

***Již viděné*** – tak zní otrocký překlad francouzského *déja vu*. Že se tento výraz ocitl přímo v názvu překladatelského nástroje, který vám nyní představíme, není náhoda. Produkt totiž umožňuje automaticky vyvolat překlad textových sentencí, které už byly přeloženy někdy dříve. Jak po nějaké době jeho používání stoupne produktivita překladu, snad není třeba zdůrazňovat.

## Překladatel potřebuje paměť...

*Computer Aided Translation*, zkráceně CAT, je snad jedno z posledních odvětví computer science, které je na našem teritoriu známo spíše z odborných zahraničních časopisů nežli z každodenní praxe. “Zásahu” na tom mají také dovozci, hnaní vidinou rychlých a co možno největších zisků. Kdejaký málo užitečný či naprosto neužitečný software se dováží, i když objem prodeje je posléze nevalný, na nástroje CAT však zrak dovozců zatím nepadl.

A přitom je – v současném stadiu našeho snažení o vstup do Evropy – potenciál zdejšího trhu nezanedbatelný. Státisíce slov všech možných materiálů EU, které je zapotřebí přeložit před vstupem, a potom desetitisíce, které se budou měsíc co měsíc překládat, při používání pouhého textového editoru určitě kvalitně nepřeloží ani celé týmy překladatelských “Einsteinů”.

Druhým momentem, který mluví pro urychlené rozšíření těchto produktů v našich končinách, je globalizace. Zde lze již hovořit o informačním toku obousměrném. Jak při dovozu, tak při vývozu je nezbytně nutné, aby pokud možno maximum průvodní dokumentace, se kterou přichází uživatel do styku, bylo v jeho rodném jazyce. A nesmí přitom jít o překlad podprůměrný, který pak kupujícího spíše odradí.

### Jak může CAT pomoci?

CAT přináší do překladatelského oboru hned několik důležitých funkcí:

**Stejnou** překladovou jednotku (nejčastěji větu) překládá uživatel nebo celá skupina uživatelů, kteří pracují v síti nad společnou *překladovou paměť*, pouze jedinkrát. Od okamžiku odeslání hotového překladu této jednotky do překladové paměti se při příštím výskytu ve výchozím textu buď automaticky nabídne, nebo – použijeme-li funkce předzpracování překladu – rovnou se do textu dosadí, takže překladatel zpracovává jen nové jednotky.

Z překladové paměti se vyhledávají a nabízejí překladateli k využití i **podobné** překladové jednotky. Nástrojem je fuzzy porovnávání s nastavitelnou hladinou podobnosti.

Obě tyto funkce mají společný vliv na překladatelskou práci – zvyšují produktivitu. Zejména v oblasti technických textů, které jsou sice odborně náročné, ale slohově rozhodně netrpí “jiráskovským” stylem, je jistá opakovatelnost takřka pravidlem.

Další pomůckou, jejíž výhoda vynikne zejména při práci v síti, je *terminologická databáze*. Přínos je zřejmý: zvýšení kvality překladu. Použitelnost této funkce zahrnuje veškeré odborné texty, od technických až po právníkou literaturu. Právě nehomogenní překlad dokumentace, ať je to “odkalovací ventil” v první kapitole, “odvzdušňovací” ve třetí, nebo “vypouštěcí” v poslední (stále pro tutéž součástku), může vést až k naprostému zmatení uživatele výrobku – přitom samostatně každý z těchto překladů je v podstatě správný. Pokud se paré dokumentace rozdělí mezi více překladatelů, bývá tento jev tou nejčastější vadou.

Stejně chyby nacházíme i v překladech právnických dokumentů, zejména zákonů, které agentury klidně rozdělí mezi pět a více překladatelů, jen aby splnily dodací lhůtu a nemusely platit překladatelům za spěšnost. Jediným nástrojem, který zde může zajistit potřebnou homogenitu, je důsledné používání terminologické databáze. Přitom musí tato databáze pokud možno “napovídat” sama, aniž by se jí musel překladatel vědomě dotazovat.

Stejně jako jiné systémy IT musí se i CAT systém v přechodném období svého nasazování nějak vyrovnat s dřívějšími znalostmi. Proto všechny dobré produkty tohoto typu nabízejí funkci začlenění *historických překladů*, tj. překladů, u nichž je předem znám text výchozí i cílový, do překladové paměti.

## Déja Vu Interactive

Déja Vu je (doufejme) prvním z CAT nástrojů, které čtenářům postupně představíme. Produkt není lokalizován; zůstaneme proto někde u anglických názvů funkcí, neboť pouhý překlad by mohl být zavádějící. Na úvod několik definic:

**Projekt:** Několik souborů pro téhož odběratele nebo majících společný terminologický základ. Musí být v jednom výchozím formátu (DOC, RTF, HLP atd.). Jednu překladovou paměť může sdílet více projektů. V rámci jednoho projektu může být více cílových jazyků. Do projektu se importují soubory výchozího jazyka. Tím je zajištěna ochrana výchozího souboru v každém okamžiku práce na projektu.

**Předmět:** Klasifikace překládaného tématu. Lze ji použít buď předdefinovanou, nebo si vytvořit klasifikaci vlastní.

**Klient:** Odběratel. Společně s předchozím pojmem umožňuje rozlišovat výrazy cílového jazyka buď oborově, nebo podle odběratelů.

### **Překládaný segment:**

Úsek textu mezi oddělovači (nejčastěji věta nebo její část).

## Základní kameny

Jako vospělý produkt nabízí Déj Vu samozřejmě všechny činnosti zmiňované v úvodní části článku. **Překladová paměť** (*Memory Database = MD*) je soubor, který obsahuje dvojice vět výchozích a cílových. Každá věta má příznak předmětu a klienta. Přeložené věty uživatel odesílá do MD buď po jedné, s možností odeslat celý soubor po provedené korektuře (jak pravopisné, tak i stylistické), nebo

zcela automaticky (nutno nastavit ve volbách) při přechodu na následující segment. Mimo dobu překladu je možné MD obhospodařovat programem údržby (*Memory Maintenance*). Tento program zajišťuje i spojení (*Align*) historických překladů nacházejících se v samostatných souborech, ať již pořízených v textovém editoru, nebo naskenovaných. Pořízený soubor lze k jiným účelům i exportovat.

**Terminologická databáze** (*Terminology Database = TD*) je soubor, kam se ukládají dvojice výrazů – jedno slovo až celá věta. Každá dvojice vedle příznaku předmětu a klienta může obsahovat i gramatické informace (rod, slovní druh). O údržbu TD se stará *Terminology Maintenance*. Lze jím importovat v několika formátech (TXT, XLS, MDB a dalších) slovníkové soubory, které překladatel pořídil za dobu práce s textovým editorem. Ve výše uvedených formátech lze i exportovat. Samozřejmě lze do TD vkládat i v průběhu překladu (zajištění homogenity překladu názvů voleb, funkcí ap. při lokalizaci produktu).

## Funkce

**Rozšíření překladu** (*Propagate*) rozšíří již hotový překlad věty buď do daného souboru, nebo do celého projektu. Slouží k zrychlení překladu, neboť shodné věty se přeloží na všech místech svého výskytu.

**Předzpracování překladu** (*Pretranslate*) před započítím překládání souboru dosazuje z překladové paměti již hotové překlady. Šetří čas v případě, kdy přijdeme k rozpracovanému projektu s dalšími soubory (je to stejné, jako když v překladu systematicky používáme postupně funkci *Propagate*). V rámci této fáze je možné do cílového textu kopírovat buď jen překlady identických výchozích textů, nebo i texty splňující nastavitelné kritérium podobnosti (fuzzy porovnávání).

**Sestavení cílového segmentu** (*Assemble*) nastupuje v okamžiku, kdy není k dispozici již hotový překlad celého segmentu nebo jeho části, na něž by zabralo fuzzy hledání. Přitom v TD jsou k dispozici části (sousloví) a zbytek segmentu obsahuje více slov, která se nebudou překládat (např. v návodech k použití je to identifikace součástí apod.). Za této situace funkce přináší zrychlení. V běžných textech čistě lexikálních není vhodná, neboť provádí vlastně jen substituci.

**Automatické vyhledávání** (*AutoSearch*) je funkce, která s narůstajícím počtem přeložených segmentů stále zefektivňuje práci. Zapíná se ve volbách. Při přechodu na následující segment automaticky nabízí nalezené položky, a to jak z MD, tak z TD.

Ale pozor – s rostoucí dimenzí obou databází postupně rostou i nároky na výkonnost počítače! Na Pentiu 100 je použití funkce skoro vyloučeno. Na Pentiu II 233 se překládaly texty do 4000 slov (MD) s celkem snesitelnou odezvou. Jakmile však byl do TD importován slovník o rozsahu přes 15 000 dvojic v obou jazycích, PII/233 začalo mít citelné prodlevy (u věty 23 slov dlouhé přes 16 sekund) a pracovní pohodu navrátil teprve přechod na Celeron 400 MHz se 128 MB paměti.

Další podmínkou efektivního využití této funkce je velikost obrazovky. Má-li totiž přinášet zrychlení práce, musí být okna viditelná současně, bez jakéhokoli přepínání z lišty aplikací. Obrazovka by tedy měla mít minimálně 17palcovou úhlopříčku, optimální by byla 19palcová.

Další funkce, *Learn*, **vyhledá v MD** sousloví přeložené již dříve a u CAT programů je zatím ojedinelá. Ostatní produkty vyžadují, aby sousloví bylo v TD.

Funkce **Najdi/nahrad'** (*Find/Search*), běžná v každém textovém editoru, má zde jednu zvláštnost, která u ostatních produktů nebývá vždy k dispozici – hledání ve výchozím jazyce a opravy v jazyce

cílovém. Je velmi vhodná, jestliže po přeložení několika souborů dodatečně zjistíme, že určitý odborný termín má odlišný překlad, a potřebujeme realizovat tuto opravu v celém projektu. Lze to provést i v případě, že původní překlady nebyly stejné.

**Oddělovače, definice pravidel a výjimek** slouží k dělení výchozího textu. Nechceme-li např. po importu text procházet a znovu spojovat věty rozdělené tečkou za pořadovým číslem uprostřed věty, definujeme výjimku “nedělit, pokud tečce předchází číslice”.

V těchto výrazech lze jako proměnné použít symboly pro libovolnou číslici či písmeno, pro malé/velké písmeno nebo libovolný znak.

**Spojení/rozdělení překládaného segmentu** – pokud se ve výchozím textu vyskytne oddělovač (např. tečka) na jiném místě, než je podchyceno pravidly a výjimkami z předchozího odstavce, lze spojit dva a více segmentů v jeden.

**Obrácení databáze** je funkce nesmírně užitečná při překladu oběma směry. Při opačném směru překladu pak máme k dispozici terminologii původně pocházející z výchozího jazyka, a lze tedy předpokládat, že by měla být správná.

**AutoText** – podobně jako v MS Office je k dispozici funkce umožňující náhradu zkratky rozvinutým tvarem. Zde je však sada zkratk vázána na jeden jazyk, takže nemůže docházet ke kolizím, které někdy známe z Wordu.

## Uživatelské rozhraní

Uživatelské rozhraní je velmi flexibilní. Svůj oblíbený způsob práce zde najdou jak ctitelé myši (tlačítek), tak rozbalovacích nabídek či klávesových zkratk.

Při začátku práce na projektu nejprve sloučíme historické výchozí a cílové soubory (pokud existují) a založíme překladovou paměť. Potom importujeme soubory, které se mají překládat – k dispozici jsou formáty Word, RTF, Help, PowerPoint, FrameMaker, PageMaker, QuarkXpress, Interleaf, Java Properties, HTML, SGML, C/C++, IBM (tm), Trados Workbench.

Po skončení práce se hotový překlad exportuje zpět do původního formátu včetně veškerých řídicích znaků. K dispozici je ještě jedna funkce, která však pro češtinu jako cílový jazyk příliš nepomůže. Jedná se o automatickou kontrolu konzistence překladu – pro položky z TD lze zkontrolovat, zda v celém projektu je dané sousloví přeloženo shodně. Bohužel pro češtinu s její bohatostí pádových koncovek je tato funkce jen částečným přínosem.

Zcela nezávislou součástí, která je vlastně bonusem i pro uživatele jiných editorů, je program **TermWatch**, který umožňuje používat terminologické databáze i při práci v jiném programu – zkoušen byl Word a Excel. Dlouhou databázi (skoro 4 MB) k mému překvapení TermWatch prohledával dosti rychle.

## Podporované jazyky

Jako výchozí i cílové jazyky podporuje Déjř Vu celkem přes 130 jazyků, včetně takových exotů, jako je urdština, tatarština nebo maorština. Příslušná písma závisí na operačním systému. Jak produkt podporuje jazyky s netypickým směrem zápisu (zprava doleva, seshora dolů), jsem vzhledem

k "středoevropskému" operačnímu systému vyzkoušet nemohl. Při instalaci lze vybrat kontrolu pravopisu pro jeden z osmi jazyků; čeština, bohužel, mezi nimi zatím není.

## **Závěr**

Déjř Vu lze rozhodně doporučit překladatelům textů, kde se vyskytují opakující se sekvence nebo kde záleží na homogenitě překladu, např. při lokalizacích dokumentace veškerých výrobků. Poměrně vysoká pořizovací cena (na webu je uváděno 990 EUR za jednu licenci, což v době psaní článku představovalo více než 36 000 Kč) se však překladateli nebo skupině překladatelů vrátí jen při intenzivním využívání. Není to však chyba výrobce; uvedená cena je u produktů této kategorie naprosto obvyklá. Příčinou je spíše nedostatečné ocenění práce odborných překladatelů, které u nás představuje zlomek evropské úrovně.

Důležitou vlastností produktu je možnost osahat si skutečně plně funkční produkt. Na adrese [www.atril.com](http://www.atril.com) je možné si produkt stáhnout. Po jeho nainstalování se vypíše klíč produktu, který zašleme firmě, a zcela zdarma dostaneme registrační klíč, který zajišťuje na jeden měsíc plnou registraci – tedy i podporu. U takto komplexních produktů je to jediný způsob, jak zjistit, zda právě tento produkt vyhovuje naší překladatelské letoře a typu textů, které nejčastěji překládáme. Profesionálnímu překladateli, kterému za měsíc prochází počítačem několik desítek tisíc slov, tato doba na posouzení bohatě postačuje.

*Miroslav Herold*