

# MATEMATICKÁ LINGVISTIKA V SOVĚTSKÉM SVAZU

Pavel Novák

Stále častěji užívaný výraz *matematická lingvistika* není jen dobovou náhradou za starší známý již výraz kvantitativní lingvistika, jehož význam je zcela průhledný — studium kvantitativní stránky jazyka. Musíme tu však výraz *matematická* pojímat dostatečně široce. Vedle těch matematických oborů, s jejichž některými elementárními partii jsme se seznámovali ve škole (aritmetika, statistika, teorie pravděpodobnosti, kombinatorika), patří do matematické a velmi abstraktní, většinou »nekvantitativní« obory, jako teorie množin, algebra, topologie, teorie grafů; přistupuje dále ta okolnost, že hranice mezi matematikou a některými jinými vědami, logikou, kybernetikou, je plynulá a že v matematické lingvistice musíme počítat i s matematickou logikou, logickou sémantikou, teorií informace, teorií algoritmu aj. Připomeneme-li si nyní, že spojení přídavného jména matematický se jménem příslušné vědy (srov. matematická fyzika, matematická ekonomie, matematická psychologie) znamená v podstatě uplatnění, využívání matematických metod a teorií v různých fázích badatelské činnosti, uznáme, že je rozsah pojmu matematická lingvistika širší, než se na první pohled zdá.

Setkáme se dále dnes s výrazem *aplikovaná* (užitá) lingvistika, jehož protějškem je výraz teoretická lingvistika. Výsledky jazykovědného zkoumání se dosud uplatňovaly při výuce jazykům, cizím i mateřskému (spisovnému), v jazykové kultuře, při překladatelské praxi, po-

případě při sestavování pomocných mezinárodních jazyků. K tomu bychom mohli připojit pomocí, kterou lingvistika poskytuje historii, literární historii, psychologii aj. Možnosti aplikace lingvistiky se radikálně rozšířily vznikem moderních strojů na zpracování informací, především samočinných počítačů, s nimiž se objevily takové úkoly, jako strojový překlad, různé druhy automatického sbírání a zpracování údajů obsazených ve vědeckých, technických a odborných textech, ovládnání automatických provozních linek řeči; tyto úkoly souvisí s automatickým zapisováním mluvené řeči a automatickým čtením psaného textu i s dlouho již prováděným průzkumem řeči pro potřeby spojovací techniky. Situace není zdaleka taková, že bychom pro splnění každého takového nového úkolu prostě sáhli do zásobárny našich poznatků o jistém jazyku nebo o jazycích vůbec, a pouze potřebné věci vybrali. Obvykle naopak zjišťujeme, že je třeba pro nové náročné požadavky naše znalosti doplnit, opravit, zpřesnit, usoustatnit, lépe organizovat. Nyní se do- stává do popředí nutnost přesně formálního popisu jazyka, a to — což se může zdát paradoxní — i jeho složek významových. A vzhledem k vývoji nových aplikací jazykovědy i vzhledem k tomu, že právě ty vyžadují aplikace matematických metod v lingvistice, znamená výraz aplikovaná *matematická lingvistika*. Toto zužování výrazu matematická lingvistika má však jisté nevýhody.

Poslední z našich úvodních terminologických poznámek bude zmínka o tzv. *strojové lingvistice*. Samočinné počítače, stroje na děrné štítky i některé jiné stroje a přístroje umožňují vedle uvedených již nových aplikací lingvistiky jednak mechanizaci a automatizaci nejrůznějších oblastí lingvistovy práce, jednak zatím ne zcela vyjasněné zavádění experimentu v jazykovědě, a to i v takových jejích oblastech, jako je teorie jazykového vývoje. I tato strojová lingvistika, jejíž rozsah se zčásti tedy překrývá s aplikovanou lingvistikou, nutně ovšem vyžaduje aplikace matematických metod.

★

Můžeme nyní konečně přejít k *matematické lingvistice v SSSR*, o jejímž dnešním postave-

ní, stavu i vývoji bývají u nás představy někdy dost nejasné nebo i zcela nesprávné. Matematická lingvistika má především v sovětské (a ruské) vědě dlouhou tradici.

Uvedme několik charakteristických, byť spíše »zevních« skutečností z posledních let. V roce 1955 se v moskevském Ústavu přesné mechaniky a výpočtové techniky začíná s pracemi na strojovém překladu. Koncem roku 1956 se na moskevské a leningradské univerzitě organizují první vědecké semináře z matematické lingvistiky a je založena Společnost pro otázky strojového překladu se sídlem v Moskvě. V roce 1957 se v Leningradě koná symposium o statistice řeči, na leningradské univerzitě se zavádějí přednášky z matematické lingvistiky a vycházejí první čísla sborníku Bjuulleten Ob'jedinenija po problemam mašinного перевода, který má od 8. čísla název Mašinnyj perevod i prikladnaja (= aplikovaná) lingvistika (dosahá dodnes celkem 14 svazků). R. 1958 je při sekci výzkumu řeči Komitétu akademie věd SSSR zřízen Komitét aplikované lingvistiky a v Moskvě organizována všesvazová konference o strojovém překladu, v časopise Voprosy jazykoznanija se poprvé objevuje rubrika, jejíž název se po jistém počátečním kolísání ustáluje jako »aplikovaná a matematická lingvistika«, vychází první svazek sborníku Problemy kibernetiki (dodnes 7 svazků) se stálým oddílem »otázky matematické lingvistiky«. V r. 1959 se v Leningradě koná symposium o matematické lingvistice, v AV SSSR je zřízena vědecká rada pro koordinaci práce v oblasti kybernetiky, v níž je i několik lingvistů, a vychází druhý poválečný logický sborník Logičeskije issledovanija, v němž poprvé publikují i lingvisté.<sup>1)</sup> V roce 1960 je v Červicích na Ukrajině organizována konference o aplikované lingvistice a v AV SSSR je zřízena vědecká rada pro aplikovanou lingvistiku, v roce 1961 se v Moskvě koná konference o zpracování informací, strojovém překladu a automatickém čtení textu, v Gorkém symposium o použití matematických metod při studiu uměleckých děl a o matematické lingvistice se jedná na 4. všesvazovém matematickém sjezdu.

Kdybychom chtěli systematicky probírat bo-

hatství otázek, o nichž byla na těchto konferencích a v těchto publikacích řeč, museli bychom k známým již hlediškům (podle aplikovaných matematických oborů, podle praktického účelu, popřípadě podle způsobu a stupně využití strojů na zpracování informací) připojit ještě jistou systematizaci celé problematiky jazykovědné. Takový přehled však zde udělat nemůžeme, a soustředíme se proto především na dvě tři zajímavé a dosud málo známé novinky.

O teoretických i praktických otázkách *strojového překladu*, který má spolu s přípravou automatického sbírání a zpracovávání informací z odborných textů klíčové postavení v celém problémovém komplexu matematické lingvistiky, jsou naši čtenáři zpraveni článkem B. Palka a P. Sgalla Automatizace překladu (Slovanský přehled 46, 1960, č. 5).

Především si ukážeme, jak je rozvoj matematické lingvistiky spjat s obrodou sovětské obecné jazykovědy, jejíž vývoj byl koncem čtyřicátých let (frontálním útokem marxistické doktríny) a začátkem padesátých let (v období kultu osobnosti) nemile zpomalen. Z několika možností, které se tu k ilustraci nabízejí, například obecná teorie vztahů jazykových systémů, kterou propracovává V. V. Ivanov, obecná teorie překladu V. J. Rozencovejga aj., vyběráme příklad tzv. *typologie jazyků*, a to proto, že no- vější pojetí typologie patří díky dílu našeho prof. V. Skalického k tomu nejdůležitějšímu, co přinesla československá lingvistika lingvistice světové.<sup>2)</sup> J a jsme tu tedy v oblasti, kde může být spolupráce sovětské a československé jazykovědy zvláště plodná.

Typologie je obecná jazykovědná disciplína zkoumající obecné principy, zákony jazykové stavby srovnávaním jazyků bez ohledu na jejich genetickou příbuznost. Zjišťuje, které vlast- nosti, rysy jazykové stavby se vyskytují, které se mohou a které se musí vyskytovat pospolu, které nikoli, a proč. Například již téměř sto padesát let je známo, že existuje jistá rovnováha mezi bohatstvím tvarosloví a vázanností slo- vosledu; jazyky s chudým tvaroslovím mívají pevný, gramatikalizovaný slovosled, jazyky s bo- hatým tvaroslovím nikoli.

Takové rysy jazyků, a na to se přišlo velmi brzo, mají ovšem vliv na problematiku, s kterou

<sup>1)</sup> Podíl matematické lingvistiky vzrůstá v třetím sborníku Primenenije logiki v nauke i technike, Mos- kva 1960. Matematická lingvistika je v jistém smyslu zastoupena i ve sborníku Logiko gramatičeskije očerki, Moskva 1961.

<sup>2)</sup> Srov. jeho »Zur ungarischen Grammatik«, Praha 1935, »Typ češtiny«, Praha 1951, a »Typologie slovan- ských jazyků, zvláště ruštiny«, v knize »Vývoj jazyka«, Praha 1960, s. 93—110.

se lingvista setkává při sestavování algoritmu strojevého překladače, tj. jistých přesných předpisů pro rozbor textů jazyka originálu a sestavení souznačných textů v jazyce překladu. Umíme si představit, že je třeba jiného postupu pro analýzu anglických vět, s jejich pevnými slovosledem a minimálním tvaroslovím, a jiného postupu pro analýzu vět většiny slovanských jazyků, v nichž jsou poměry právě opačné. Moskevský lingvista I. A. Meščuk, známý u nás i z překladu jeho článků, přišel na myšlenku sestavit schéma jakéhosi univerzálního algoritmu analýzy jazyků, v němž by se počítalo se všemi jejich možnými zvláštnostmi. Stavba jakéhokoli konkrétního jazyka by se pak mohla velmi přesně charakterizovat udáním těch částí schémat, jejichž zpracování je nutné pro analýzu tohoto jazyka. Prozatím je hotova morfologická část tohoto schématu, svého druhu dotazníku pro charakteristiku jazyků.<sup>4)</sup>

Typologická problematika je také spojena s tzv. převodním jazykem, umělo jazykovou soustavou plněci zprostředkující úlohu při strojevého překladače, avšak tuto souvislost musíme ponechat stranou, neboť bychom se nemohli při výkladu vyhnout přílišným podrobnostem.

Potřebami strojevého překladače bylo vyvoláno studium jazyka s pomocí tzv. *matematických modelů jazyka*, tj. jistých matematických konstrukcí, jejichž vlastnosti jsou v určitém směru blízké vlastnostem reálných jazyků.<sup>5)</sup> K takovým modelům, jejichž rozeznáváme několik druhů, patří například model O. S. Kulaginova.<sup>6)</sup> Kulaginová vlastně popisuje jistou skupinu velmi jednoduchých hypotetických jazyků a matematickým způsobem zkoumá jejich vlastnosti. S pomocí jistých vztahů zaměřitelnosti mezi slovy vět oněch jazyků tvoří formální obdoby známých slovních druhů apod. Vlastnosti konkrétních jazyků se pak zkoumají srovnáváním s oněmi umělými jazyky. Setkáváme se tu s jevem ve vědě dosti častým, že se totiž jisté překonané

5) K otázkám strojevého překladače v Moskvě, slovo a slovesnost 20, 1959, s. 295–299, a strojevé překladači v SSSR, Pokroky matematiky, fyziky a astronomie 5, 1960, 212–215.

6) I. A. Meščuk, »Morfologická analýza při strojním překladači«, Problemy kibernetiki 6, 1961, 207–212.

7) G. S. Cejtin, »K voprosu o postrojeniji matematičeskij modelij jazyka«, Doklady na konferencii po obrabotke informacii, mašinomu perevodu i avtomatičeskomu čteniju teksta, Moskva 1961, vypusk 3, s. 1. A. Jaurisová–M. Jauris, »Učiji teorie množin v jazykovědě«, Slovo a slovesnost 21, 1960, č. 1.

názory, postupy atd. opět vracejí, avšak na novém, vyšším stupni. Bylo ve své době pokrokovým poznáním, že jazyky s extrémním nakupením jistých vlastností, tzv. čisté typy, ve skutečnosti neexistují. Jak je vidět, dospívá se však nyní k názoru o užitečnosti i různé typové krajnosti konstruovat. Model Kulaginova dále pracovával I. I. Revzin, který také začal spojovat podobná studia s klasickou problematikou typologickou.<sup>7)</sup> O dalších výsledcích těchto snah dozvíme se sborníku Strukturno-tipologičeskije issledovanija, vydávaného Ústavem slavistiky AV SSSR v Moskvě, jehož první svazek nedávno vyšel.

Jiným směrem, kterým se děje zpřesňování typologické teorie — pokusy číselné charakterizovat míru jazykových vlastností, například stupně volnosti slovosledu, směrem, který však nemá zatím v Sovětském svazu výrazné zástupce, se dostáváme do oblasti kvantitativní.

\*

*Kvantitativní lingvistika* je nejstarší částí matematické lingvistiky. Základním problémem tu je vždy vzájemná vazba různých kvantitativních (číselných) charakteristik jazykových jevů a vazba těchto charakteristik s vlastnostmi kvalitatívními. Některé vztahy jsou dost prozkoumány. Například již A. M. Peškovskij (zemřel 1933) věděl, že se gramatické prvky jazyka (například koncovy, předložky) vyskytují v jazykových projevech mnohem častěji než prvky lexikální (slovní základny). To je základní poznatek, na němž například N. D. Andrejev buduje návrh algoritmu automatického sestavení gramatiky z dostatečně dlouhého textu.<sup>8)</sup> Se vznikem teorie informace přibýly k známým kvantitativním, statistickým charakteristikám textu i některé charakteristiky další, entropie, redundance aj. Např. entropií některého jazyka se obvykle míní číslo udávající, kolik »informace« nese průměrně písmeno v textech tohoto jazyka. Některé z takových charakteristik jsou prozatím známy jen pro několik málo jazyků. Sovětské zkušenosti ukazují podle našeho názoru celkem na to, že pro lingvistu je tu úkolem především

8) »O některých pojatijach tak nazývanej teoretiko-množestvennoj koncepcii jazyka«, Voprosy jazykoznanija 9, 1960, č. 6, s. 88–94.

9) »Modelirovanije jazyka na baze ego statističeskoi i teoretiko-množestvennoj struktury«, Rezisy sovětskaja i matematičeskoi lingvistike, Leningrad 1959, s. 15–22.

jazykovědná interpretace těchto charakteristik, výzkum vztahů těchto charakteristik k různým rysům a vlastnostem jazykové stavby a jazykových projevů. Víme například zatím jen zcela globálně, že je jisté známá neekonomičnost jazyků dána potřebou zabezpečit spolehlivost v dorozumívání styku.<sup>9)</sup> V otázkách metod zjišťování entropie a jiných kvantitativních charakteristik je naproti tomu těžké nějaký přínos od lingvistů očekávat.

Těmto a mnoha dalším příbuzným otázkám věnuje v poslední době mnoho pozornosti přední sovětský matematik a teoretik kybernetiky A. N. Kolmogorov. Souvisí to s jeho nedávno vydanými výstupy o protievolučném, v podstatě idealisticky motivovaným snahám apriorně vymezit kybernetice nějaké hranice.<sup>10)</sup> Kolmogorov upozorňuje, že je třeba zmocnit se kybernetickými metodami nejsubtilnějších strážek lidské duševní činnosti, a odtud jeho aktivní zájem o zkoumání jazyka, zejména jazyka uměleckých děl a poezie. Navrhl také nedávno nový způsob experimentálního zjišťování entropie jazyka, který má proti starším metodám C. E. Shannona, zakladatele teorie informace, značné výhody.<sup>11)</sup>

Na některé obecné otázky týkající se matematické lingvistiky mají sovětské jazykozpytci různé názory.<sup>12)</sup> Bývají například vyslovovány obavy, aby se při formálním popisu jazyka neztrácela specifická jazyka jako společenské-historické vzniklého a vyvíjejícího se jevu. Na takovou námitku lze odpovědět poměrně snadno: formálnost popisu je vlastností metody, nikoli předmětu, a není na místě směšovat cha-

9) Před časem proběhla v našem denním tisku zpráva, v níž se v jisté souvislosti tvrdilo, že entropie je matematický výraz charakterizující celou strukturu jazyka. Toto tvrzení je od základu chybné.

10) Vlastně jde o spolupracovnice N. Ryčkově, »Automatizacija životičeskijch jevov v prekladu I. Tondla K. Kulbata jako dodatku k prekladu knihy Z. Rovenčského, A. Ujomova a I. Ujomovovej stroj a myšlení, Praha 1962, s. 157–175.

11) M. Ryčková, »Lingvistika i matematika«, Nauka i život 1961, č. 9.  
12) Viz například Trv. AN SSSR, old lit. i jaz. 19, 1961, č. 1, s. 60–76. Ze starších jazykovědných prací k podobným matematickým lingvistickým otázkám viz A. A. Refor-matskij, M. I. Steblin-Kennanskij, L. R. Zinder.

rakter předmětu vědy a metody vědy. Takové směšování vědy vede nakonec k metodickému dualismu, jehož nejdůležitějším popřením je historický materialismus. Ten, jak známo, »vznikl jako rozšíření dialektického materialismu na společnost, jako aplikace dialektického materialismu na poznání společenského života«.

Další takovou spornou otázkou je vztah nových metod a obecné jazykové teorie. Můžeme se ovšem lišit v odhadu dalšího vývoje jazykové vědy, avšak těžko se lišit v uznávání oboustranně aktivního vztahu teorie a praxe, v tom, že každé poznání je relativní a že prostě žádná věda není pojištěna proti nutnosti stálých — a někdy hodně radikálních — modifikací své teorie.

To jsou ovšem zásadní otázky *metodologické* a je nutno řešit je ve spolupráci s filozofy. Přehlédneme-li tematicky plán časopisu Voprosy filosofii, uveřejněný v I. čísle 16. ročníku, kde čteme taková témata, jako vnitřní logika vývoje soudobé vědy, úloha filosofických předpokladů ve vytváření vědeckých teorií, o vztahu oboustranných a formálních momentů v rozvoji vědeckého poznání, úloha znaků v konstrukci formálních modelů jazyka aj., musíme sovětským jazykovědcům trochu záviďet.<sup>13)</sup>

13) O matematické lingvistice v jiných slovanských zemích pojednáme jindy.

#### Vybraná sovětská a naše literatura:

O. S. Achmanova, I. A. Meščuk, Je. V. Padučeva, R. M. Frumkina, »O točnyh metodach issledovanija jazyka (O tak nazývanej »matematičeskoi lingvistike)«, Moskva 1961. — N. D. Andrejev, L. R. Zinder, »Osnovnyje problemy prikladnoj lingvistiki«, Voprosy jazykoznanija 8, 1959, č. 4, 3–19. — R. L. Dobrušin, »Matematičeskije metody v lingvistike«, Matematičeskije prosvěščanije 6, 1961, 37–60. — V. V. Ivanov, heslo »Matematičeskaja lingvistika«, Bol'saja sovětskaja enciklopedija, tom 51. — V. V. Ivanov, S. K. Šaumjan, »Lingvističeskije problemy kibernetiki i strukturalizacija lingvistiki«, sb. Kibernetiku na službu komunizmu I, Moskva 1961, 218–234. — L. I. Revzin, »O sootnoženii strukturyh i statističeskijch metodov v sovremennoj lingvistike«, Voprosy statističeskijch reči, Leningrad 1958, 45–57. — L. I. Revzin, »Modelij jazyka«, Moskva 1962. — P. Šval, »Novoje otroženie matematičeskijch metod v jazykovědě«, Slovo a slovesnost 20, 1959, č. 1. — S. K. Šaumjan, Nasuščnije zadatki strukturalnoj lingvistiki, Trv. AN SSSR, old lit. i jaz. 21, 1962, č. 2. — A. A. Zinov'jev, »O matematičeskoi lingvistike«, Voprosy filosofii 13, 1959, č. 9, 133–139. — A. A. Zinov'jev, I. I. Revzin, »Logičeskaja model' kak sredstvo naučnogo issledovanija«, Voprosy filosofii 14, 1960, č. 1, 62–80.

● Časopis »Kurier Polski« uspořádal čtenářskou anketu na nejlepší román minulého roku. Nejvíce hlasů dostal román Jaroslawa Iwaszkiewiczze »Stawa i chwala«, jehož 3. díl vyšel v provincii minulého roku.