10 Praktické využití a kritéria pro hodnocení systémů strojového překladu

Pro světové jazyky existuje bohatý výběr systémů nejrůznějších parametrů, jakými jsou např. koncepce, kvalita, určení, optimální využití, požadavky na hardware, požadavky na obsluhu, cena.
Za základní lze považovat rozdělení systémů podle míry účasti člověka, od zcela automatických systémů (ideál: FAHQT – Fully Automatic High Quality Translation) až po systémy s různou mírou účasti člověka (HAMT – Human-Aided Machine Translation nebo MAHT – Machine-Aided Human Translation).

Yehoshua Bar-Hillel (1958): FAHQT není v principu možný. Příklad: *The box was in the pen.* Důvody:

(a) Je těžké si představit, že se někdy podaří zabudovat do systému strojového překladu znalosti o světě, potřebné k rozlišení dvou významů slova *pen*.
(b) Znalosti tohoto typu jsou pro FAHQT nezbytné.

Žádné z obou tvrzení však neplatí absolutně, neboť:

(a) Existují pokusy o reprezentaci a využití znalostí o světě v MT ('umělá inteligence').
(b) 'Porozumění' na takové úrovni možná není pro překlad vždy nutné.

Panuje však shoda v tom, že alespoň v současnosti je možný buď plně automatický překlad, nebo velmi kvalitní překlad, ale ne obojí najednou. V praxi je tedy třeba hledat kompromisní řešení.

10.1 MAHT

postup překladu zůstává v rukou člověka

obecné nástroje pro práci s textem , týká se jen cílového jazyka: textový editor, kontrola pravopisu (neumí odhalit chyby ve shodě), kontrola syntaxe (shoda, opakování slov, věty bez slovesa v určitém tvaru), kontrola stylu (klišé, věty začínající spojkami, končící předložkami, věty příliš krátké nebo dlouhé)

příručky v elektronické podobě: glosáře, výkladové, dvojjazyčné, synonymické slovníky, tezaury, encyklopedie

překladatelská stanice: okna pro zdrojový a cílový text, příručky, existující překlady (překladová paměť, např. Deja Vu, Trados); vstup zdrojového textu elektronicky, přes skener; automatická tvorba glosářů k danému tématu, využití existujících (firemních) glosářů; automatické vyhledávání odborných termínů, manuální nebo automatické vkládání nalezených ekvivalentů – 'překladový polotovar'

10.2 HAMT

postup překladu je řízen počítačem, člověk zasahuje před vlastním překladem (preeditace), v jeho průběhu (interakce) a po skončení překladu (posteditace)

10.2.1 Preeditace

– vyhledávání možných problémů z hlediska systému ve vstupním textu a pokusy o jejich odstranění.

- Preeditaci může provádět člověk na vstupním textu jen na základě znalosti omezení systému, nebo sytém může sám upozorňovat na problematické jevy a pasáže.
- Typické operace: identifikace názvů, označení slovního druhu u víceznačných slov, vymezení vedlejších vět, uzávorkování koordinací, nahrazení neznámých slov. (SUSY)
- Preeditace v extrémní podobě: regulovaný jazyk – autor textu se musí držet omezeného repertoáru syntaktických konstrukcí, používat omezený slovník s přesně definovanými významy a snažit se o co nejkratší a nejjednodušší věty (Systran ve firmě Xerox). Omezený jazyk je něco jiného než jazyk určitého oboru.

10.2.2 Posteditace

– úprava výstupu na dohodnutou úroveň kvality, odlišnou pro účely povšechné informace a pro účely publikační.

- Ve starších systémech se posteditace prováděla červenou tužkou, později textovým editorem, ale systém sám neupozorňoval na možné problémy, chyby bylo nutné opravit přepisováním. Základním předpokladem i zde je však možnost současného zobrazení vstupu i výstupu a možnost nahradit jedno slovo jiným v celém textu.
- Pro časté opravy je možné zavést makroinstrukce (např. změny slovosledu, slovesného vidu).
- Systém může upozornit na pasáže, které mohou být špatně, neboť se nepodařilo vyřešit víceznačnost, najít ve slovníku nějaký výraz nebo analyzovat nějakou konstrukci. Může také umět opravit všechny výskyty chyby podobného typu v celém textu.
- Lingvisticky inteligentní textové editory by měly umět upozornit na strukturní víceznačnosti a nabídnout alternativní konstrukce, upravit všechny výrazy, které vykazují mluvnickou shodu s nahrazeným slovem, apod.

10.2.3 Interakce

– systém žádá o spolupráci v průběhu překladu (řešení víceznačností, neznámá slova).

- obsluha musí znát oba jazyky
- obsluha zná jen cílový jazyk, pak může pomoci jen při stylistických úpravách v syntéze, případně může vyloučit interpretace v kontextu nepravděpodobné
- obsluha zná jen jazyk vstupu; je-li vstupní text již hotový, systém se pokouší odstranit dotazy víceznačnosti (též překladové typu *wall*); systém však může spolupracovat přímo s autorem textu ve chvíli jeho vytváření. Některé systémy pak pracují s předem definovaným repertoárem frází s 'proměnnými' (obchodní dopisy). Sestavení textu může probíhat tak, že uživatel odpovídá na dotazy systému, nebo tak, že se systém pokouší z repertoáru frází sestavit parafrázi vstupního textu.

10.3 Systémy pro překlad typově omezených textů

- Omezení může být dáno oborem a žánrem textu: meteorologické zprávy, softwarové manuály, abstrakta literatury o hydraulických strojích, apod.
- Někdy nejde o omezení, ale o jazyk jiného druhu, jak syntakticky (lékařské anamnézy *The patient presented with a cold.*), tak odlišným užitím běžných slov (*lavička* ve stavebnictví). Někdy je orientace na texty určitého typu nikoli zjednodušením úkolu, ale komplikací: právnické texty, patenty.
- Rozdíly mezi jazyky: zápisy ze schůzí v minulém (A,Č) nebo přítomném (F) čase, interpunkce (právnická A nerada čárky), preference určitých konstrukcí (určitý člen se v A manuálech vynechává, infinitiv místo imperativu ve F manuálech).
- Omezení na určitý typ textů lze kombinovat s regulovaným jazykem.

10.4 Využití méně kvalitního strojového překladu

Strojový překlad bez lidského zásahu je obvykle poměrně laciný a může dobře posloužit pro účely povšechné informace

- odborníkům zběhlým v dané problematice,
- těm, kdo znají základy jazyka vstupu,
- těm, kdo mají rozhodnout, zda text stojí za to přeložit pořádně, zda vydat peníze na nákup knížky apod.,
- překladatelům jako polotovar.

10.5 Hodnocení systémů strojového překladu

Neexistuje všeobecně uznávaná metoda. Hodnocení většinou provádějí na zakázku lidé vzdálení od problematiky MT a výsledky jsou důvěrné. Tvůrci systémů nejsou objektivní, mají zájem na bezchybné demonstraci systému a skrývají nedostatky pečlivě připravenými podmínkami.

10.5.1 Hodnocení v jednotlivých stádiích vývoje systému

Vývoj prototypu: přinesla dílčí změna či vylepšení zamýšlený efekt? jsou jednotlivé procesy a data ve vzájemném souladu?; hodnocení prototypu jako celku bez přísných požadavků na rychlost a uživatelský komfort

Vývoj prakticky použitelného systému: podařilo se vyvinout systém odpovídající původním představám?; nejen lingvistické, ale i provozní parametry s ohledem na budoucí uživatele, komerční perspektivy

Hodnocení systému potenciálním uživatelem: odpovídá výkon systému parametrům, které jeho tvůrce udává? je systém použitelný a ekonomicky přínosný v daném prostředí? (kolik lidské práce a jak kvalifikované je třeba k dosažení požadované úrovně překladu, jaký hardware je třeba, nakolik je systém slučitelný s existujícím zařízením a metodami, jak snadno se dá systém vylepšovat) Důležitý je názor překladatelů, kteří budou se systémem spolupracovat, a konečných uživatelů výsledku.

10.5.2 Lingvistické hodnocení nezkorigovaného výstupu

– hodnocení jádra systému. Podle možnosti brát v úvahu vnitřní strukturu systému lze rozdělit na 'glass-box evaluation' a 'black-box evaluation'. Podle rigoróznosti metody na obecné hodnocení kvality a analýzu chyb.

Hodnocení kvality:

- věrnost originálu, přesnost, zachování významu
  testy: obsahuje originál nějakou informaci navíc? proveď to, co říká překlad! zpětný překlad
- srozumitelnost, jasné vyjádření významu, jak snadno se to čte
  testy: 'známkování'; jak dlouhé jsou věty, jak moc složitých nominalizací; doplň zakryté slovo!
- styl odpovídající obsahu originálu
  testy: jen subjektivní hodnocení, známkování

Chybová analýza:

- počet (procento) chyb, t.j. počet zásahů do textu nutných k tomu, aby se dospělo k přijatelnému překladu (přidání, vymazání, záměna slova, každá změna pořadí u dvou slov) – málo objektivní
- klasifikace chyb podle jazykových jevů a relativní složitosti úpravy systému (zásahy do slovníku, do gramatiky, do celkové koncepce)
- standardní soubory příkladů ('benchmark tests') – zatím neexistují

10.5.3 Hodnocení během tvorby prototypu

– t.j. během výzkumných prací. Odhalování chyb v systému pomocí podrobného sledování 'výpočtu' s mezivýsledky.

- programátorské chyby
- lingvistické chyby: opomenutí, neúplnost, chyba ve slovníku nebo gramatice ve fázi analýzy, transferu nebo syntézy

Rozsáhlejší testy celého systému mají za cíl odhalit chyby, které vyžadují zásah do jádra systému.

10.5.4 Hodnocení během vývoje praktického systému

Navíc se sledují různé obslužné funkce, jako vstup textu, doplňování slovníku, redakce výstupu. Cílem je zjistit, co systém zvládne a na co už nestačí. S tím souvisí specifikace podmínek, za nichž je systém schopen dávat uspokojivé výsledky. V testech se systému stále znovu předkládají rozsáhlejší soubory textů. Nejdříve se opravují ty chyby systému, které vyžadují nejméně práce. Pokud je vynaložená námaha neúměrná výsledku, je třeba uvažovat o důstojném pohřbu.

10.5.5 Hodnocení potenciálním uživatelem

Někdy uživatel při vývoji spolupracuje, nebo je dokonce systém určen přímo pro něj. Ale většinou je pro něj systém černá skříňka. Uživatel zvažuje využití systému v širším kontextu: vstup textu, úprava textu, (pre-editace, doplňování slovníku,) překlad, (posteditace,) formátování, výstup (elektronicky, tisk, publikace). Uživatel si musí být vědom toho, co vlastně potřebuje: jaký typ textu chce překládat, z jakého oboru, pro jakého čtenáře a k jakému účelu je výstup určen, jaké objemy textů se budou zpracovávat, jak moc se bude na překlad pospíchat?

Lingvistické hodnocení:

- lexikální problémy mohou nastat, je-li vstup oborově neomezený, je-li v něm hodně složitých nominálních spojení
- syntaktické problémy může vyvolat výskyt konstrukcí, s nimiž gramatika nepočítá

Proto je vhodné zadat k překladu reprezentativní vzorek a analyzovat výsledek. Problém je v tom, že uživatel není většinou schopen odlišit chyby, které vyžadují nepatrný zásah do systému, od těch, které vyžadují náročnou přestavbu. Uživatel si často také není vědom toho, že něco neumí vyřešit žádný systém, nebo že některé chyby je mnohem snadnější opravit na výstupu, než se je pokoušet řešit v systému (zájmenné odkazy). Na základě pečlivě vybraných příkladů by si však uživatel mohl udělat představu o výkonnosti a omezeních systému.

Další aspekty:

- rozšiřitelnost
  - doplňování slovníku
  - rozšíření na další typy textů
  - rozšíření o další jazyky
- nároky na hardware
- omezení vlastního systému (velikost slovníku, rozsah vstupu)
- nároky na personál
- kromě nákladů též přínosy: rychlost, konsistentní terminologie, simultánní výstup ve více jazycích, možnost vývoje