

Pokroky matematiky, fyziky a astronomie

Eva Benešová

O matematické lingvistice

Pokroky matematiky, fyziky a astronomie, Vol. 9 (1964), No. 6, 335--343

Persistent URL: <http://dml.cz/dmlcz/139554>

Terms of use:

© Jednota českých matematiků a fyziků, 1964

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>

O MATEMATICKÉ LINGVISTICE

EVA BENEŠOVÁ, Praha

Rozvoj jazykovědy se až donedávna obešel bez propracovaného formálního aparátu. Zatímco v mnohých jiných vědách se stále více využívalo na jedné straně matematiky a logiky, na druhé straně experimentálního výzkumu s přesnými metodami, v jazykovědě se někdy operovalo s formulacemi a závěry neujasněnými a nejednoznačnými (srov. [1]).

Pronikání matematických a formálně logických metod v posledních desetiletích zasáhlo i lingvistiku. Vhodnou půdou pro tuto význačnou proměnu, která je dnes v nejpřednějších rozvoji, začala připravovat na počátku 20. století strukturní lingvistika, za jejíhož zakladatele se obecně považuje F. DE SAUSSURE. V jeho *Cours de linguistique générale*, vydaném r. 1916, jsou obsaženy základy nového chápání jazyka. Jazyk je považován za systém, jehož prvky jsou navzájem spolu spjaty určitými vztahy, a právě tyto vztahy mají být pomocí strukturní komparativní analýzy synchronicky (v jedné časové rovině) zkoumány. Strukturní lingvistika se zabývá také srovnáváním jazyků nepříbuzných (na rozdíl od předešlého období), aby lépe poznala zákonitosti jazykové struktury. Zasluhou F. de Saussura, J. BAUDOINA DE COURTENAY a jiných badatelů počal být tento postup spojován s chápáním funkčním (jde o fungování jazykových prvků ve smyslu jejich platnosti, úlohy), tj. se studiem komunikativní úlohy jazyka. Tento rys se projevil hlavně ve studiu systému hlásek a jejich funkce pražskou školou koncem 20. let. (O tomto období podrobněji viz MATHESIUS [2].)

Dnes je již nepochybné, že je možno jazyk studovat přesnými metodami podobně jako jiné systémy, které mohou být rozebírány se zřetelem ke vzájemným vztahům mezi jejich prvky. Vznik *matematické lingvistiky* plně poskytuje možnost rozvíjení takových postupů. Sám název tohoto nového odvětví jazykovědy není příliš šťastný, jak ukazuje BAR-HILLEL [3]. Slovo „matematický“ je správné, pokud je chápeme ve smyslu „teorie formálních systémů“. Avšak je nepřesné a zamlžující, spojujeme-li tento pojem pouze s čísly a kvantitativními postupy. O kvantitativních problémech pojednává jen část matematické lingvistiky — tzv. *kvantitativní lingvistika*. Pro oblast jazykového zkoumání zabývajícího se z nového hlediska *teorií gramatiky*, která se tak stává předmětem matematického exaktního a abstraktního popisu, se užívá název *algebraická lingvistika*. Konečně tzv. *lingvistika aplikovaná* provádí aplikaci matematických metod na praktické řešení aktuálních úkolů, zejména v souvislosti s kybernetickými možnostmi samočinných počítačů.

1. KVANTITATIVNÍ LINGVISTIKA

Tato disciplína využívá pro své účely statistických postupů, teorie pravděpodobnosti a teorie informace. Statistické metody mají v lingvistice již určitou tradici. Pomineme zde používání kvantitativních postupů v širším slova smyslu, které jsou, jak vystihl DOBRUŠIN [1], vlastní práci všech jazykovědců a byly vlastně používány všeobecně a odedávna, neboť „vždy jsou v jazyce jevy častější a řidší“; statistiky přímo bylo již ve 30. letech i dříve využíváno jak ve fonologii, tak v oblasti slovní zásoby. Frekvenční slovníky měly svůj význam pedagogický pro sestavování učebnic (výběr minimální slovní zásoby aktivní a pasivní) a staly se též podkladem pro poznávání lexikální stránky jazyka vůbec [4]. V kombinaci s teorií pravděpodobnosti matematická statistika umožňovala a stále umožňuje další výzkum v oblasti lexikální, stylistické i v typologii a v historické jazykovědě.

Rozsah zkoumaného materiálu při statistickém výzkumu slovní zásoby má být co největší, aby bylo možno využít zákona velkých čísel. Relativní frekvence slov se pak blíží pravděpodobnosti jejich výskytu, což se týká hlavně formálních (gramatických) slov. Slovní zásoba se zkoumá jednak z hlediska funkce, kdy rozlišujeme slova plnovýznamová a gramatická; ukázalo se, že gramatická slova (předložky, spojky) jsou nejvíce frekventovaná; četnost je v nepřímém poměru k jejich počtu. Jinak lze zkoumat slovní zásobu z hlediska délky slov, kdy délka a frekvence užití slova jsou rovněž v nepřímém poměru. Na základě experimentálního měření dospíváme pak k pojmům „index opakování“ (průměrné opakování slov v textu), „distribuce“ (rozložení slov v textu), „pořadí slova“ (umístění slova ve frekvenčním seznamu). Stoupání frekvence a s ním souběžné klesání počtu slov vyjádřil G. K. ZIPF v zákonu $r \cdot f = \text{konst}$ (f = frekvence slov, r = pořadí ve frekvenčním seznamu). Tento vzorec nevystihoval přesně jazykové zákonitosti a byl opraven B. MANDELBROTEM na tzv. kanonický zákon: $p_r = P(r + q)^{-\beta}$ (p = pravděpodobnost výskytu slova, P = charakteristika rozsahu výběru, β = konstanta pro daný text, q = konstanta korigující frekvenci slov s nízkým pořadím). Statistickým výzkumem slovníku se též zabýval YULE. Podrobnější informace o těchto postupech viz ve [4] a [5].

Je samozřejmé, že pravděpodobnostních a statistických postupů se dosti značně využívá též ve stylistice, jak pro komparaci jednotlivých jazykových stylů (styl odborný, umělecký apod.), tak pro studium stylu jednotlivých autorů nebo jednotlivých literárních děl. I když samotné kvantitativní postupy rozhodně nikdy nestačí k vyřešení tak složitých a jemných otázek, jako jsou otázky stylistické, a právě tento jednostranný přístup bývá slabinou některých studií, přece jsou nutnou součástí vytvoření vstřížné a pravdivé charakteristiky stylistického tématu, které zkoumáme.

V této oblasti se např. P. GUIRAUD zabývá slovníkem autora z hlediska tzv. „koncentrace“ (užívání omezeného počtu slov s velkou frekvencí) $C = \text{součet frekvencí 50 nejvíce frekventovaných slov ku délce textu } N$ a „bohatosti“ (větší počet slov s malou frekvencí) $R = V/\sqrt{N}$. Yule se pokusil o vypočtení tzv. charakteristické konstanty, která by byla typická pro styl určitého autora a napomáhala by iden-

tifikaci literárních děl:

$$K = \sum_{r=1}^s r^2 n_r / N^2 - 1/N,$$

kde n_r je počet slov o frekvenci r a N délka textu (viz [5] a [6]).

O řešení obecných otázek historicko-srovnávacích se v USA pokouší tzv. glottochronologie (lexikální statistika) — jejím předním představitelem je SWADESH —, která se zabývá určováním chronologie jazykových rodin. Na základě srovnávání formálně podobných slov základního slovního fondu snaží se určit období, v němž se jazyky od sebe oddělily. Typologickými problémy se zabývá GREENBERG. HERDAN ve svém díle *Language as Choice and Chance* se snaží o vyřešení starého lingvistického problému „langue“ a „parole“. Posuzuje, co v jazyce je věcí svobodného výběru mluvčího a co je podmíněno strukturou jazyka. Podle jeho závěrů je shoda mezi členy jazykového kolektivu nejen v tom, že používají stejného systému fonémů¹⁾, slovníku a gramatiky, ale i v relativní frekvenci používání těchto prostředků. Individuální projevy jsou tedy určité výkyvy konstantní pravděpodobnosti. O těchto otázkách píše SGALL [6] a ZVEGINCEV [7].

Nové perspektivy v kvantitativní lingvistice otevřela teorie informace. Značná pozornost se věnuje zejména výzkumu entropie²⁾ a redundance³⁾ prvků jazykového systému. C. SHANNON se zabýval stanovením entropie a redundance pro psanou angličtinu ($H = -\sum p(A) \lg p(A)$; $R_n = 1 - H_n/H_0$, kde A = jeden z jevů, p = pravděpodobnost výskytu jevu). Jeho metody bylo použito též u nás (viz práci L. DOLEŽELA [8]). Na základě rozložení četností grafémů⁴⁾ a jejich digramových kombinací v textech byla vypočítána redundance nultého, prvního a druhého řádu v češtině ($H_0 = 5,39$, $H_1 = 4,67$, $H_2 = 3,87$ bitu) a metodou experimentální predikce stanovena mezní entropie ($H_\infty \varepsilon \langle 1,04 - 1,94 \rangle$). (Srovnej v angličtině: $H_0 = 4,76$, $H_1 = 4,03$, $H_2 = 3,32$, $H_\infty = 1,40$.) Jsou známy hodnoty entropie pro ruštinu, němčinu, francouzštinu i některé jiné jazyky. Podle L. Doležela by entropie numerická mohla sloužit jako informačně teoretická charakteristika určitého jazyka, entropie predikční jako charakteristika jednotlivých stylů.

Redundance je pro reálné jazykové sdělení nutná, neboť je základem dobrého porozumění sdělovanému textu. V různých jazycích je různá: v ruštině 72%—82%, v angličtině 70%—81%, v češtině 64%—81%, v němčině 66% ap. Metody zjišťování entropie a redundance nejsou dosud jednotné, takže teprve další výzkum může přinést obecně lingvistické závěry. Protože ve srovnání s ostatními kódy má jazyk význačnou specifičnost, bude třeba, aby k rozřešení těchto otázek přispěly též jiné vědy, např. fyziologie, psychologie ap. [5].

1) *Foném* je hláska s významotvornou schopností. O systému fonémů pojednává fonologie.

2) *Entropie* slouží jako míra „neurčitosti“ a průměrného množství informace v jazykovém sdělení.

3) *Redundance* odráží nadbytečnost znaků v jazykovém sdělení.

4) *Grafém* je jednotka grafické roviny jazyka.

2. ALGEBRAICKÁ LINGVISTIKA

Snaha formulovat pomocí matematických metod popis jazyka a přispět tak k vytvoření exaktní a formální teorie gramatiky, z níž by bylo vyloučeno spoléhání se na jazykové povědomí mluvčích, je pro teoretickou lingvistiku z oblastí matematické lingvistiky nejdůležitější. Základem takového postupu je zjištění těsného vztahu mezi novou teorií gramatiky a matematickou teorií automatů, a tedy i teorií rekurzivních funkcí a teorií algoritmů. Vzniká tak úsek bádání společný lingvistům i matematikům, přinášející cenné poznatky oběma stranám.

Velmi podstatným přínosem pro rozvíjení počátečních pokusů o vytvoření formálního aparátu pro popis přirozených jazyků — o tyto počátky se zasloužili HARRIS, HOCKETT, HJELMSLEV, ULDALL aj. — byl též výzkum logické syntaxe jazyka (CARNAP, AJDUKIEWICZ). Z hlediska matematického přispěli k těmto pokusům POST a CURRY. Přehled tohoto vývoje podává BAR-HILLEL [3].

Protože je jazyk — podle pojetí dnes v algebraické lingvistice běžného — nekonečná množina vět, nedají se jeho prvky (tj. věty, které jsou konečnými posloupnostmi symbolů příslušného jazyka) vymezit výčtem, nýbrž pouze procedurou. Úkolem procedury (systému pravidel, jehož fungování je obdobné činnosti matematického automatu) je pak buď přecíslit všechny gramaticky správné věty jazyka a generovat kteroukoliv z nich, ale žádnou jinou (procedura generativní), nebo po konečném počtu kroků rozhodnout, zda je daná posloupnost symbolů co do své syntaktické struktury gramaticky správnou větou (procedura rekognoskativní). V obou případech se zároveň přiřazují větám strukturní charakteristiky.

Americký lingvista N. CHOMSKY ([9], [10], [11]) stanovil si jako cíl generativní popis jazyka. Generativní gramatika vymezuje jazyk jako množinu gramaticky správných vět. Pravidla mají tvar $\varphi \rightarrow \psi$, kde \rightarrow čteme „přepiš“ a φ a ψ jsou řetězy symbolů. Rozlišují se symboly koncové (terminální), odpovídající jednotkám popisovaného jazyka (např. jednotlivým morfémům⁵), a symboly pomocné, tvořící vnitřní abecedu gramatiky. Určitý řetěz symbolů je označen jako řetěz výchozí. Při generování věty postupujeme tak, že vyjdeme z výchozího řetězu a uplatňujeme pravidla gramatiky, tzn. řetěz symbolů upravujeme tak, že levou stranu pravidla v něm obsaženou nahrazujeme pravou stranou, až dojdeme k řetězu, který podle pravidel dané gramatiky dále přepisovat nelze. Je-li gramatika adekvátní pro daný jazyk, je možno tímto způsobem generovat právě všechny koncové řetězy (řetězy složené z koncových symbolů), odpovídající gramaticky správným větám tohoto jazyka.

N. Chomsky vymezil různé typy generativních gramatik a ukázal, jak jsou uspořádány co do generativní síly.

I. Gramatika s konečným počtem stavů: jestliže $\varphi \rightarrow \psi$, pak existují A, ω, a, B takové, že $A \rightarrow \omega, \omega = aB$ nebo $\omega = a$. (Velké písmeno označuje vždy jeden pomocný symbol, malé písmeno symbol koncový, řecké písmeno řetěz symbolů.) V grama-

⁵) *Morfém* je jednotka tvaroslovné roviny jazyka. Je to nejmenší nesamostatná významová část slova.

tice tohoto typu je možno dojít k libovolně dlouhým větám, obsahují-li pravidla rekurzivní symboly (např. $A \rightarrow aB$, $B \rightarrow bA$).

II. Frázová nekontextová gramatika: jestliže $\varphi \rightarrow \psi$, pak existují A , ω takové, že $A \rightarrow \omega$.

Tato gramatika je generativně silnější, neboť umožňuje vytváření vět s tzv. sebezapouštěním (rekurzivní symbol je z obou stran obklopen jinými symboly). Např. $A \rightarrow aAa$, nebo $A \rightarrow aB$, $B \rightarrow Ac$. Protože všechny přirozené jazyky obsahují věty se sebezapouštěním, nemohou být, jak ukazuje Chomsky, adekvátně popsány gramatikami s konečným počtem stavů.

Derivace jednoduché věty „Sestra složila zkoušku“ by měla v gramatice podobné gramatikám Chomského asi tuto podobu:

$$\begin{aligned} S &\rightarrow NP_1 + VP \\ VP &\rightarrow Verb + NP_2 \\ Verb &\rightarrow složila \\ NP_1 &\rightarrow sestra \\ NP_2 &\rightarrow zkoušku \end{aligned}$$

Příklad se sebezapouštěním: „Zpráva, že sestra složila zkoušku, potěšila matku“.

$$\begin{aligned} S &\rightarrow NP_1 + VP \\ VP &\rightarrow Verb + NP_2 \\ NP_1 &\rightarrow NP_1 + C + S \\ Verb &\rightarrow složila, potěšila \\ C &\rightarrow že \\ NP_1 &\rightarrow zpráva, sestra \\ NP_2 &\rightarrow zkoušku, matku \end{aligned}$$

Derivace vět je možno názorně vyjadřovat grafem stromu (obdržíme tzv. strukturní charakteristiku). Pravidla syntaktická jsou doplněna pravidly morfonologickými, která přispívají ke konečnému správnému zápisu věty v přirozeném jazyce.

III. Obsahuje-li frázová gramatika pravidla typu: jestliže $\varphi \rightarrow \psi$, pak existují A , φ_1 , φ_2 , ω takové, že $\varphi = \varphi_1 A \varphi_2$ a $\psi = \varphi_1 \omega \varphi_2$, jde o gramatiku kontextovou, symbol A je možno přepsat na ω jen v kontextu $\varphi_1 \dots \varphi_2$. Touto gramatikou je možno generovat i věty, kde je nutno změnit pořadí symbolů v řetězu. Přestože kontextové gramatiky umožňují tuto změnu, téměř všichni badatelé, kteří vycházejí z frázových gramatik, dávají přednost gramatikám nekontextovým, zejména proto, že kontextové nezachycují dobře gramatické struktury vět.

Avšak Chomsky nepovažuje žádnou z doposud zmíněných postupů za plně adekvátní k popisu přirozeného jazyka. Ve svých pracích přesvědčivě dokazuje, že vytčenému cíli nejlépe odpovídají gramatiky ještě obecnějšího typu, tzv. transformační. Jejich matematický aparát, který není ještě zcela propracován, je příliš složitý, než aby mohl být v tomto článku popsán. Tyto gramatiky obsahují jako jednu ze složek

pravidla frázová, je jich však použito pouze k derivaci terminálních řetězců, které přísluší nejzákladnějším větám. Aplikujeme-li na tyto terminální řetězce některé základní transformace (tzv. obligatorní) a morfonologická pravidla, dospíváme k tzv. větám jádrovým, které obsahují již určitá selektivní omezení. Poslední fází jsou pak transformace fakultativní, které umožňují např. převod vět z aktiva do pasíva, koordinaci, vytváření složitější věty ze dvou jednodušších ap. Dovolují škrtnání prvků, což frázové gramatiky vylučují, i permutace (se zachováním strukturní charakteristiky).

Elementární příklady fakultativních transformací: a) Převod aktiva „Bratr píše dopis“ do pasíva „Dopis je psán bratrem“:

$$\begin{aligned} NP_1 + k_1 + Verb_{tr} + NP_2 + k_4 &\Rightarrow \\ \Rightarrow NP_2 + k_1 + ps + Verb_{tr} + part + NP_1 + k_7 \end{aligned}$$

(k_i = pád, ps = pomocné sloveso, $part$ = koncovka přičestí trpného, $Verb_{tr}$ = přechodné sloveso).

b) Tvoření složitější věty: „Bratr píšící dopis přijde pozdě“.

$$\begin{aligned} NP_1 + k_1 + Verb_{1intr} + Ad \\ NP_1 + k_1 + Verb_{2tr} + NP_2 + k_4 &\Rightarrow \\ \Rightarrow NP_1 + k_1 + A_{verb_2} + NP_2 + k_4 + Verb_{1intr} + Ad . \end{aligned}$$

K určení adekvátnosti gramatik je nutno zkoumat nejen jejich generativní, ale i explikativní sílu; je to požadavek spjatý těsně s lingvistickým zkoumáním strukturní stavby vět, jak již bylo ukázáno při srovnání kontextových a nekontextových gramatik. Otázky týkající se srovnání postupů frázových a transformačních zůstávají dosud otevřené, neboť i transformační gramatiky mají své nevýhody a mnohá pracoviště se doposud vědomě bez nich obešla.

Rekognoskativní procedurou se na rozdíl od Chomského zabývá Y. BAR-HILLEL [3]. Je tvůrcem tzv. kategoriální gramatiky, která má za úkol určovat syntaktické struktury daných řetězců prvků daného jazyka. Vychází z předpokladu, že „každý z konečně mnoha prvků daného přirozeného jazyka má konečně mnoho syntaktických funkcí“, a zabývá se „vytvářením vhodné notace pro tyto funkce (čili kategoriemi) a konstruováním algoritmů na základě notace“. Základem postupu jsou tzv. „škrtačí pravidla“

$$[\varphi_i/\varphi_j], \varphi_j \rightarrow \varphi_i \quad \text{a} \quad \varphi_1[\varphi_i \setminus \varphi_j] \rightarrow \varphi_j,$$

jimiž se redukuje jednotlivá spojení bezprostředních složek věty, až se dospěje ke konečnému tvaru (symbolu pro větu), který potvrdí, že šlo o větu správně gramaticky utvořenou.

Chomsky i Bar-Hillel vycházejí tedy při své práci z tzv. koncepce bezprostředních složek. Koncepce závislostní si také všímá seskupení slov ve větě, ale na rozdíl od koncepce bezprostředních složek jde vždy jen o seskupení dvojčlenné syntaktické dvojice, v němž vždy jeden člen je řídící, druhý závislý. S tímto pojetím pracuje kon-

figurační metoda T. N. MOLOŠNÉ, dále postup HARPERŮV, HAYSŮV, GAIFMANŮV aj. [3]. Obě koncepce spojuje gramatika O. S. KULAGINOVÉ, založená na teorii množin. Jazyk je v tomto pojetí tvořen konečnou množinou slov (slovníkem) s daným systémem okruhů slov (disjunktních podmnožin) a danou množinou správných vět. Cíl Kulaginové je rekognoskativní, jde jí o formulování pravidel rozboru větné struktury, a to zejména z hlediska podkladu pro strojový překlad. Tento přístup popsali u nás A. a M. JAURISOVI [12]. Autory dalších prací tohoto zaměření jsou mj. J. J. REVZIN [13] a S. MARCUS [14]. V nedávné době bylo dokázáno, že gramatiky kategoriální, závislostní, stejně jako tzv. prediktivní analýza (zpracovávaná na Harvardské universitě a založená na poznatku, že každé slovo obsahuje určitou predikci o možnostech syntaktické struktury slov po něm následujících) jsou co do generativní síly ekvivalentní frázovým gramatikám nekontextovým. Tyto závěry shrnuje BAR-HILLEL — viz [3].

Nejzávažnějším nedostatkem všech popsaných systémů je doposud nezpracovaný vztah mezi gramatickou strukturou a sémantikou vět. Objasňování tohoto vztahu zůstává nejaktuálnějším současným úkolem.

3. APLIKOVANÁ LINGVISTIKA

Myšlenka strojového překladu z jazyka do jazyka, která vznikla už ve třicátých letech, získala reálný podklad v souvislosti s rozvojem elektronických samočinných počítačů. O využití těchto strojů pro překlad se začalo uvažovat v USA v r. 1946, kdy došlo k diskusi o hranicích použití automatických počítačích strojů; výměna názorů o možnostech vytvoření vyhovujícího kódu a mechanického slovníku, o otázkách homonymie⁶⁾, nutnosti předredaktora a postredaktora a jiných tehdy zcela neurčitých problémech vyústila v 1. konferenci o strojovém překladu organizované BAR-HILLEM v r. 1952 a konečně v prvním pokusu na počítači v New-Yorku v r. 1954 [15]. V příštím roce došlo k pokusu v Sovětském svazu a potom postupně v řadě dalších zemí.

Je samozřejmé, že stroj, na který je možno se dívat jako na přibližný model činnosti překladatele, postrádá velmi důležitý moment lidské činnosti při normálním překladu — povědomí a zkušenost. Nemůže se spoléhat na nic víc než na čistě formální stránku vstupního textu. Proto musí čelit mnoha potížím vyplývajícím z odlišnosti znakového charakteru jazyka, bezprostředně spojeného s myšlením, od jiných znaků. Nejobtížnějším problémem je homonymie při analýze a synonymie⁷⁾ při syntéze textu. Tyto potíže, které se dosud nepodařilo zcela překonat, vedou k dohadům, zda je vyhovující, výstižný a správný překlad vůbec možný. Plné rozřešení této otázky přinese teprve budoucnost, ale zdá se, že za splnění podmínky vhodného omezení strojového

⁶⁾ *Homonymie* je souzvucnost. Homonyma jsou slova mající stejnou formální podobu, ale odlišný význam.

⁷⁾ *Synonymie* je souznačnost. Synonyma jsou slova se stejným nebo přibližně stejným významem, ale odlišnou formální podobou.

překládání na texty, které nemají velkou stylistickou náročnost, tj. zejména na texty odborné a z nich zase vědecko-technické, může mít strojový překlad při ohromném množství odborné literatury v současné době významnou úlohu. Otázka postředaktora není doposud rozhodnuta.

Algoritmus procesu překladu je do stroje vložen v podobě programu pro analýzu vstupního jazyka a syntézu jazyka výstupního. Tento program je rozčleněn dále na jednotlivé části, které však nemají přesné hranice a jedna zapadá do druhé. Tak např. algoritmus analýzy je možno rozdělit na fáze: vstup textu do stroje, analýza lexikální, frazeologická, morfologická, syntaktická a sémantická. Slovník v podobě slovních základů stejně jako ostatní údaje (slovní spojení, gramatické tabulky apod.) jsou uloženy v paměti stroje. Pro úsporné hledání ve slovníku se zkoušejí různé metody, např. tzv. přímá metoda. Různé typy syntaktické analýzy se staly základem různých pojetí gramatik (metoda bezprostředních složek, závislostní, prediktivní — viz kap. 2). Homonymie se řeší hlavně omezením slovníku na určitý obor a dále pomocí kontextu, tj. výrazů, které předcházejí příslušnému slovu nebo následují po něm.

Nutnost zpracování značného počtu algoritmů analýzy a syntézy při binárních překladech ($n^2 - n$, kde n = počet jazyků) dala vznik myšlenke převodního jazyka (počet algoritmů $2n$). Tento zprostředkující jazyk by měl být zvolen tak, aby co nejlépe odrážel vlastnosti přirozených jazyků s výjimkou homonymie a synonymie [16]. Pro tuto úlohu se nehodí žádný z přirozených jazyků, ani na druhé straně umělý formální jazyk blízký predikátovému kalkulu. Existuje několik návrhů sestavení převodního jazyka, v Sovětském svazu návrh MEL'ČUKŮV (převodní jazyk jako síť vztahů mezi prvky všech přirozených jazyků), zatím nejpropracovanější ANDREJEVŮV (převodní jazyk má svou vlastní gramatiku), u nás SGALLŮV (převodní jazyk s vlastní generativní nekontextovou frázovou gramatikou). Zvláštní důležitosti nabývají v tomto ohledu otázky sémantických slovních druhů a významové výstavby věty. Kromě strojového překladu zabývá se aplikovaná lingvistika dalšími druhy automatického zpracování textu, ať už je to šifrování, automatické třídění, indexování pro dokumentační účely, které spočívá zejména ve zpracování nejpodstatnějších pojmů (hlavně titulů) z určitého textu, automatické referování a pořizování výtahů z odborné literatury na základě slov závažných pro daný obor ap. Některé z těchto úkolů lze provádět i na děrnoštítkových strojích. Objevují se pokusy o sestavování informačních jazyků sloužících k tomu, aby stroj uchovávací informace byl schopen informaci vydat v podobě odpovědi na otázku [5].

V ČSSR se na otázkách matematické lingvistiky začalo pracovat v roce 1959, a to na Karlově universitě, kde dnes existují dvě pracoviště tohoto oboru — lingvistická skupina Centra numerické matematiky při matematicko-fyzikální fakultě a oddělení algebraické lingvistiky a strojového překladu na filosofické fakultě. V roce 1961 vzniklo při Ústavu pro jazyk český ČSAV oddělení matematické a aplikované lingvistiky, které se zaměřuje na lingvistiku kvantitativní (otázky teorie informace, statistická stylistika), zatímco skupina KU se orientuje spíše na algebraickou lingvistiku a její

aplikace. Kromě těchto dvou pracovišť se otázkami matematické lingvistiky zabývají různé skupiny menší, jako skupina MÚ ČSAV (matematická teorie jazyků zpracovávaná K. ČULÍKEM), VÚMS, mimo Prahu Ústav slovenského jazyka v Bratislavě a skupina brněnské university.

Dosavadní i nové výsledky naše i zahraniční je možno sledovat zejména v časopisech Slovo a slovesnost, Acta Universitatis Carolinae a dnes již i v časopise Kybernetika.

Literatura

- [1] DOBRUŠIN R. Z.: Matematické metody v lingvistice. Matematicke prosvetšenie 6 (1961), 37.
- [2] MATHESIUS V.: *Čeština a obecný jazykozpyt*. Praha 1947.
- [3] BAR-HILLEL Y.: *Four Lectures on Algebraic Linguistics and Machine Translation*. Jerusalem 1963.
- [4] TĚŠITELOVÁ M.: K statistickému výzkumu slovní zásoby. Slovo a slovesnost 22 (1961), 171.
- [5] SGALL P. a kolektiv: *Cesty moderní jazykovědy*. Praha 1964.
- [6] SGALL P.: Nové otázky matematických metod v jazykovědě. Slovo a slovesnost 20 (1959), 44.
- [7] ZVEGINCEV V. A.: *Očerki po obščemu jazykoznaniju*. Moskva 1962.
- [8] DOLEŽEL L.: Předběžný odhad entropie a redundance psané češtiny. Slovo a slovesnost 24 (1963), 165.
- [9] CHOMSKY N.: On the Notion "Rule of Grammar". Structure of language and its mathematical aspects. PSAM 12, Providence (1961), 6.
- [10] CHOMSKY N.: *Syntactic Structures*. 's-Gravenhague 1957. (Ruský překlad — sborník Novoje v lingvistike 2, Moskva 1962).
- [11] CHOMSKY N.: Formal properties of grammars (v ruském překladu: O někotorych formal'nych svojstvach grammatik. Kibernetičeskij sbornik 5, Moskva 1962, 279).
- [12] JAURISOVI A. a M.: Užití teorie množin v jazykovědě. Slovo a slovesnost 21 (1960), 34.
- [13] REVZIN I. I.: *Modeli jazyka*. Moskva 1962.
- [14] MARCUS S.: *Linguistică matematică. Modele lingvistice în lingvistică*. Bukurešť 1963.
- [15] LOCKE W. N., BOOTH A. D. (vyd.): *Machine Translation of Languages*. New York 1955 (ruský překlad Mašinnyj perevod, Moskva 1957).
- [16] SGALL P.: Převodní jazyk a teorie gramatiky. Slovo a slovesnost 24 (1963), 114.

GALVANOMAGNETICKÉ JEVY V POLOVODIČÍCH A JEJICH TECHNICKÝ VÝZNAM

LADISLAV ŠTOURÁČ, Praha

Studium transportních jevů v polovodičích zaujímá v současné fyzice pevných látek významné místo. Tyto jevy, které jsou spojeny s pohybem elektronů a fononů v polovodičích za přítomnosti vnějších polí, mají však i značný technický význam. Mnohé transportní jevy se staly základem pro konstrukci polovodičových součástek a čidel