

Několik prací uvádějících do matematické lingvistiky

V posledních letech vyšla řada publikací, které byly myšleny jako úvody do některých oblastí jazykovědného bádání nebo přístupů k němu, zejm. do matematické lingvistiky. Jsou to:

Ju. D. Apresjan, *Ideji i metody sovremennoj strukturnoj lingvistiky (kratkiĭ očerk)*, Moskva 1966, 302 s. (franc. překlad pod názvem *Idées et méthodes de la linguistique structurale contemporaine* má vyjít v knižnici Monographies de linguistique mathématique v nakl. Dunod, Paříž);

A. V. Gladkij - I. A. Meľčuk, *Elementy matematičeskoj lingvistiky*, Moskva 1969, 192 s.;

M. Gross - A. Lentin, *Notions sur les grammaires formelles*, Paříž 1967, 197 s. (angl. překlad *Introduction to formal grammars*, Berlín - Heidelberg - New York 1970, XI + 231 s.);

R. Gunzenhäuser (ed.), *Nicht-numerische Informationsverarbeitung*, Vídeň - New York 1968 (zvl. VI. Sprachbearbeitung und Dokumentation mit Rechenanlage, s. 277—399);

S. Marcus - E. Nicolau - S. Stati, *Introducere in lingvistica matematică*, Bukurešť 1966, 336 s.;

Ch. Muller, *Initiation à la statistique linguistique*, Paříž 1968, 248 s.;

M. Novotný, *Einführung in die algebraische Linguistik*, Bonn 1967, 56 s.;

Z. Pawlak, *Gramatyka i matematyka* (Biblioteczka matematyczna 20), Varšava 1965, 112 s.

Cílem této zprávy je charakterizovat jejich obsah, zaměření a přístupnost.¹

¹ O některých jiných se v SaS podrobněji už referovalo: O. S. Achmanova - I. A. Meľčuk - R. M. Frumkina - Je. V. Padučeva, *O točnych metodach issledovanija jazyka*, Moskva 1961 (vyšlo též v angl. překladu *Exact methods in linguistic research*, Berkeley, California 1963, a v českém překl. *Exaktní metody v jazykovědě* (skriptum filos. fak. UJEP), Praha 1965), rec. P. Nováka - B. Palca a M. Těšitelové v SaS 23, 1963, 131—137; D. G. Hays, *Introduction to computational linguistics* (Mathematical linguistics and automatic language processing 2), New York 1967, rec. K. Paly přinese SaS; G. Herdan, *Quantitative linguistics*, London 1964, rec. oddělení matem. lingvistiky v ÚJČ v SaS 28, 1967, 58—64; S. Marcus, *Introduction mathématique à la linguistique structurale* (Mo-

Nejprve však jednu poznámku terminologickou. V názvech těchto prací se vyskytují pojmenování, jejichž české ekvivalenty jsou: *strukturální lingvistika*, *matematická l.*, *algebraická l.*, *statistická l.* (vlastně téměř s ním synonymní *lingvistická statistika*) a *počítačová lingvistika*.² Se všemi těmito výrazy jsou spojeny jisté nevýhody: a) Sugerují existenci jakýchsi samostatných, vzájemně oddělených disciplín. Kde však např. „končí“ strukturální a „začíná“ matematická lingvistika?³ b) Jsou — aspoň potenciálně — systematicky víceznačné. Mohou se totiž pojímat jako název jistého více či méně dominujícího metodologického nebo instrumentálního přístupu ke zkoumání a popisu jazyka anebo jako název jisté oblasti zkoumání charakterizované vlastní problematikou, metodami a výsledky. Nejmarkantněji se toto dvojí chápání uplatňuje u pojmenování *matematická lingvistika*, u něhož v prvním případě jde spíše o chápání obdobné jako např. u výrazu *matematická psychologie*, kdežto v druhém případě spíše o chápání obdobné jako u výrazu *matematická fyzika*.⁴ c) U některých autorů pozorujeme tendenci k vytváření nevhodných synonymických řad (*matematická lingvistika* = *kvantitativní l.*; *kvantitativní l.* = *statistická l.*). Tato tendence je dokladem toho, jak i ve vědeckém jazyce může momentální rozsah aplikace pojmenování ovlivňovat jeho významové užití.

Část těchto nevýhod souvisí s tím, že v současné době přežívá úzus několika období předchozích a že se mnohdy nebere zřetel na původní podněty pro vznik příslušných pojmenování (např. *kvantitativní l.* — „třeba zkoumat nejen kvalitu“). Zmíněné výrazy jsou nejméně problematické ve funkci spíše pragmatické, orientační, tak v názvu knihy, časopisu, oddílu bibliografie, názvu instituce apod. Jinak je vhodné zacházet s nimi tak, aby se jejich nevýhody co nejvíce omezily.

Z uvedených publikací zabírá do nejvíce oblastí úvod S. Marcuse - E. Nicolaua - S. Statiho, určený lingvistům, inženýrům a matematikům, kteří se chtějí seznámit s novým uplatněním svých oborů, i jiným zájemcům. Má 11 kapitol: (1) strukturální lingvistika [historický přehled], (2) obecný pohled na matematickou lingvistiku [ta se chápe jako etapa vývoje strukturální lingvistiky (s. 59) a rozumí se jí studium kvantitativních a formálních aspektů jazykových jevů (s. 42)], (3) matematické modely v lingvistice, (4) algoritmické modelování jazyka, (5) informační jazyky, (6) kosmická řeč [o Freudenthalově lincosu⁵], (7) automatický překlad, (8) pravděpodobnost, informace a entropie v jazyce, (9) jiné statistické metody v lingvistice, (10) jazyk a kód, (11) automatické rozpoznávání obrazů. Klade se důraz na motivaci zaváděných pojmů, ty jsou osvětlovány na

nogographies de linguistique mathématique 1), Paříž 1967, rec. L. Nebeského v SaS 30, 1969, 333—334; P. Sgall a kol., *Cesty moderní jazykovědy* (Jazykověda a automatizace), Praha 1964, rec. L. Řeháčka v SaS 27, 1966, 147—152, aj.

² To je výraz, který navrhuji jako ekvivalent pro angl. *computational linguistics*, něm. *computer-orientierte Linguistik*. Výraz *strojová lingvistika* byl původně míněn za angl. *mechanolinguistics*, která má širší význam.

³ Srov. Gladki - Meičuk, *Elementy...*: jakýkoli vědecký popis [...] musí být logicky důsledný [...] jednoznačný [...] a plně explicitní [...]. A takový popis je také formálním, tj. — při dosti vysoké úrovni formalizace — v podstatě matematickým popisem (s. 9).

⁴ Srov. A. V. Gladki, *O formálních metodach v lingvistice*, VJaz 15, 1966, č. 3, s. 52 a Gladki - Meičuk, *Elementy...* s. 16. — G. A. Miller, *Mathematics and Psychology*, New York-London-Sydney 1964, praví o matematické psychologii, že je více postojem než předmětem (s. 1) a M. Bunge, *Scientific research I (The search for system)*, Berlin-Heidelberg-New York 1967, konstatuje o termínu *matematická fyzika*, že označuje zkoumání čistě matematických problémů, které se objevily v teoretické fyzice a které mají pro fyziky instrumentální hodnotu (s. 468).

⁵ Viz zprávu v SaS 23, 1962, 77—78.

příkladech; teorémy se uvádějí bez důkazů. Ke každé kapitole je připojena výběrová bibliografie.

Apresjanův nástin předpokládá čtenáře, který zvládl některý ze standardních úvodů do jazykovědy, a chce ho uvést do problematiky a metod současné strukturální lingvistiky a připravit ho k studiu speciální literatury. Strukturální lingvistika se přitom chápe jako jazykovědná disciplína studující jazyk jako systém znaků a jako kód (s. 36). Práce má 5 částí: (I) z dějin strukturální lingvistiky: (kap. 1) proč vznikla strukturální lingvistika [vnější a vnitřní podněty], (2) bezprostřední předchůdci strukturální lingvistiky [Baudouin de Courtenay, F. de Saussure], (3) klasické školy strukturální lingvistiky; (II) lingvistické modely: (1) pojem modelu, (2) typy lingvistických modelů [srov. názvy a členění dalších částí], (3) elementární matematické pojmy; (III) modely [lingvistického] zkoumání: (1) modely dešifrování [ty se uplatňují na textu, popř. už částečně analyzovaném; probírají se Suchotinovy metody vydělení vokálů a konsonantů, zjištění vztahu syntaktické závislosti a Harrisova metoda zjištění hranic morfů], (2) experimentální modely [uplatnění technik dodávání, vynechávání a substituce prvků, změny jejich pořádku, transformace forem na jiné a překladu k řešení otázek morfologie, tvoření slov a syntaxe]; (IV) modely jazykové činnosti: (kap. 1) generativní modely, (2) syntaktické modely analýzy, (3) sémantické modely analýzy; (V) pojem metateorie [modelu hotového lingvistického popisu, zde se uvádí Meičukův návrh hodnocení lingvistického popisu⁶ a některé výsledky teorie generativních gramatik]. Bibliografie obsahuje kolem 390 položek.

Algebraické lingvistice jsou věnovány čtyři práce, i když názvy dvou z nich by mohly ukazovat na záběr širší. Nejprístupnější je knížka polského matematika Z. Pawlaka, která vyšla v knižnici odpovídající Škole mladých matematiků nakladatelství Mladá fronta a Matematické knižnici Státního pedagogického nakladatelství. Pawlakův úvod obsahuje tyto kapitoly: (1) základní pojmy, (2) nekontextové gramatiky, (3) kategoriální gramatiky, (4) gramatiky s konečným počtem stavů, (5) popis činnosti [pojem programu], (6) genetické gramatiky [struktura genetických kódů, tato oblast je dotčena pouze v této knížce] a (7) jazyky matematických strojů. Výklad je omezen na uvedení a ilustraci pojmů.

Náročnější je úvod Gladkého a Meičuka, určený lingvistům a matematikům zajímajícím se o otázky jazyka, jehož obsah tvoří kapitoly: (1) obsah pojmu matematická lingvistika [chápe se jako nekvantitativní matematická disciplína „obrácená“ k přirozeným jazykům a lingvistice (s. 16), jako teorie struktury teorií konkrétních jazyků a jazyka vůbec (s. 175)], (2) formální gramatiky, (3) třídy generativních gramatik [Chomského gramatiky typu 1, gramatiky frázové, nekontextové, binární nekontextové, gramatiky s konečným počtem stavů], (4) generativní gramatiky a přirozené jazyky, (5) o formálních vlastnostech generativních gramatik [problémy řešitelnosti a hodnocení složitosti odvozování], (7) některé jiné pojmy a problémy matematické lingvistiky: (A) modelování jazyků [kategoriální gramatiky, zásobníkové automaty, logická analýza jazyka], (B) modelování lingvistických zkoumání [některé základní pojmy analytických modelů jazyka]. Teorémy se uvádějí bez důkazů. Ke knížce je připojena výběrová bibliografie.

Zatímco úvod Gladkého a Meičuka věnuje většinu místa generativním gramatikám a analytických modelů si všímá jen okrajově, věnuje se oběma směry m algebraické lingvistiky stejná pozornost — navíc z jednotčího hlediska — v krátké práci brněnského matematika M. Novotného,⁷ psané pro potřebu

⁶ O standartnoj forme i količestvennyh charakteristikach nekotorych lingvističeskich opisanih, VJaz 12, 1963, č. 1, s. 113—123.

⁷ M. Novotný je též autorem skripta filoz. fak. UJEP Matematika pro lingvisty, Praha 1965, 69 s.

vydávající instituce. Kapitoly: (1) elementární pojmy, (2) množinový model, (3) obecná gramatika [tj. neomezený přepisovací systém] a její jazyk, (4) gramatika a její jazyk. Teoremy se uvádějí s důkazy.

Nejobsáhlejší z této skupiny je úvod Grossův a Lentinův, který vznikl z universitních přednášek konaných pro matematiky a zčásti i pro lingvisty. Probírá se v něm látka odpovídající zhruba Chomského přehledu Formal properties of grammars,⁸ avšak je více rozvedena. V stručném úvodu, jehož autorem je N. Chomsky, se algebraická lingvistika pojímá jako studium formálních vlastností přirozeného jazyka s odhlédnutím od jejich realizací v jednotlivých jazycích. Práce obsahuje tři části a doplněk: (I) logicko-algebraický úvod: (1) slova, pologrupy, jazyky, (2) obecné poučení o formálních systémech, (3) kombinatorické systémy, (4) Turingovy automaty, (5) vyčíslitelnost, rozhodnutelnost, (6) kombinatorické systémy a Turingovy automaty, nerozhodnutelné problémy; (II) některé významné třídy jazyků: (7) jazyky Chomského [jazyky generovatelné nekontextovou gramatikou], (8) nerozhodnutelné vlastnosti nekontextových gramatik, (9) zásobníkové automaty, (10) jazyky Kleeneho [jazyky generovatelné gramatikou s konečným počtem stavů] a konečné automaty, (11) jazyky definované soustavami rovnic, (12) kontextové gramatiky, lineárně ohraničené automaty; (III) algebraický pohled: (13) homomorfismus pologrup, (14) doplňky k jazykům Kleeneho, (15) doplňky k jazykům Chomského, (16) algebraické jazyky. Stručný dodatek pojednává o transformačních gramatikách, a to podle stavu znalostí o nich z hlediska spíše popisného a obsahového. Je připojena výběrová komentovaná bibliografie.

Do statistické lingvistiky uvádí práce Mullerova, určená lingvistům, jimž chce autor předvést a v principech i prakticky objasnit běžné statistické operace uplatňované na faktech jazyka a stylu. Kniha je rozdělena do dvou částí. V první části (o principech a metodách lingvistické statistiky) se v 17 kapitolkách uvádějí některé ze základních pojmů a vět teorie pravděpodobností a matematické statistiky (teorie odhadu a testování hypotéz). Část druhá (o lexikální statistice) obsahuje v prvních šesti kapitolkách základní pojmy, v dalších pak seznamuje s některými konkrétními problémy lexikální statistiky (většinou na materiálu francouzské klasické literatury).

7. kap. se týká jednoduchých vztahů mezi délkou textu a počtem různých slov v textu (počtem prvků „slovníku“, f. vocabulaire); 8. kap. rozložení četností prvků slovníku v textu, 9. kap. pokusů o přesnější vystižení tohoto rozložení četností (Zipfův zákon, Waringovo rozložení), 10. kap. toho, co lze říci o počtu prvků slovníku v části textu na základě znalosti délky a počtu prvků slovníku v textu celém, 11. kap. se týká toho, co lze obdobně říci o rozložení četností prvků slovníku v části textu na základě znalosti rozložení četností prvků slovníku v celém textu, 12. kap. růstu počtu prvků slovníku v závislosti na délce textu, 13. kap. problematiky prvků s četností 1, 14. kap. výskytu jednotlivých slov nebo jejich skupin v různých částech textu, 15. kap. problematiky pravděpodobnosti prvků množiny abstraktních výrazů, které se vyskytují ve všech uvažovaných textech (f. lexique) a 16. kap. způsobů zjišťování vztahu slovníku dvou textů. Jsou připojena cvičení (některá řešená), potřebné numerické tabulky a nezbytná literatura.

Zbývající publikace je věnována počítačové lingvistice. Sborník vydany R. Gunzenhäuserem a určený i pracovníkům různých technických a vědeckých oborů, kteří nemají znalosti z programování a obsluhy samočinných počítačů, podává přehled nových aplikačních možností v oblasti nenumerického

⁸ Srov. *Handbook of mathematical psychology 2* (R. Luce - R. Bush - E. Galanter, eds.), 1964, 328—418 (dostupné v ruském překladě: *Kibernetičeskij sborník (novaja serija)* vyp. 2, 1966, 121—230), srov. rec. K. Paly v Sas 27, 1966, 74—78.

zpracování údajů. Ze sedmi oddílů se naší problematiky týká VI. (zpracování jazyka a dokumentace s pomocí počítačů), který obsahuje 6 kapitol od různých autorů: (A) D. Krallmann, Strojová analýza přirozeného jazyka [přehled, do něhož je kromě oblastí, o nichž pojednávají další kapitoly, zahrnuto i zkoumání jazykových jednotek, postupy automatické syntaktické analýzy a systémy otázek a odpovědí (angl. question-answering systems, něm. Frage-Antwort-Systeme); autor pokládá počítačovou lingvistiku, computer-orientierte Linguistik, za část matematické lingvistiky, s. 277], (B) U. Winkler, Stavba a aplikační možnosti elektronických slovníků [sestavování seznamů slov různých typů, přepracovávání slovníků], (C) H. Schnelle-G. Engelen, Strojové překládání jazyků [obsahuje vedle přehledu, v němž se věnuje pozornost systému A. Hoppeho, problematiku dat a jejich struktur, lingvistického zpracování dat a zkušenosti z hlediska programování], (D) D. Krallmann, Stylistický popis textu s pomocí statistických metod [popis analýzy, podání výsledků a jejich interpretace], (E) F. W. Kistermann, Dokumentace a elektronické systémy zpracování dat [konvenční a nekonvenční dokumentace, selektivní rozšiřování informace, tj. sdělování informací nikoli podle přímých dotazů, nýbrž podle trvalých zájmových profilů zákazníků]. Poslední kapitola tohoto oddílu z pera W. Lingenberga je věnována strojovému zpracování dat v knihovnách. Ke každé kapitole je připojena bohatá bibliografie (celkem víc než 350 titulů).

Nepřihlížíme-li k dostupnosti některých z těchto publikací, můžeme uzavřít, že pro „oblasti“ lingvistiky, o nichž byla řeč, jsou k dispozici už dobré úvody nebo přehledy.