

INTERCORP schůzka řešitelů 27.9.2005

Přítomni: , F. Čermák, V. Schmiedtová, M. Křen, J. Kocek, M. Vavřín, S. Cinková – švédština, P. Čermák – španělština, v. Dovalil – němčina, Forková – polština, Geiklová – francouzština, Haušildová – holandština, T. Káňa – němčina, A. Klégr – angličtina, Kolmanová – maďarština, H. Lindroos – finština, M. Nábělková – slovenština, H. Peloušková – němčina, P. Štichauer – italština, N. Kalajdzievová - bulharština

prof. Čermák - přivítání účastníků na druhém setkání paralelních korpusů a zároveň představil pana Vavřína, který má na starosti technickou stránku paralelních korpusů v ÚČNK

e-mail: martin.vavrin@ff.cuni.cz – zde se můžete obrátit s dotazy, které se týkají Vaší práce s paralelním korpusem

- na konci setkání ukázka německého paralelního korpusu z Brna – ukázka možností jak znázorňovat paralelní korpus na webových stránkách
- peníze na výzkumný záměr došly na účet později, proto se později nakoupilo hardwarové vybavení – počítač + skenery podle zadaných požadavků a následná instalace, bohužel nakoupení tohoto vybavení bylo uskutečněno v době prázdnin, tím došlo k posunu zaškolení do skenování
- pokud bude potřeba nakoupit počítače a skenery je možnost do konce roku
- pokud bude zájem je možno si domluvit další školení na skenování
- zásada je, aby neměly malé obory svůj skener, ale aby se domluvily na spolupráci s některými dalšími malými obory. Můžeme nakoupit také diskety, flash disky a CD.
- byla zřízena webová stránka s diskusním forem – trnka.ff.cuni.cz/ucnk/intercorp
- byl nakoupen software ParaConc – multilicence, kde každý obor - jazyk získá po jedné kopii.
- zásady při práci s textem- I. fáze – oskenování textu
 - II. fáze – zarovnání textu po odstavcích – p. Vavřín rozeslal podrobný návod jak postupovat při zarovnávání textu, aby se s ním dalo pracovat v programu ParaConc
- finance: pro každého řešitele – koordinátora za jazyk je počítáno s odměnou
 - peníze pro skenující jsou připraveny k vyplacení – předpokládá se, že si koordinátoři vytipují studenty, kteří se zaučí do skenování, a pokud možno vydrží delší čas
- zahraniční cesty – každý jazyk má naplánovanou jednu cestu do zahraničí
- naskenované knihy odevzdávat, pokud je to možné, zarovnané s českou verzí
- nákup knih (e-booky) přes platební kartu

Celková rekapitulace podle jazyků:

Srbština – Adamovičová – omluvena, na služební cestě v zahraničí
naskenováno 10 knih, vytvořený seznam knih, které v budoucnu zahrne do paralelního korpusu

Slovinština – Blažek – bude se skenovat do konce roku cca 10 knih

Nizozemština – Boková - nepřítomna

Švédština – Cinková – vytvořila seznam knih, které je možno zahrnout do paralelního korpusu

- překlady ze švédštiny do češtiny
naskenováno cca 25 knih, 3 rukopisy z češtiny do švédštiny
- překlady z češtiny do švédštiny – 320 knih v knihovně, 200 knih by přicházely v úvahu
- e-book možno koupit originál (cena 1 knihy – 250,- Kč)
- skenování přes digitalizační centrum v Akademii věd lhotak@lib.cas.cz
- knihy pouze naskenují 2,50 / stránku , velký objem 6 000 stránek 2,-/stránku

- doba zpracování - týden , předávání na DVD- neopravené soubory

Španělština – Čermák – naskenováno 13 knih v obou verzích, 10 knih v české verzi
- vytvořen přehled titulů, které budou zahrnuty do paralelního korpusu

Otázka: pokud se podaří získat elektronické verze, jak bude placena?

Prof. Čermák: Pokud bude potřeba zaplatit poplatek za poskytnutí textů, jsme schopni tuto částku zaplatit.

Němčina - Dovalil – naskenováno 6 dvojic knih, potřebují zaučit nové studenty pro skenování – je možno se domluvit přes e-mail. Vavřín - max počet 7 lidí
- objeven zdroj elektronických textů, který musí ještě prozkoumat

Pozn. prof. Čermák: Bylo by dobré prozkoumat také elektronické zdroje mezinárodní: např. Oxfordský textový archiv (OTA)

Polština – Forková – naskenováno 10 knih – polské texty
- navazují spolupráci s nakladatelstvími, s polskými korpusy –

Otázka: Jak využít spolupráci, zda pouze získat text, nebo pokračovat ve spolupráci?

Prof. Čermák: Idea – korpusy chceme zveřejnit na internetu, uvážit, zda nakladatelství povolí zveřejnění. Projekt je výzkumný, budou výzkumné výstupy, studie, články, texty by měly být zveřejnitelné pro akademickou obec.

Otázka: Jaké texty zahrnout do výběru?

Prof. Čermák: Texty – počty nejsou omezené, časové omezení: od druhé světové války – beletrie, přírodní vědy, historie

Francouzština – Geiklová – od Sváška 10 naskenovaných knih
- chystají se navázat spolupráci s nakladatelstvími

Bulharština – Kalajdžijevová – 2 páry knih odevzdány
2 páry připraveny k odevzdání
18 knih chtějí naskenovat
Elektronické podoby textů – nejsou, zatím není bulharský korpus
Skenují se obě verze

Diskuse:

Reakce na skenování – příklad: 300 stran ve Wordu, vychází 50 Kč na hodinu, studenti nechtějí za tyto peníze skenovat – nemají zaplacenou cestu do knihovny, dlouho trvá naskenování a korektura textu, a další dokončovací práce.

Káňa: V Brně se platí 70 Kč za hodinu- studenti jsou ochotni za tyto peníze pracovat.

Prof. Čermák: Zvážit, zda se jedná o korekturu cizího textu, nebo českého textu. Je potřeba nastavit cenu, aby studenti byli ochotni tuto práci vykonávat.

Vavřín: Je potřeba zvážit počáteční obtíže při zaškolování.

Brno nabízí hodinovou mzdu jako optimální řešení.

Angličtina vyplácí podle objemu textu.

Prof. Čermák: Jak honorovat? Zda podle hodin, nebo podle objemu. Zda také honorovat práci, která je potřeba kolem.

Schmidtová: Přislíbila zlepšit hodinovou mzdu.

Dánština – Haušildová – má zapůjčen skener pro práci v Dánsku
- sestaven seznam knih, které budou zahrnuty do paralelního korpusu 170 knih
v současné době v Dánsku, kde získává překlady z češtiny do dánštiny
- letos naskenováno 5 knih

Schmiedtová: Je možno zapůjčit skener – pokud by bylo potřeba skenovat v knihovně v cizině

Portugalština – Jindrová - omluvena

Chorvatština – Jirásek – vytvářejí seznam překladů, které jsou do češtiny a do chorvatštiny, vycházejí ze stránek obce překladatelů, bibliografie Dušana Karpatský

Bibliografie slovanské knihovny

Rejstřík české národní knihovny, Záhřebské národní knihovny

Oslovily překladatele v Chorvatsku – zda by nemohli získat texty z Chorvatska od překladatelů, čekají, zda se podaří získat elektr. verzi textů, někteří neodpověděli, někteří přislíbili spolupráci zatím čekají se skenováním, mají 26 knih naskenováno z chorvatské verze, do konce roku 40 knih, oddělena fáze od skenování od korektur, jak honorovat korekturu cizojazyčného textu – studenti nejsou ochotni pracovat za 50,- Kč

Prof. Čermák připomněl, že produkt práce je majetkem ÚČNK.

Němčina – Brno – Káňa – připraveno 15 dvojic textů, zařazeno 211 párů textu
Česko – německo paralelní korpus má 3,5 milionů slov

Angličtina – Klégr – spousta textů v elektr. podobě, databáze obsahuje 152 dvojic, k dispozici 82 dvojic, 164 textů, zbývá doskenovat 70 textů, 10 milionů slov, texty jsou pouze spárovány, doplňují texty do databáze, skenovat, upravovat texty, cíl krátkodobý- začistit 15 – 20 párů pro publikační výstup,

Prof. Čermák: Zdroje textů – ústava EU - ÚČNK musí vyzkoušet možnost, jak texty stáhnout

Maďarština – Kolmanová – začne pracovat v blízké době, mají zpracovanou vlastní literaturu, v národní knihovně, možno vyžádat efekt. texty v knihovně, dávají dohromady studenty

Finština – Lindroos – 2 knihy naskenovány

Slovenština – Nábělková - 5 párů knih do konce roku, spolupracují se Slovenským národním korpusem

Prof. Čermák: Učebnice – nepoužívat – nemají kontext

Makedonština – Kiril – naskenováno v Akademii věd, problém půjčování knih

Čínština - nepřítomna

Litevština - nepřítomna

Italština – Štichauer – seznam literatury, vycházel z databáze ÚČNK, oslovil vydavatele, čeká na odezvu, zaslán oficiální dopis + žádost, pokud nakladatelství neposkytnou texty, pak budou skenovat, připravena 1 kniha nasekovaná

Arabština – nepřítomna

Japonština – nepřítomna

Lotyština – nepřítomna

Ruština – nepřítomni

Ukrajinaština – nepřítomni

Na závěr schůzky byla prezentována ukázka německého korpusu – Káňa, Peloušková.