizace sémantické struktury. Rod slovesný. Způsob slovesný. Osoba. Encyklopedický slovník češtiny (eds. P. Karlík, M. Nekula, J. Pleskalová).
56. Štícha, František (2002): Česko-německá srovnávací gramatika.
57. Štícha, František (2002): Český národní korpus. Úvod a příručka uživatele. - Praha, Filosofická fakulta UK 2000. Slovo a slovesnost 63, pp. 73-74.
58. Uhlířová, Ludmila (2002): E-mail as a new electronic medium in Prague Language Consulting Services. Referát přednesený na zasedání Komise pro gramatickou stavbu slovanských jazyků (Virrat, Finsko, 30.8. až 4.9.).
59. Uhlířová, Ludmila (2002): Jazyková poradna v měnící se komunikační situaci u nás. Sociologický časopis 38, pp. 443-455.
60. Urešová, Zdeňka; Řezníčková, Veronika (in print): K syntaktické anotaci textu z Českého národního korpusu: od analytické k tektogramatické rovině. Aktuálne otázky slovenskej syntaxe.
61. Vidová-Hladká, Barbora (in print): Pražský závislostní korpus aneb Co tady před padesáti lety nebylo. Pokroky matematiky a fyziky. MFF UK, Prague.
62. Zeman, Daniel (2002): Can Subcategorization Help a Statistical Dependency Parser? Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002), Taibei, Taiwan.
63. Zeman, Daniel (in print): How to Decrease Performance of a Statistical Parser. PBML 78. Charles University, Prague.

RECENZE

64. Blažek, David; Řezníčková, Veronika (2002): review of K. Böttger, S. Dönninghaus, R. Marzari (eds.): Beiträge der Europäischen Slavistischen Linguistik. Polyslav 4. Verlag Otto Sagner, München 2001. 292 p.. Slovo a slovesnost 63, pp. 227-232
65. Hajičová, Eva (2002): recenze knihy: Studie z korpusové lingvistiky (eds. F. Čermák, J. Klímová, V. Petkevič). Slovo a slovesnost 63, pp. 65-68.
66. Holub, Martin (2002): review of R. Harald Baayen: Word Frequency Distributions. PBML 77.
67. Ribarov, Kiril; Vidová-Hladká, Barbora (in print): Review of: L. Lebart, A. Salem, L. Berry: Exploring Textual Data. In: Text, Speech and Language Technology series, volume 4. Kluwer Academic Publishers. 1998. PBML 78. Charles University in Prague.
68. Štěpánek, Jan (in print): Review of: Building on Frege (eds. A. Newen, U. Nortmann, R. Stuhlmann-Laeisz). PBML 78. MFF UK, Prague.