

Možnosti elektronického zpracování velkých esperantských slovníků

Předneseno na konferenci KAEST v Dobřichovicích na podzim r. 2004

Jednou z podmínek dalšího rozvoje esperanta, zejména jeho užívání v odborné sféře a také nedílnou součástí průkazu jeho použitelnosti jako mostového jazyka, potřebného zejména v současné době, kdy je hledáno levné a při tom demokratické řešení jazykové problematiky ve sjednocující se Evropě, je m.j. existence dobrých a dostatečně obsáhlých slovníků výkladových a také překladových z esperanta do národních jazyků a naopak. Bylo velkým handicapem esperanta, že velký slovník, tedy slovník obsahující více jak 100 tis. hesel dosud nemělo.

Dal jsem si tu práci a spočítáním hesel na každé dvacáté stránce jsem zjistil přibližné počty slov obsažených v některých slovnících. Kraftův Česko-esperantský slovník má asi 52 tis. hesel, Nový PIV asi 50 tis. hesel a Krausův Esperantsko-německý slovník asi 80 tis. hesel.

Esperanto bohužel nemá ani oficiální výkladový slovník, jaký má většina národních jazyků. Oficiálním esperantským lexikografickým dokumentem je Akademický slovník. Ten je ovšem jen oficiálním, velmi úzkým seznamem vybraných slov, chybí mu zcela výkladová část. Jediný nový střední výkladový slovník esperanta je PIV (Plena Ilustrita Vortaro - Úplný ilustrovaný slovník), který je proto převážnou většinou esperantistů považován a vydáván za oficiální, i když s Akademickým slovníkem je v řadě případů v rozporu. Také rozsahem zmíněných 50 tis. slov ho nelze porovnávat s oficiálními národními jazykovými slovníky. Například největší český výkladový slovník, Příruční slovník jazyka českého vycházel od r. 1935 a dokončen byl Ústavem pro jazyk český v roce 1957. Má 9 svazků a obsahuje přes 250 tis. hesel. Autorský kolektiv byl tehdy oceněn státním vyznamenáním. Výkladové slovníky některých evropských národů jsou údajně ještě obsáhlejší. Sjednocující význam PIV pro esperanto je ale nepochybný a jistě všichni oceňujeme jeho existenci.

Je také pravdou, že střední esperantské slovníky jsou pro běžnou jazykovou praxi postačující. To ale neplatí pro práci vědeckou a odbornou vůbec, pro precizní vyjadřování právních a politických vztahů a pod. Zde je zapotřebí, aby každé slovo, každý odborný termín měly svůj jednoznačný oficiální význam. Nestačí tedy, aby např. v případě mezinárodního právního sporu bylo rozhodováno podle slovníku, který je většinou považován za oficiální.

Jaké jsou příčiny tohoto neutěšeného stavu? Asi jich bude ještě víc, ale připomeňme alespoň tyto:

1. Esperanto nemá oficiální lexikografickou instituci pověřenou zpracováním oficiálního výkladového slovníku. Akademie, která má k plnění tohoto úkolu nejblíže ho neplní a dokonce se nepokládá za kompetentní ho plnit. (viz

článek Wery Blanke: Terminologia Esperanto-Centro: spertoj, problemoj, perspektivoj - Esperantské teminologické centrum: zkušenosti, problémy, perspektivy).

2. Práce slovníkářů ani tvorba terminářů není nijak řízena, nekonají se žádná mezinárodní setkání, výměna zkušeností je nulová.

Metody práce na slovnících a terminářích jsou zastaralé. Počítač je využíván většinou jen jako psací a typografický stroj. Knižně vydávané slovníky jsou drahé, velké překladové slovníky, které by měly více set stran nemohou být zaplacený esperantisty jednoho národa. Knižní slovníky neumožňují účinnou a širokou diskusi o obsahu, sporných výrazech a termínech, protože případné změny nebo opravy jsou prakticky nemožné. Vidíme to například na PIV, kde připomínky jsou sice zveřejňovány na internetu, možná že si ojedinelí esperantisté provedou úpravy ve svém exempláři, ale masa slovníků již zůstane tak, jak byly vydány.

Velkou nadějí esperanta pro řešení této problematiky jsou proto slovníky elektronické. Náklady na pořízení nosiče dat, kopírování a distribuci elektronického slovníku jsou nesrovnatelně nižší než náklady na papír, tisk, vazbu a distribuci papírového slovníku. Mimo to elektronický slovník může být v aktuální podobě umístěn na internetu nebo vydáván na CD v malém nákladu, i postupně podle odbytu, při tom mohou být průběžně prováděny opravy a doplňky z podnětu autora i na základě připomínek uživatelů. To umožňuje m.j. dosud neběžnou spolupráci uživatelů s autorem nebo autorským kolektivem.

Aby mohla být taková práce prováděna, mělo by být nejprve vybráno vhodné programové vybavení. V té souvislosti bych vás rád seznámil s dosavadními výsledky své práce na elektronickém slovníku, zejména z hlediska použitého programového vybavení. Jak mnozí z vás vědí, zabývám se již několik let tvorbou elektronického "Velkého esperantsko-českého slovníku" ve všeobecně známém tabulkovém programu Microsoft - Excel. Tento program jsem si pro tvorbu slovníku původně vybral proto, že jsem ho dobře znal ze svého zaměstnání, kde jsem v něm zpracovával velké rozborové a plánové databázové tabulky. Zpočátku jsem ze slovníkářského hlediska na program oceňoval zejména přehlednost uspořádání v tabulce, možnost automatického alfanumerického řazení dat, resp. slov, a snadné vyhledávání vřazených esperantských nebo českých slov. Postupně jsem ale objevoval a ještě stále objevuji další přednosti a možnosti tohoto programu.

Pro tvorbu esperantsko-českého slovníku jsem si stanovil cíl sebrat co největší zásobu esperantských slov včetně odborných termínů a přinést jejich překlad do češtiny. Slovník je určen zatím především pro Čechy, neobsahuje gramatické údaje pro česká slova, ale naopak obsahuje gramatické údaje k esperantským slovům, zejména kmeny slov a tranzitivnost sloves. Slovník usiluje o podrobné označování příslušnosti slov k jednotlivým oborům činnosti, resp. vědy, aby

mohl posloužit jako východisko pro další tvorbu odborných překladových slovníků odborníky příslušných oborů. K málo známým slovům jsou připojeny vysvětlivky v češtině. Kmenové slovo je vytištěno tučně a může je uvedeno na počátku příslušného hnízda slov. K četným slovům jsou připojeny odkazy na synonyma nebo příbuzná slova. Při tvorbě slovníku používám pro svoji potřebu ještě některé další údaje, např. údaj o zdroji ze kterého byl překlad slova získán, u mnohých slov odkaz na gramatiku, kde je o slově podrobněji pojednáno, údaj zda kmen slova je nebo není v PIV nebo akademickém slovníku a pod. Pro verzi připravovanou k prezentaci jsou některé údaje, resp. sloupce tabulky odstraněny a naopak k pohodlnému a rychlému vyhledávání slov přidány sloupce s uvedením esperantských slov se spřežkami a předřazenými tečkami nebo čárkami.

Esperantská a česká slova a všechny doplňující údaje jsou v tabulce uvedeny tak že každý z údajů zaujímá jednu buňku tabulky. O každém slově je tedy uváděno několik údajů, takže slovníkem, jehož slovní zásoba je nyní asi 145 tisíc hesel je obsazeno více jak milion buněk. Doufám, že počet slov ještě významně vzroste. Program a moje nepřilíš moderní hardware toto množství dat zvládají bez problémů. Pro podrobnější popis práce na slovníku bohužel zde není místo. V zásadě se dá říci, že je pracováno běžnými databázovými metodami. Uvedu proto jen některé z možností využití těchto metod pro lexikografickou práci.

Kromě již zmíněných výhod Excelu, tedy všeobecnou dostupnost a zvládání velkého množství dat jsem se již zmínil o bezproblémovém alfanumerickém řazení dat. Jsou-li vloženy potřebné třídící údaje, je možné je pozměnit podle přání tvůrce slovníku například využít

- prosté abecední řazení esperantských slov,
- abecední řazení podle slovních kmenů a uvnitř skupiny každého kmene opět abecedně, popřípadě s tím, že kmenové slovo bude jako první,
- řazení termínů podle oborů a uvnitř každého oboru opět abecedně,
- spojit některé obory, např. anatomii, lékařství a farmacii a vytvořit tak širší oborový slovník.

Dalšími možnostmi, které Excel tvůrcům slovníků poskytuje jsou zejména

- snadný import dat z databázových souborů a zejména dat vytvořených v textovém editoru Word nebo do něho převedených nebo naskenovaných textů,
- snadný export dat do databázových souborů a do textového formátu Word, pokud by byl zájem výsledný produkt prezentovat jako databázi nebo text,
- možnost různých grafických úprav v buňkách, zejména volba fontů, typů a velikostí písma, barvy písma a podkladu,

- možnost hromadného kopírování, transportu a mazání obsahu buněk, očíslování řádků a pod.,
- texty z jedné buňky je možné různě rozdělit do více buněk a naopak obsah více buněk spojit do jedné buňky atp.

S využitím jmenovaných, případně dalších úkonů lze sestavit různé, pro tvorbu slovníku vhodné postupy automatizované práce nebo postupy k usnadnění ruční práce, lze např.:

- posoudit soubor esperantských slov, zda jednotlivá slova v něm obsažená jsou obsažena v jiném souboru slov, např. zda slova určitého slovníku jsou již obsažena ve vytvářeném slovníku, patřičně je označit. Proveďte se to tak, že se slova v novém souboru označí, oba soubory se sloučí do jednoho, seřadí podle abecedy a vloží funkce označující slova, která se vyskytnou za sebou opakovaně,
- provést snadnou přestavbu slovníku na slovník jiného druhu nebo jiné formy; např. není velkým problémem ze slovníku vytvořit specializovaný oborový slovník nebo seznam slov pro kontrolu gramatické správnosti esperantského textu vytvářeného v textovém editoru,
- relativně snadno provést excerpci slov obsažených v určitém textu. V národních jazycích byla dosud pro velkou pracnost provedena excerpce jen z mála knih. V esperantu je pro jeho pravidelnost s využitím počítačové techniky excerpce v textu použitých slov poměrně snadná. Např. excerpce slov z esperantského překladu Čapkova románu „Válka s mloky“ jsem v programech Word a Excel provedl asi za 3 hodiny a zjistil, že bylo použito celkem 70 tis. slov a z toho 13 tis. různých slov. Zatím sice pro Velký slovník nehrozí vyčerpání slovní zásoby již zpracované v jiných slovnících a terminářích. Přesto snad jednou slovní zásoba Velkého slovníku vzroste natolik, že hledání dalších slov bude možné jedině excerpce textů. V současné době excerpce odborných textů může být významným zdrojem pro terminologickou práci.
- provést eventuální přestavbu esperantsko-českého slovníku na česko-esperantský. Esperantsko-český slovník, kde ke každému esperantskému slovu je uvedeno více možných českých významů, lze sice použít i k hledání esperantských výrazů pro česká slova, ale jen za cenu, že úkon hledání provedeme vícekrát, protože jednomu českému slovu může odpovídat více esperantských slov, uvedených ve více buňkách. V česko-esperantském slovníku by pro jedno české slovo byly všechny esperantské ekvivalenty uvedeny v jedné buňce. Zatím jsem přestavbu slovníku takto neprováděl, ale věřím že je v programu Excel proveditelná také s využitím více automatizovaných postupů, tedy minimálně v vynaložení mnohem menšího množství práce.

Samozřejmě Excel nebyl vytvořen pro zpracování slovníků a esperantských tím méně. Proto se při zpracování slovníku musíme v některých věcech přizpůsobit programu. Podle mého názoru ale toto přizpůsobení je nepodstatné v relaci ke zmíněným výhodám, které program poskytuje.

Např. Excel není uzpůsoben pro přímou práci s esperantskými znaky. Při tvorbě slovníku musíme použít spřežky s X, které zajišťují správné abecední řazení.

Tabulka je pro prezentaci slovníku zbytečně široká, má 230 sloupců, stačila by jich méně než pětina, ale má jen 65.536 řádků, zatím co slovník by jich potřeboval až pětkrát tolik. Musíme to zatím obejít tak, že pro prezentaci slovníku položíme více skupin sloupců vedle sebe.

Aby vyhledávání slov probíhalo správně, nelze v žádném elektronickém slovníku používat zkratky slov, tedy ani v knižních slovnících obvyklé vlnovky k označení kmenu slova.

V esperantském prostředí využitím elektronické formy mohou být vytvořeny podmínky pro práci na velkém mezinárodním výkladovém a překladovém slovníku. Konkrétně by to bylo možné tak, že na základě dosud zpracovaných esperantsko-národních elektronických slovníků by byla vypracována co možná nejobsáhlejší levá (esperantská) strana slovníku, ke které by esperantisté různých národů doplňovali (podle svých možností postupně) pravou část ve svém národním jazyce. Tím by byla dodržena zásada zpracování slovníku do určitého jazyka rodilým mluvčím.

Mimoto by mohla být zpracovávána a postupně doplňována nebo upravována výkladová část k jednotlivým esperantským slovům a termínům. Začít by se mohlo, pokud by autoři a vydavatel souhlasili, využitím Nového PIV. Byly by tak položeny základy pro tvorbu oficiálního výkladového slovníku esperanta.

Takovou mezinárodní spoluprací by vznikl soubor vzájemně provázaných elektronických slovníků, který by nad svou normální užitkovou hodnotu dokládal mimořádnost esperanta jako mostového jazyka. Slovníky by mohly být dále doplňovány a upravovány podle vývoje jazyků. Vznikl by tak nejenom samotný slovník, ale také široký mezinárodní autorský kolektiv, který by se mohl stát základnou pro ustavení oficiální lexikografické instituce. Samozřejmě by byla vítána i spolupráce uživatelů, zejména upozorňování na chyby a na slova, která ve slovníku ještě chybí.

Pro tvorbu velkého esperantského mezinárodního slovníku, jak jsem o něm hovořil, považuji za velikou přednost rozšířenost programu Excel po celém světě. Protože Excel je prodáván a zpravidla instalován jako součást programového balíku Microsoft - Office současně např. s velmi rozšířeným textovým editorem Word, má ho mnoho esperantistů již nainstalovaný ve svém počítači, i když ho třeba ještě nepoužívá. Mimoto programový balík Office je prodáván na celém světě uzpůsobený pro příslušné národní podmínky, resp. je

možné jej na příslušné národní podmínky nastavit. Excel je dále vyvíjen, jeho funkce jsou dále rozšiřovány a zlepšovány. Existuje také prohlížeč verze programu Excel volně bezplatně nainstalovatelná z internetu, ovšem jen pro pasivní využití.

Závěry:

1. Esperanto potřebuje, aby byly vytvořeny oficiální velké slovníky výkladové i překladové z esperanta do národních jazyků a z národních jazyků do esperanta.
2. Využitím možnosti vytvářet elektronické slovníky lze překonat hlavní dosavadní problémy, pro které dosud nebyly velké esperantské slovníky dosud vytvořeny.
3. Využitím vhodného programu lze vytvořit podmínky pro vypracování jedinečného mezinárodně provázaného systému velkých slovníků.
4. Pro všeobecnou rozšířenost a technické možnosti doporučuji pro tuto práci použít tabulkový procesor Microsoft-Excel.