

Memorandum 1949

Warren Weaver

Toto dílo se snažilo podpořit statistickou sémantiku a snažilo se inspirovat ostatní, aby se jejímu studiu věnovali. Autorem díla je Warren Weaver, který se o své úvahy dělil s kybernetikem Norbertem Wienerem. Sám autor v úvodu píše: *Toto dílo, které je o překladu z jedné řeči do druhé a o možnosti, že se na něm budou podílet počítače, bylo napsáno s jedinou nadějí: že toto dílo inspiruje někoho, aby se tím zabýval.*

Byl jsem nervózní, že mé ideje jsou naivní. Ale protože mi toto téma přijde velmi důležité, jsem ochoten vystavit svou ignoranci na odiv doufajíc, že má snaha tuto ignoranci zastíní.

V memorandu jsou zmíněny čtyři hlavní problémy, kterými se Weaver zabýval. Těmi problémy je mnohoznačnost slov, kryptografická metoda, existence “základního jazyka” a logika jazyka.

Kryptografická metoda

Vše začalo válečnou příhodou, kterou vyprávěl Weaverovi jeden známý. Jeho kolega prý rozluštil kód, který byl vytvořen v turečtině. A přitom ten kolega turecky neuměl a dokonce ani nevěděl, v jakém jazyce je kód vytvořen. Tato příhoda přivedla Weavera k myšlence, že kryptografie je nezávislá na jazyce, ze kterého překládá. Z toho vyvodil, že existuje něco společného pro všechny jazyky.

Weaver došel k přesvědčení, že je možné, že jazyk je jen kryptografie. Podrobně to rozvádí v dopise, který napsal kryptografovi Norbertu Wienerovi. Píše: *Přemýšlel jsi někdy o tom, že jazyk je vlastně kryptografie? Že text v ruštině je vlastně napsán v angličtině, jen je zakódován zvláštními symboly? (Tím naráží na azbuku.) Kdybychom měli nějaký odborný text, u kterého nezáleží na slohu a eleganci, mohl by být přeložen jen nějakým počítačem.*

Wiener jeho nadšení zmírňuje a upozorňuje ho na různé problémy. Například v angličtině slovo get má mnoho forem a my je musíme rozlišovat.

Kryptografickou metodou se zabýval také Claude Shannon. Ten publikoval svou matematickou teorii komunikace, ve které se zmiňuje mim jiné i o kryptografii. Uvažuje takto: *Člověk velmi touží říct, že čínský text je jen zakódovaný v angličtině. A vzhledem k tomu, že máme metody, které rozluští téměř jakýkoliv kryptografický problém, pak nemáme už zároveň i metodu na překlad?*

Avšak kryptografickou metodou nikdo k žádným konkrétním výsledkům zatím nedošel.

Existence “základního jazyka”

Ze všeobecného hlediska je pravděpodobné, že všechny jazyky vznikly podobným vývojem. Přece všechny byly vytvořeny člověkem- stejným živočišným druhem. Weaver uvádí srovnání se stromem. Přece všechny stromy na světě jsou různé a přesto mají stejné kvality (treeness= “stromovství”). Tyto kvality jsou shodné pro stromy v Polsku i na Cejlonu. A co je důležité: i jihoameričan nemá problém s tím, aby řekl, že norský strom je strom.

Je tedy možné, že abychom si všichni rozuměli, musíme “klesnout” ze své řeči na nějakou základní úroveň. Pro lepší představivost přirovnává jazyky k věžím. Bydlíme na vrcholcích věží. S okolními věžmi se domlouváme pouze křikem. Domluva je omezena vzdáleností. Některé věže nás už nemají šanci slyšet. Všechny věže ale mají společný sklep. Abychom se mohli domluvit se všemi obyvateli věží, musíme sejít do samotných základů budovy (= k základům jazyka).

Problém je, že zatím nemáme tušení, jak se k tomu základnímu jazyku dostat.

Mnohoznačnost

Problémem všech jazyků jsou vícevýznamová slova. Ta dělají překladu největší potíže. Touto mnohoznačností se zabývali doktor Andrew Booth a doktor Richens (křesní jsem nenašla). Andrew Booth dokonce napsal v roce 1948 své memorandum, ve kterém se zabýval strojovým překladem. Chtěl, aby stroj zacházel se slovem, které má přeložit, jinak než dosud.

1. Strojový překlad

Navrhoval tento postup: Stroj prve zjistí, jestli jeho paměť obsahuje to slovo, které má přeložit. Pokud ne, oddělá poslední písmeno slova a znovu hledá v paměti. Poslední písmena oddělá tolikrát, dokud se mu vzniklé slovo nenajde v paměti. Poté vezme tu část slova, kterou vypustil, a vyhledá ji ve slovníku přípon. Tam zjistí, co přípona znamená a spojí ji se slovem.

Pro představu slovo running. Stroj oddělá g,n,i,n a dojde ke slovu run. To už má v paměti. Poté vezme -ning a hledá to ve slovníku přípon. Jak stroj pozná, že je ve slově zdvojené n a že ho má z koncovky vypustit, již popsáno nebylo.

Obecně se tento způsob překládání nedá využít u literárních děl, kde je důležitý styl a najdeme zde idiomy a mnohovýznamová slova.

2. Vyřešení mnohoznačnosti

Na strojový překlad navázalo pracování s kontextem. Pokud by správně pracoval strojový překlad, nebyl by problém ho rozšířit. Tedy aby stroj nebral jen slovo, které má přeložit, ale i N slov z obou stran slova. Kdyby měl stroj blízký kontext slova, lépe by se určoval jeho význam. Avšak otázka je: Kolik je N? Víme, že záleží na typu článku, například v odborném textu by stroj nepotřeboval tak veliký kontext, jako například v článku zabývajícím se sociologií. Dalším problémem je, že i kdybychom věděli, jaké N slov musíme brát, dostali bychom se k obrovskému počtu souborů slov/ frází, které by musely být někde uloženy.

Mnohoznačnost ale je do budoucna řešitelná. Využívání nějakého mikrokontextu je pravděpodobné.

Logika jazyka

V řeči jsou určité prvky, které nejdou jednoduše přeložit. Důležitou roli v řeči hrají totiž kromě faktů i emoce a styl promluvy. Lidská řeč není tak logická, jako například jazyk matematiky. Avšak svým způsobem trochu logická je.

Perfektního překladu dosáhnout nelze. Ale proces, při kterém by se dosáhlo překladu, který obsahuje jen pár procent chyb, je možný.

Memorandum 1949 bylo unikátní v tom, že popisovalo metody a cíle strojového překladu v době, kdy lidé ještě netušili, čeho všeho jsou počítače schopny.