

Pokroky matematiky, fyziky a astronomie

I.A. Melčuk

Strojové překládání v SSSR

Pokroky matematiky, fyziky a astronomie, Vol. 5 (1960), No. 2, 212--215

Persistent URL: <http://dml.cz/dmlcz/137057>

Terms of use:

© Jednota českých matematiků a fyziků, 1960

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://project.dml.cz>

ze způsobů této spolupráce má být též výměna časopisů a knih, a to jak učebnic středoškolské matematiky tak i knih vědecké povahy. Vědecká společnost věnovala vedoucím delegací osnovy nových učebnic středoškolské matematiky pro potřeby členstva matematických společností jednotlivých zemí; zároveň přislíbila, že bude soustavně zasílat nové publikace, které vyjdou nákladem Vědecké společnosti.

Rud. Zelinka, Praha

STROJOVÉ PŘEKLÁDÁNÍ V SSSR¹⁾

I. A. MELČUK

Každým rokem přibývá počtu vědeckých a technických publikací v nejrůznějších jazycích. V souvislosti s tím je třeba vynakládat stále více prostředků i času na to, aby se dostaly vědcům a technikům do rukou překlady těchto publikací. Odtud plyne snaha zautomatizovat vědecko-technické překládání, a to cestou redukce překládání na řadu určitých logických operací, jež lze vkládat do matematického stroje, schopného řešit známé logické úlohy.

S řešením problémů strojového překládání se v SSSR začalo v roce 1955 v ústavu přesné mechaniky a výpočtové techniky, a v Matematickém ústavu V. A. Stěklova Akademie věd SSSR. Později se do této práce zapojily také Institut jazykovědy Akademie věd SSSR, Leningradská universita, která vybudovala experimentální laboratoř strojového překládání, výpočtářské středisko Akademie věd arménské SSR, Institut elektroniky, automatizace a telemechanizace Akademie věd gruzínské SSR a jiné kolektivy.

Hlavním obsahem práce první etapy bylo sestavení tak zvaných algoritmů pro strojové překládání. Aby stroj mohl přeložit text z jednoho jazyka do jiného, je třeba mu vše potřebné k tomuto úkonu zadat ve formě speciálního souboru pravidel. Tato pravidla musí být přesně formulována, musí být jednoznačná, musí připouštět mechanickou realizaci a musí tvořit systém logicky uzavřený, zahrnující všechny možné případy.

Během let 1955—57 byla sestavena řada algoritmů: francouzsko-ruský, dva anglicko-ruské, čínsko-ruský, německo-ruský, japonsko-ruský, maďarsko-ruský. Francouzsko-ruský a jeden anglicko-ruský algoritmus byly programovány a vyzkoušeny na strojích, to jest podařilo se přeložit úryvek vědeckého (matematického) textu z francouzštiny a angličtiny do ruštiny. Ostatní algoritmy jsou ve stadiu programování.

Práce na strojovém překládání jdou v současné době třemi hlavními směry: zkoumání možných způsobů strojového překládání a volba nejlepších způsobů, vypracování přesných (především matematických) metod popisu jazyka v těsné spolupráci matematiků a jazykovědců, zkoumání vzájemné souvislosti mezi strojovým překladem a jinými praktickými aplikacemi lingvistiky pro zobecnění a plnější využití výsledků, dosažených v přílehlých oblastech.

Jaká je hlavní problematika těchto prací?

Před třemi lety bylo sestavení algoritmů a jejich realizace hlavním úkolem. Dnes jsou algoritmy již sestaveny, věnuje se proto největší pozornost zobec-

¹⁾ И. А. Мельчук, Работы по машинному переводу в СССР, Vestnik AN SSSR, č. 2, 1959.

nění dosažených výsledků. Jako první byl vytyčen úkol vytvořit obecnou teorii sestrojování překladových algoritmů (zejména vypracování universální formy překladového algoritmu jako takového a pravidel pro přizpůsobení této formy konkrétním jazykům). Úspěšné vyřešení tohoto úkolu umožní pak uložit stroji samému, aby sestavoval překladové algoritmy na podkladě předem připravených slovníků a paralelních textů.

Velký význam pro obecnou teorii překladových algoritmů má systém operátorů, vypracovaný v Matematickém ústavě. Překládání se dělí na řadu standardních aktů, které se vyskytují při překládání z kteréhokoli do kteréhokoli jazyka. Takový elementární akt spolu se standardním programem realizace, jemu odpovídajícím, se nazývá operátor.

Každý algoritmus může být vyjádřen ve formě posloupnosti operátorů. Operátorový zápis umožňuje automatizovat velmi pracné programování překladových algoritmů. Tak při programování části maďarsko-ruského textu bylo během pěti minut sestaveno pět programů; k jejich ručnímu napsání by bylo třeba dvaceti až třiceti pracovních dnů (jednoho pracovníka — *čeloveko-dněj*). Operátorový záznam je důležitý také teoreticky; ulehčuje sjednocování algoritmů a vypracování jedné universální formy překladového algoritmu.

Strojové překládání je v současné době pouze první etapou na cestě k řešení obecnějšího a důležitějšího úkolu: jak možno lepším využitím matematických strojů jako nástrojů lidského myšlení dosáhnout toho, aby stroj měl nejširší možnosti pro práci s texty v různých jazycích, aby mohl nejen překládat, ale provádět i redakci textu, referovat, dělat sám bibliografické a jiné odkazy atd. Všechny tyto operace se redukuje na vybrání určité informace z textu a na zápis této informace v nějaké jiné formě. Pro uskutečnění těchto operací je nezbytný speciální „jazyk“, kterým by se zapisovala informace získaná z textu.

Takový jazyk musí jednak zabezpečit jednoduchý a přesný zápis získané informace, jednak musí být vhodný pro to, aby bylo možno do něho překládat texty, napsané reálnými jazyky. Takový jazyk je možno využít jak pro zápis a uchování informace ve stroji²⁾, tak i jako prostředníka při strojovém překládání. Při překládání z mnoha jazyků do mnoha jiných je možné v obou směrech překládat z jazyka původního do jazyka informace, a z tohoto do jazyka druhého. To umožňuje nejen zmenšit počet algoritmů, nezbytných pro přímé překládání z jazyka do jazyka, ale ulehčuje také unifikaci těchto algoritmů. Sestavení jazyka — prostředníka je dnes hlavním problémem většiny prací o strojovém překládání v SSSR.

Z mnoha možností, jak sestavit jazyk-prostředník, se zkoumají dvě, o nichž v dalším promluvíme.

Jednou z možných cest je sestavit jazyk-prostředník jako umělý jazyk, který má svou slovní zásobu, morfologii a skladbu (analogicky běžným jazykům nebo jazykům umělým typu esperanta). Jeho skladba se určí statistickým rozbohem jazyků, z nichž se překládá; bude obsahovat jen ty prvky, které se vyskytují dostatečně často ve všech nebo ve většině těchto konkrétních jazyků, přičemž každému prvku se přisoudí specifická váha, úměrná počtu lidí, kteří tímto jazykem mluví.

Takový jazyk-prostředník by byl jakýmsi „množinovým průnikem“ několika reálných jazyků. Bylo pro něj vypracováno schéma symbolického zápisu. V dalším se předpokládá, že se této formálně logické soustavy použije jako jazyka informace.

²⁾ Jazyk, v němž se zapisuje informace pro uchování ve stroji, nazýváme „jazykem informace“.

Druhá cesta spočívá v tom, pokládat jazyk-prostředník jen za soustavu korespondencí mezi reálnými jazyky. Korespondence se určí ve třech oblastech: lexikální korespondence (korespondence mezi slovy a frazeologickými obraty v různých jazycích), morfologická korespondence a korespondence syntaktická (korespondence mezi elementárními syntaktickými konstrukcemi).

Celky, tvořené slovy různých jazyků, odpovídajících si jedno druhému při překládání (svazky lexikálních korespondencí), pokládají se za slova jazyka-prostředníka; jeho syntaktickými vztahy jsou svazky syntaktických korespondencí, to jest soubory vzájemně si odpovídajících elementárních syntaktických konstrukcí různých jazyků. Morfologie jako zvláštní systém prostředků pro regulární a obligátní vyjádření určitých rozdílů v tomto jazyku-prostředníku nebude. Podstata věci spočívá v tomto: předpokládá se, že jazyk prostředník představuje sám „množinové sjednocení“ všech jazyků, z nichž nebo do nichž se překládá, to znamená, že v něm mohou být vyjádřeny jakékoli rozdíly, vyskytující se ve všech těchto jazycích. Zdůrazněně „mohou“, nikoli „musí“, vyjadřují se nikoli obligátně, nýbrž příležitostně, pokud mají v jazyku prostředníku místo.

Svazky morfologických korespondencí se berou za slova jazyka-prostředníka (zároveň se svazky lexikálních korespondencí). Tato slova se mohou a nemusí vyskytovat. Tak na příklad kategorie počtu podstatných jmen bude vyjádřena v jazyku-prostředníku při překladu z jazyků, ve kterých tato kategorie je (jazyk ruský, anglický, arménský, maďarský ap.), nebude však vyjádřena např. při překládání z čínštiny, kde tato kategorie chybí. Cílem je tu zabránit ztrátě potřebné informace a nutnosti vypracovávat informace nepotřebné — ať jde o překládání v rámci kterékoli dvojice jazyků.

Obligátní vyjádření v jazyku-prostředníku se liší jen ve dvou směrech: slovně (lexikálně) a syntakticky (vztahy mezi slovy), to jest v tom, co se vyskytuje ve všech jazycích lidstva a bez čehož žádný jazyk není myslitelný.

V celku je navrhovaný jazyk-prostředník do jisté míry analogický jednak tak zvaným prajazykům srovnávacího jazykozpytu, které samy rovněž představují soustavu korespondencí mezi jazyky, jednak matematické logice (výrovkovému počtu): „slova“ a „syntaktické vztahy“ jazyka-prostředníka odpovídají abecedě a pravidlům tvoření formálně-logických jazyků.

Jazyk-prostředník jako systém korespondencí může být vytvořen v podobě tabulek s řádky a sloupci. Sloupce se týkají různých jazyků. V každém řádku jsou zapsány „jednotky“ různých jazyků, odpovídající si při překládání. Pořadová čísla řádků (indexy řádků), které obsahují lexikální a morfologické korespondence, představují slova jazyka-prostředníka. Pořadová čísla řádků, které obsahují syntaktické korespondence, tvoří jeho syntaktické vztahy.

Překlad pomocí jazyka-prostředníka se děje ve dvou etapách: analýsa, neboli překlad z daného jazyka do jazyka-prostředníka. Různým jednotkám daného jazyka se přiřazují speciálními pravidly indexy odpovídajících řádků v tabulce jazyka-prostředníka; syntéza, neboli překlad z jazyka-prostředníka do jazyka druhého. Jednotky, vybrané ze svazků korespondencí, se přetvoří a rozestaví v soulase se zákony druhého jazyka.

Dnes je vypracován model jazyka-prostředníka pro nepřilíh dlouhé matematické texty. Sestavují se algoritmy nezávislé analýsy a syntézy pro řadu jazyků. Bylo započato s prací na zhotovení lexikálních korespondencí mezi hlavními evropskými jazyky.

Ústřední částí algoritmu pro strojové překládání pomocí jazyka-prostředníka je syntaktická analýza. Touto analýzou se určí syntaktické vazby mezi všemi slovy překládaného textu. Provádí se to pomocí soupisu elementárních syntaktických konstrukcí (konfigurací), které jsou v textech daného typu a pravidel jejich zjištění v textu. Proto je nezbytné mít pro sestrojování algoritmů dostatečně obsáhlé soupisy konfigurací pro všechny jazyky, určené k překládání.

Existuje mnoho jiných lingvistických problémů, jejichž řešení je pro sestavování algoritmů nezbytné. Tyto problémy je třeba samostatně zkoumat. Patří sem zejména vymezení „zbytných“ prvků v jazyku, historicky vzniklých kategorií, které dnes takřka pozbyly jazykového významu. Tak na příklad zbytečným se ukázal rod ruského slovesa, v mnohém se ukázalo zbytečným skloňování ruských a francouzských přídavných jmen, vždy bylo zbytečné rozlišování ruského dativu a lokálu, skoro vždy nominativu a akusativu atd. Problém „zbytnosti“ v jazyku má velký význam také pro spojovací techniku, neboť odstranění zbytečných sdělení umožní mnohonásobně zvýšit efektivnost sdělovacího a přijímacího zařízení. Proto se na tuto otázku soustřeďuje úsilí odborníků pro strojové překládání a spojovací techniku.

Jazykovědné práce v oblasti strojového překládání se musí opírat o důkladný statistický výzkum textu. Statistika je nutná pro vymezení zkoumaného materiálu, pro určení okruhů jevů, jež je třeba popsat a systematizovat. Kvantitativní charakteristiky umožňují hodnotit specifickou váhu různých jazykových jevů. Toto hodnocení umožní soustředit se na podstatné a odsunout stranou druhořadé. Kromě toho jsou kvantitativní charakteristiky nutné pro hodnocení ekonomičnosti toho či onoho řešení. Nelze ovšem absolutně přesně řešit všechny jazykové problémy; statistika však umožňuje najít řešení přibližná, nejpravděpodobnější. Statistický popis řeči je velmi důležitý nejen pro strojové překládání, ale i pro spojovací techniku, pro polygrafii, metodiku výuky jazykům atd. Proto jedním z předních úkolů je rozvinout potřebný statistický výzkum různých jazyků, především ruštiny.

Pro úspěšné provádění těchto výzkumů jsou nezbytné matematické stroje; to zase vyžaduje těsný kontakt jazykovědců s jinými odborníky z příslušných oblastí.

Na závěr připomeňme ještě jednu oblast spolupráce jazykovědců, matematiků a inženýrů: projektování speciálních překladových a informačních strojů, určených přímo pro práci s jazykem³⁾.

Ke koordinaci výzkumů ve všech těchto směrech přispěla nemálo první vše-svazová konference pro strojové překládání, která se konala v květnu 1958 v Moskvě. Všechny výzkumy v oboru strojového překládání je třeba vidět jako počáteční etapu širšího okruhu prací. Jejich konečným cílem je „naučit“ stroje lidskému jazyku. Realizace tohoto záměru způsobí pravou revoluci ve vědě a v technice. Řešení problémů strojového překládání znamená krok vpřed na cestě za tímto cílem.

Přeložila Irena Merglová

³⁾ Dnes se v SSSR i v jiných zemích provádějí experimentální překlady zatím na universálních matematických strojích, neupravovaných pro tento účel.