

# **PLIN041 Vývoj počítačové lingvistiky**

## *Algebraická lingvistika*

Mgr. Dana Hlaváčková, Ph.D.  
od 2. pol. 50. let 20. st.

# Algebraická lingvistika

matematika – algebra



logika – formální logika



logická analýza jazyka



algebraická lingvistika, formální lingvistika

# Algebraická lingvistika

- **algebra**
  - z arabštiny – *znovuspojení rozbitých částí*, perský matematik a astronom Al-Chorezmí, 8./9. st., v angličtině až v 15. st.
  - objevuje se už u Babylóňanů
- část matematiky, která studuje matematické **symbols** a **operace** s nimi x aritmetika (čísla)
- nepoužívá kvantitativní metody (statistiku, pravděpodobnost)
- rozvoj matematiky a logiky = vysoký stupeň abstrakce

# Algebraická lingvistika – kontext

- hledání jazyka vědy, metajazyk, 2. pol. 19. st. – přesný a jasný
  - *přirozený jazyk* – nevhodný např. pro nejasnost, nelogičnost, homonymii, synonymii
  - **matematická (symbolická) logika**, symbol zastupuje určitý jev, kalkul = soustava pravidel operací se znaky
    - logické kalkuly (Leibnitz, 17. st.)
    - systém – algebra logiky (booleovská algebra, Boole, 19. st.)
    - výroková a pravdivostní fce, kvantifikátor (Frege, 19. st.)
    - Principia Mathematica (B. Russell, poč. 20. st., 3 díly)
- **Ludwig Wittgenstein** – *Tractatus logico-philosophicus*, 1921
  - úkolem filozofie je logická analýza jazyka
  - **novopozitivismus** – Vídeňský kruh, **R. Carnap** – *Logická syntax jazyka*, 1934

# Algebraická lingvistika

- **formální logika** (v SSSR odsouzena, později akceptována)
  - studuje podmínky, za nichž jedny výroky vyplývají z druhých, používá symboly a abstrakce
  - *výrokový kalkul* – konjunkce, disjunkce, implikace, ekvivalence
  - *predikátový kalkul* – predikát, argument, operátor, kvantifikátor
- **algebraická lingvistika** – Y. Bar-Hillel (izraelský matematik a lingvista)
  - *teorie jazykových modelů* v Sovětském svazu
- dnes **formální lingvistika** (formální popis gramatik a jazyků)
- význam pro **strojový překlad** a **strojovou lingvistiku**
- **vše, co je v lingvistice přesně popsáno, lze strojově zpracovat**

# Matematické modely v lingvistice

- *modely* – využívají se pro popis reálných jevů/objektů, jejichž studium je nesnadné
- *jazyková univerzália* (společné vlastnosti)
  - převodní jazyk
- popis gramatiky jednoho jazyka
- popis jedné složky jazyka (syntax)
- popis komunikačního procesu
- model vývoje jazyka
  - věta – lineární řetězec jednotek
  - teorie množin
  - teorie grafů – nelineární pojetí
- *dnes mezinárodní projekt Universal Dependencies*

# Matematické modely v lingvistice (syntax, morfologie)

- **Noam Chomsky** – generativní a transformační mluvnice
- **Yehoshua Bar-Hillel** – kategoriální (rekognoskativní) mluvnice
- **Sebastian Konstantinovič Šaumjan** – aplikačně generativní model
- **Olga Sergejevna Kulagina** – teorie analytických modelů
- závislostní gramatika a teorie grafů (syntax)

# Matematické modely v lingvistice

- **generativní a transformační mluvnice** – **Noam Chomsky** (\*1928), americký lingvista
- nejpropracovanější systém
- *Syntaktické struktury*, 1957 (česky 1966)
- nejdříve ignoruje obsah, později přijímá i složku sémantickou
- generování gramaticky správných vět z výchozího **symbolu** na základě souboru **pravidel** = gramatika jazyka
- omezený počet pravidel – neomezený počet vět
- jádrové věty – ostatní se z nich generují na základě transformačních pravidel
- ve druhém pojetí, fonetická složka, povrchová a hloubková struktura
- formální gramatika, hierarchie formálních jazyků, teorie automatů



# Matematické modely v lingvistice

- **Y. Bar-Hillel** – kategoriální (rekognoskativní) gramatika
- nejvýznamnější žák Rudolfa Carnapa
- opačný postup než u Chomského
- v souvislosti se strojovým překladem (pro jazyky s málo rozvinutou morfologií a pevným slovosledem – angličtina)
- pojem – *kategorie* (sémantická kategorie v logice a filozofii)
- *S* – sentence, *N* – noun (neterminály)
- množiny slov (terminály)
- tvary slov nahrazují **symboly**, věta = **řetězec symbolů**
- zjišťuje se (rekognoskuje) struktura věty a její gramatická správnost
- bezkontextová gramatika

# Matematické modely v lingvistice

- **Sovětský svaz** – v 1. pol. 20. st. vliv marxismu a tzv. *marrismu* (*Nikolaj Marr – odmítá vývoj ide. jaz. z prajazyka, zavrhl J. V. Stalin*)
- řada lingvistických směrů odmítána
- ideologizace lingvistiky, lingvistika v SSSR v izolaci
- strukturalismus až v 50. letech (Vinogradov, Šaumjan)
- přechod k pomezním disciplínám
- rozvoj algebraické lingvistiky a strojového překladu
- aplikace jazykovědných teorií na ruštinu + nové přístupy
- část lingvistů v emigraci (Jakobson, Trubeckoj)

# Matematické modely v lingvistice

- **Sebastian Konstantinovič Šaumjan (1916–2007)**
  - původem Armén, pocházel z Tbilisi (Gruzie)
  - mluvil arménsky, gruzínsky, rusky, anglicky a německy
  - vystudoval filologii na univerzitě v Tbilisi
  - 2. sv. v., zpravodajská jednotka, člen KSSS (pomáhal kolegům, kteří byli v nemilosti)
  - pozice na univerzitě v Moskvě
  - podporoval dílo Jakobsona a Trubeckého
  - **1975** – povolená emigrace židů (Izrael) do USA (Yale University)
  - *...ever smiling but never complaining...*
  - *geniální lingvista a člověk s velkým srdcem*

# Matematické modely v lingvistice

- **S. K. Šaumjan** – aplikačně-generativní model
- *Generative Grammar of Russian, 1958*
- *Applicational Generative Model and Transformational Calculus of Russian, 1963*
- *Strukturnaja lingvistika, 1965*
- **spojení strukturalismu a generativní gramatiky**
- jazykové jednotky (symboly) a vztahy mezi nimi se odvozují metodami matematické logiky, generují se pomocí *aplikace*
- univerzální typy – *term (T)*, *sentence (S)*, sloveso = operátor
  - *genotypický* (univerzální jazyk)
  - *fenotypický* jazyk (přirozený jazyk)
- model použit pro parsing u strojového překladu

# Matematické modely v lingvistice

- **Olga Sergejevna Kulagina – teorie analytických modelů**  
(analytická metoda)
  - Chomsky, Bar-Hillel, Šaumjan – syntetické modely
- *O jednom způsobu určování gramatických pojmů na základě teorie množin, 1958*
- pro **slovanské jazyky** – rozvinutá morfologie a volný slovosled
- **teorie množin** (1. využití v lingvistice)
  - výchozí množina = gramaticky správné věty
  - podmnožiny = soubory základních jednotek (lexikologie, morfologie, syntax)

# Matematické modely v lingvistice

- **Solomon Marcus** (rumunský lingvista a matematik, dostává se až na úroveň fonémů)
  - *Matematická lingvistika*, 1963
  - *Gramatika a konečné automaty*, 1964
  - (u nás pod názvem *Algebraické modely v lingvistice*, 1969)
- **Ladislav Nebeský** (teorie grafů, **binární básně**)
- *Analytický směr v algebraické lingvistice* (SaS 1967, s. 161–7)

# Matematické modely v lingvistice

- **závislostní gramatika a teorie grafů** (nelineární zobrazení věty, graf)
- nezávisle na sobě konstruují závislostní gramatiku:
  - **američtí lingvisté** – David G. Hays, K. E. Harper; strojový překlad (*Použití strojů při konstruování gramatiky*, 1959)
  - **sovětští lingvisté** – D. S. Cejtin, L. N. Zatorina (*O vyčlenění konfigurací v ruské větě*, 1961)
- **závislostní syntax** (druhotně morfologie, jiné jazykové roviny ne)
  - závislostní pravidla
  - závislostní strom – uzel a hrany
  - projektivní a neprojektivní věty (*dálnice ucpaná auty x auty dálnice ucpaná*)

# Matematické modely v lingvistice

- myšlenka **závislostních vztahů** v jazyce již dříve
  - L. Tesnière – slovesná valence, aktanty a cirkumstanty
  - u nás V. Šmilauer – grafické znázornění větného rozboru, Daneš (syntax), Dokulil (slovotvorba)
- ČSSR – **funkční generativní popis, FGD** (Panevová, Sgall)
  - Prague Dependency Treebank, PDT