

Komunikace a jazyková správnost

Předmět Základy odborného stylu
(VB000)

K. Pala, Centrum ZPJ FI MU
podzim 2016

O čem bude řeč?

- Komunikace v PJ
- Zvuky a znaky
- Pravopisné systémy
- Pravidla (českého) pravopisu
- Jazyková správnost
- Pravopisné chyby a jejich opravování
- Korektory – pravopisné, gramatické
- Komunikace H-C, jednocestná : dvoucestná
- Počítače a PJ, nástroje pro PJ

Pozitivní komunikace

- Výchozí axiom: nelze nekomunikovat
- Naše civilizace stojí na komunikaci – je to základ pro vědu, techniku, kulturu, pro veškeré znalosti lidstva
- Musí splňovat jisté standardy, aby byla efektivní
- Pro úspěšnou komunikaci potřebujeme vhodná pravidla, normy
- Většina lidské komunikace probíhá v přirozeném jazyce
- Má dvě základní podoby: mluvenou a psanou
- Mluvený jazyk je výchozí
- Potřeba norem pro přechod od mluvené podoby k psané

Mluvený a psaný jazyk

- Mluvený jazyk je primární (cca 5000 jazyků světa)
- Psaný jazyk je až sekundární – je to ale paměť lidstva
- Psané texty v přirozeném jazyce vyžadují přiřazení zvuků jazyka (hlásek) psaným znakům (písmenům)
- Úplná reprezentace zvuků v jednotlivých jazycích světa – k tomu slouží International Phonetic Alphabet (IPA)
- Reprezentace zvuků – různé pravopisné systémy
- Jsou základem veškeré civilizace a kultury obecně
- Abecední písma – latinka, cyrilice a mnohá další
- Ideografická písma (logografická) – čínština, cca 50 tis. znaků (základní soubor čítá 9-12 tis. znaků, standard)
- Slabičná písma – japonština, korejština, jiná - gruzínština

Přiřazení zvuků znakům

- Kolik hlásek (fonémů) má čeština?
- Kolik písmen má čeština?
- Kolik hlásek (fonémů) je v angličtině?
- Kolik písmen je v angličtině?
- 40-42: 36 (piš, jak slyšíš)
- 40-44: 26
- Uvedená čísla předurčují povahu daného pravopisného systému – fonetický (čeština) vs. historický (angličtina – 14. stol.)
- Spřežkové a diakritické systémy (v češtině)

Funkce pravopisných systémů

- **Zaznamenávací** – aby se to dobře psalo
- **Vybavovací** – aby se to dobře četlo
- Která funkce je důležitější? Vyváženost?
- Posílíme-li jednu funkci, oslabíme druhou a naopak
- Jaká je situace v praxi?
- Historické systémy (typicky angličtina, 14. stol.)
- Fonetické systémy (Husova reforma, čeština, slovenština)
- Který systém je obtížnější k naučení?
- Spřežkové systémy – stará čeština, dnes zčásti i polština
- Lze získat přes milion liber za úspěšnou reformu anglického pravopisu? (nadace Bernarda Shawa, 1910)
- Pravopisné systémy jsou velmi konzervativní (viz spory kolem reformy v němčině – Ústavní soud)

Jazykové chyby

- Jazyková správnost a gramotnost – podmínka úspěšné komunikace a profesionální úspěšnosti obecně
- Pravopis není gramatika (popis struktury jazyka), pravidla pravopisu – převod zvuků na znaky
- V čem se nejvíc chybuje? (styl, interpunkce)
- Typy pravopisných chyb v textech
- Překlepy (*prgram, studijní, ...*)
- Morfologické chyby – koncovky (*hloupejma, kerí*)
- Syntaktické chyby – shoda (psaní y/i), vazby (valence)
- Styl a stylistické chyby – *provedení nařízení*, opakování slov jako *prostě, teda*, chybná formulace myšlenek
- Typografické chyby – mezery, pomlčky, spojovníky, uvozovky, fonty, jednopísmenové předložky na konci řádků
- Spisovnost : nespisovnost, formálnost : neformálnost

Opravování chyb v textech

- Snahou je, aby v našich textech bylo minimum chyb
- Lidé – korektoři nejsou dokonalí, v textech zůstávají vždy nějaké chyby
- Co považujeme za chyby?
- Vznikly pomůcky obsahující informace o jazykové správnosti – u nás Pravidla č. p., v angličtině slovníky
- Knižní příručky – slovníky – u nás norma: SSČ a SSJČ
- Elektronické nástroje odhalující překlepy a jiné chyby v textech: různé typy korektorů (gramatické, stylistické)
- Korektory dnes najdeme v produktech, jako jsou Microsoft Office (Word), Open Office aj.
- Co dovedou? Mohou být chytřejší než uživatel?

Pravidla českého pravopisu I

- Je český pravopis obtížný?
- Kombinace fonetického a historického principu, fonetický převažuje – piš, jak slyšíš, ang. je těžší
- Je dobré vědět, v čem se nejvíc chybuje?
- Stylistické chyby – cca 23 %, vznikají při snaze formulovat myšlenky, to je nesnadná úloha
- Interpunkční chyby – asi 20 %, mají synt. povahu
- Ostatní chyby – překlepy, *y/i*, velká písmena, tvary *mne/mě*, koncovky, typografické chyby aj.
- Reformy českého pravopisu – pokusy

Pravidla českého pravopisu II

- Pravidla č.p. existují v knižní a elektronické podobě (poslední sporné vydání ÚJČ – 1993)
- Internetová jazyková příručka – ÚJČ a FI MU
- Má dvě části – slovníkovou a normativní
- Slovníková část pokrývá cca 60 000 čes. slov
- IJP běží na serveru Centra ZPJ, denně téměř 40 tis. přístupů (ukázka) , též kniha Ak. přír. č. j.
- Příručka obsahuje automatickou morfologii a je nově doplněna o dva normativní slovníky
- Používá se standardně ve školách a institucích

Korektory – pravopisné a jiné

- Jak jsou tyto nástroje konstruovány?
- Kolik slov má čeština? Kolik slovních tvarů?
- PSJČ – cca 250 000 základních tvarů slov
- Slovních tvarů v češtině je kolem 60 milionů
- Morfologická analýza a morfologický analyzátor je základem pro konstrukci korektoru překlepů
- Pro češtinu – morf. analyzátor Majka, cca 400 000 českých kmenů (vytvořen v Centru ZPJ, používá jej firma Seznam ve svém vyhledávači)

Gramatické korektory I

- Dovedou opravovat slovní spojení v kontextu, např. *studentka šel do školi když pršeló.*
- Na rozdíl od korektoru překlepů gram. korektor Wordu podtrhává zeleně a opravy jen doporučuje
- Chyby v gramatické shodě a slovesn. vazbách
- Upozorňují také na chyby v interpunkci (ne na všechny, úspěšnost kolem 50 %)
- Povaha české interpunkce je syntaktická
- Gramatický korektor je k dispozici v české verzi Wordu (autoři: Oliva, Květoň, Petkevič)

Gramatické korektory II

- Nástroj: Grammaticon (od firmy Lingea) – co umí? Nízké pokrytí, kolem 10 % – falešné hlášky
- Co gramatické korektory nedovedou?
- Nakolik se na ně lze spolehnout?
- Evaluační parametry: přesnost a pokrytí
- Přesnost (úspěšnost) se pohybuje do 60 %, jde o velmi těžkou úlohu, patří do obl. UI
- Principy fungování gramatických korektorů?
- Automatická syntaktická analýza – parsery (Set)
- Heuristická pravidla – negativní příklady

Současný stav českého pravopisu I

- Poslední reforma čes. pravop. proběhla v r. 1993
- Pokus o tzv. „demokratizaci“ čes. pravopisu?
- Úprava psaní slov cizího původu (kurs/kurz)
- Ref. vedla k malé pravopisné válce (filos/zofie) – láska k moudrosti vs. láska k temnotě
- Výsledek: špinavý kompromis (větší fonetizace)
- spornost reformy si lze ověřit na datech – dnes na velkých souborech textů – korpusech

Jaký je tedy současný stav? – je vidět, že norma je rozkolísaná – to je nejhorší možný výsledek

Současný stav II

- Korpusy (CzTenTen12, 5 miliard) to potvrzují
kurs: 65,360 vs. kurz: 581,913, feminis/zmus:
11,895 vs. 332 výskytů
- Rozkolísanost standardů je obecně nežádoucí,
komplikuje plynulost komunikace, reforma?
- Bezbolestná úprava by např. byla *ú/ů*
- Za problém se pokládají velká písmena,
volnost je značná, lze vidět vliv angličtiny
- Pokud jde o *y/i*, situace pro úpravu není zralá
- Případná inspirace slovenštinou, chytré řešení

Předpokládaný vývoj

- Institucí, která se stará o jazykovou kulturu v ČR, je Ústav pro jazyk český AV ČR
- Komunikuje s veřejností prostřednictvím Jazykové poradny (a nyní též IJP)
- Jejich aktuální přístup k problematice českého pravopisu je spíše liberální a zbytečně opatrný
- Lze očekávat nějaké reformy? Spíše ne.
- Pro absolventy FI je žádoucí, aby psali kultivovaně a bez chyb (bakalářské, diplomové práce)

Počítače a PJ I

- Komunikace mezi člověkem a počítačem je dnes primárně jednocestná
- Její kvalita ve skutečnosti závisí na tom, jak dobře uživatel zná programové vybavení svého počítače (jeho OS)
- Počítače s námi zatím nedovedou přirozeně komunikovat – úloha je obtížná
- Potřeba dvoucestnosti – zpracování PJ je součástí umělé inteligence (viz obor UMI na FI)
- Jde tu o modelování některých funkcí lidského mozku na počítači

Počítače a PJ II

- UI a počítačové zpracování přirozeného jazyka
- Tři součásti – reprezentace znalostí o světě, inference (logika), znalost PJ (gramatika)
- Dialogové systémy je musí obsahovat
- Turingův test, Eliza, chatboty, každoroční soutěž o Loebnerovu cenu, roboty
- Zpracování mluvené řeči – diktovací systémy
- Dovedou přepisovat zvuky na znaky – pro češtinu: Newton Technologies, Dictate 4.5, 7 tis.
- U těchto systémů ještě nejde o porozumění PJ

Sw. nástroje pro práci s jazykem

- Elektronické slovníky české – nástroj DebDict
- Vícejazyčné elektronické slovníky – např. produkty firmy Lingea
- Google Translator
- České překladače: Eurotran, PC Translator – málo kvalitní, úspěšnost dosahuje 60-70 %
- Problematika strojového překladu obecně
- Morfologické a syntaktické analyzátoři
- Dialogové systémy
- Porozumění přirozenému jazyku (extrakce inf.)

Poznámka k předmětu ZOS

- Na FI je celkem běžné, že bc. práce obsahují víc než 10 pravopisných chyb
- Svědčí to o nízkých kvalitách bc. studentů, které je veřejně vidět v ISu
- Uvažuje se o stanovení hranice – práce s vyšším počtem chyb nebudou akceptovány
- Klíč je ve vedoucích a oponentech, kteří jsou často slepými vedoucími jiné slepé
- Vazba předmětu ZOS na SBAPR – bc. práce
- Cíl: co nejkvalitnější bc. a diplomové práce

Mgr. Jiří Materna, Seznam

- **Vysokou školu jste neprospěl pouze u počítače, ale zahrál jste si například i ve fakultním divadle. Jaké to bylo, když se informatici pustili do divadla?**

Celá řada lidí si myslí, že divadlo na fakultě informatiky nemá co dělat, ale já jsem přesvědčený o opak. Problémem českého školství je obrovská pasivita studentů a jejich neschopnost prezentovat své myšlenky na veřejnosti. Je to dáno především tím, že k tomu nejsou vedeni již od útlého věku. Často