

[JTP \[http://www.jtpunion.org/spip\]](http://www.jtpunion.org/spip) > 5. Články a zajímavosti [rubrique.php3?id\_rubrique=5] > [Recenze SW pro překladatele \[rubrique.php3?id\\_rubrique=61\]](#)

## Technologie: Užitečný blbec

*Živobytí překladatelů ohrozí až vynález umělé inteligence.*

Instituce Evropská unie utratí ročně zhruba 1,1 miliardy eur za překlady a tlumočení. V rámci unie totiž aktuálně existuje 506 možných jazykových kombinací a s každým případným rozšířením se tato rodina o další desítky kombinací rozroste. Kolem osmi set překladatelů a tlumočnicků, kteří najdou ročně práci při převodu dokumentů a jednání do mateřských jazyků členských zemí, se zatím o svou obživu obávat nemusejí. Jedním z principů unie je totiž mimo jiné důraz na udržení národní identity přinejmenším na úrovni jazykové.

### Tajné služby a spol.

Překladatele zatím neohrožují ani výsledky úsilí o dokonalý strojový překlad. „Hrozivý“ optimismus amerických tajných služeb z poloviny minulého století, že docílení plně funkčního řešení bez potřeby lidské revize je otázkou několika málo let, vzal dávno za své.

Počátkem studené války americká vláda investovala do strojových překladů nemalé částky, aby CIA a dalším zpravodajským službám ulehčila sledování Ruska a jeho satelitů. Povzbuzením byly úspěchy v luštění kódovaných zpráv během 2. světové války, vznik informační teorie amerického elektronického inženýra a matematika Clauda Shannona a spekulace o univerzálních principech základů přirozených jazyků. Na mnoha amerických univerzitách byly zahájeny výzkumy možnosti strojového překladu a již roku 1954 veřejně představila první úspěchy společnost IBM spolu s univerzitou v Georgetownu.

Rané systémy spočívaly prvotně v rozsáhlých dvoujazyčných slovnících, které nabízely k výrazům ve zdrojovém jazyce jeden či více ekvivalentů v jazyce cílovém, a v základních pravidlech pro správný slovosled výsledného výstupu. Brzy se však projevila omezení tohoto systému a začalo být zřejmé, že jazyk je natolik složitým a komplexním mechanismem, že vyžaduje systematictější metody, jak analyzovat větnou stavbu. Vědci však postupně naráželi na sémantické bariéry, znemožňující výraznější průlom, a tak optimismus upadal. V roce 1964 byl založen Poradní výbor pro automatické zpracování jazyků (Automatic Language Processing Advisory Committee), který o dva roky později dospěl ke „slavnému“ závěru: strojový překlad je pomalejší, méně přesný a dvakrát nákladnější než práce lidského překladatele a vznik použitelného strojového překladače není v nejbližší době uskutečnitelný. Zdroj peněz začal vysychat.

V 70. letech začaly strojový překlad poptávat nové subjekty, protože administrativní a komerční nároky multijazykových komunit a mezinárodního obchodu zvýšily poptávku po překladech v Evropě, Kanadě a Japonsku nad kapacitu zavedených překladatelských služeb. Cílem se staly stroji podporované překladové systémy, jež by proces překládání zlevnily. Automatické systémy pro překlad se poté začaly specializovat na úzce zaměřené oblasti – technické manuály, zdravotnické formuláře a podobně – nebo sloužily jako levný a rychlý prostředek hrubého překladu pro potřebu nadnárodních institucí, organizací a společností. Příkladem budiž systém Systran, využívaný Evropským společenstvím, NATO či firmou General Motors. Pro stolní počítače vznikaly první jednoduché nástroje CAT (Computer-aided Translation) dostupné širší překladatelské veřejnosti, které pracují na základě vytváření databáze z překladů vytvořených konkrétním překladatelem.

### Hybridní model

Práci překladatelů zjednodušila možnost fulltextového vyhledávání napříč internetem a již v roce 1997 spustila AltaVista svůj Babelfish – první internetovou službu poskytující zdarma a v reálném čase strojový překlad obsahu internetových stránek. Momentálně je nejdále vyhledávač Google. Jeho překladač je podobně jako většina podobných nástrojů založen pouze na statistické metodě, tedy pravděpodobnosti. Zjednodušeně řečeno, čím více vět obdobného charakteru obsahuje databáze, tím pravděpodobnější je správnost výsledného ekvivalentu. Přestože má Google přístup k ohromnému množství dat díky bilingvním korporátním i institucionálním webům, která dokáže rychle nasávat a vytvářet na jejich základě algoritmy, použití výsledného textu je stále omezeno

spíše na základní orientaci v překládaném textu a jeho využití jako finálního překladu je zatím v nedohlednu.

Možnou cestu do budoucna však představují takzvané hybridní modely, na nichž je postaven například evropský projekt EuroMatrix. Jeho druhá vývojová fáze s názvem EuroMatrixPlus byla zahájena letos na jaře v Praze. Účastníky projektu jsou totiž mimo jiné Ústav formální a aplikované lingvistiky Matematicko-fyzikální fakulty Univerzity Karlovy a česká překladatelská agentura CEET. Statistika sice hraje v EuroMatrixPlus významnou roli, nicméně je doplněná o lingvistická pravidla. Výsledek by tedy měl být kvalitnější. „Zatím to tak ale není, Google má nepoměrně obsáhlejší databázi dat a statistika díky tomu zatím lingvistiku hravě nahradí,“ popisuje dosavadní vývoj ředitel společnosti CEET David Matuška.

Jaký by měl být ideální výstup celého projektu, který je zaměřen na překlady v rámci agentury Evropské unie? Na překládaný text se nejprve nasadí překladový nástroj CAT. Ten určité segmenty textu přeloží ve stoprocentní shodě s překladovou pamětí a na zbylý text se nasadí strojový překladač. Poté nastupuje lidská korektura, případně překladatel dopřeloží potřebné „zbytky“. Výsledek se „nalije“ zpátky do databáze CAT a poslouží i ke zdokonalení nástroje pro strojový překlad. Hlavní výhoda? Překlad se jednak zrychlí, podstatná je ale i potenciální úspora nákladů, která by mohla činit 20 až 70 procent.

### Počítání versus hraní

Již v roce 2001 francouzský překladatel a softwarový vývojář Yves Champollion předpověděl, že budoucnost lidského překladu bude spočívat v korekci překladu vyrobeného počítačem. „I když řídím překladatelskou firmu, vidím budoucnost v tom, že se překladatelé postupně změní v editory,“ souhlasí Matuška. Přesto existují oblasti, ve kterých budou jazykové schopnosti překladatelů od počátku nenahraditelné. Beletrie, marketingové texty, ale i publicistika.

Champollion například porovnává překlad se třemi základními způsoby, jakými lidé počítají:

- ▶ **Počítání:** když máme spočítat  $145 + 133$ , rozdělíme operaci do menších, spočítáme a máme výsledek.
- ▶ **Paměť:** když máme spočítat  $8 \times 5$ , vybavíme si tabulku násobení a odpovíme okamžitě.
- ▶ **Selský rozum:** když máme říci, zda je  $1\,450\,000 \times 3789$  více nebo méně než jedna, odpovíme více, aniž bychom počítali (počítač odpoví stejně, ale nejprve provede početní operaci).

Šachový mág Gari Kasparov by jistě potvrdil, že počítače zvládají poměrně rychle první a druhou metodu. Třetí metodu bychom mohli dle Champolliona považovat za „pěknou a poetickou, nicméně neefektivní“. „To je ale zásadní chyba. Všichni seriózní IT inženýři, kteří se vážně zabývají budoucností počítačových možností, pracují právě na této třetí metodě,“ dodává Champollion.

Zatímco počítače šachy počítají, Kasparov je hraje. Zatímco počítače nikdy „nepochopí“ význam, mohou do určité míry překládat. Protože ale překlad bez pochopení významu nedává smysl, bez lidské revize se stroje nikdy neobjedou. Minimálně do té doby, než bude vynalezena umělá inteligence.

*David Kasl*

týdeník EURO 47/2009