

LINGWISTYKA MATEMATYCZNA I TEORIA AUTOMATÓW

podręcznik dla studentów kierunków:
etnolingwistyka i językoznawstwo i informacja naukowa

Włodzimierz Lapis
Instytut Lingwistyki UAM
ul. Międzychodzka 3/5
60-371 POZNAŃ

e-mail: lapisw@amu.edu.pl

POZNAŃ 2004

Ludzie! Najpiękniejsza rzecz na świecie, to uczyć się i poznawać świat. Róbcie wszystko, żeby wam belfry tego nie obrzydziły. Jak ktoś przestanie się uczyć, poznawać, to już koniec, starość. Już nic nowego nie zobaczy. Nawet jak się ma te gówniane trójki, nie przejmować się, robić swoje i kochać tę matkę niezależnie od baby, która jej uczy i chce ci ją obrzydzić.

Marek Piekarczyk
(lider TSA)

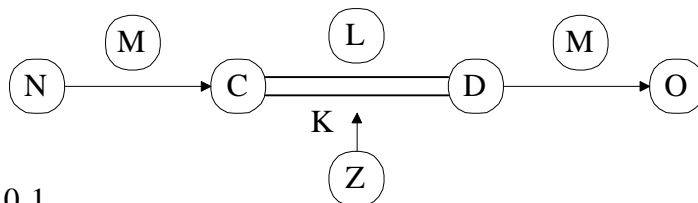
SPIS TREŚCI

Przedmowa	4
I Podstawowe pojęcia i ogólne twierdzenia lingwistyki matematycznej	6
1. Wyrażenia, języki, gramatyki	6
2. Hierarchia Chomsky'ego	13
3. Operacje na językach. Zamkniętość klas języków ze względu na operacje regularne	14
4. I twierdzenie o postaci normalnej	15
II Gramatyki bezkontekstowe	18
1. Gramatyki λ - wolne	18
2. Postać normalna Chomsky'ego	20
3. Drzewa derywacji	23
4. Rozstrzygalność problemów. Twierdzenie o pompowaniu	24
5. Postać normalna Greibach	32
6. Jednoznaczność i wieloznaczność wyводу	32
7. Semantyczne i syntaktyczne zagadnienia języka naturalnego	36
8. Notacja Backusa-Naura	42
III Gramatyki regularne	44
1. Postać normalna gramatyk regularnych	44
2. Wykresy gramatyk regularnych	47
3. Wyrażenia regularne	50
4. Języki regularne, a języki liniowe i bezkontekstowe. Gramatyki samorozszerzające	62
5. Gramatyki deterministyczne	68
IV Gramatyki kontekstowe	73
1. Monotoniczność gramatyk	73
2. Postać normalna Kurody	75
V Gramatyki struktur frazowych	77
1. Postać normalna	77
2. Drzewa derywacji	79
VI Automaty skończenie-stanowe	81
1. Podstawowe definicje	81
2. Dwa twierdzenia Kleene'go	86
3. Twierdzenie Scotta i jego konsekwencje	89
VII Automaty ze stosem	96
1. Podstawowe definicje	96
2. Automaty z pustym stosem	102
3. Automaty ze stosem, a gramatyki bezkontekstowe	105
VIII Maszyny Turinga	107
1. Definicja maszyny Turinga	107
2. Akceptowalność języków przez maszynę Turinga	109
3. Funkcje sygnalizujące i złożoność problemu akceptacji języków	112
4. Rozstrzygalne i nierozstrzygalne problemy lingwistyki	115
5. Dodatkowe uwagi i spostrzeżenia	118
Bibliografia	119

PRZEDMOWA

Niniejszy podręcznik powstał jako pomoc dydaktyczna do prowadzenia jedno-semestralnego konwersatorium na kierunkach „etnolingwistyka” i „językoznawstwo i informacja naukowa”, odbywanego w ramach 30-godzinnego kursu przedmiotu „lingwistyka matematyczna i teoria automatów”.

W opracowaniu tym przedstawione zostały zagadnienia wchodzące w zakres szeroko rozumianej teorii języków formalnych. Celem omówienia zakresu jej przedmiotu, przyjrzymy się poniższemu schematowi układu komunikacyjnego (UK) obrazującemu środowisko i proces przekazywania i odbierania informacji.



Rys. 0.1.

Tak więc pewien nadawca (N) po zakodowaniu w koderze (C) pewnego komunikatu (M) na pewien język (L) nadaje go jakimś kanałem (K), gdzie ewentualnie może on ulec zniekształceniom (Z). W takiej formie, po ponownym przetłumaczeniu (tym razem w dekodерze (D)) - przyjmowany jest on przez odbiorcę (O). Jeśli wszystko odbywa się w warunkach normalnych, tj. zakłócenia (Z) są niewielkie i nadawca oraz odbiorca nastrojeni są na jeden kod (ten sam język), to wówczas informacja nadana przez N jest w identycznej formie odebrana i prawidłowo zrozumiana przez O (UWAGA: celem lepszego zrozumienia omawianych tu pojęć, spróbuj określić elementy UK w przypadkach: bezpośrednia rozmowa ludzi, rozmowa telefoniczna).

W naszych rozważaniach nie będziemy zajmować się zagadnieniami rodzaju kanału przekazu informacji, czy też doznawanych przez nie zakłóceń. Prawie wcale nie będziemy zajmować się również konkretnymi UK. Na dobrą sprawę nie będzie nas też interesować nadawca ani odbiorca. Właściwy zakres naszych rozważań został zaprezentowany w tabeli 0.1.

Ponadto, przez cały kurs przewijać się będą zagadnienia rozstrzygalności poszczególnych problemów teorii języków formalnych, a przez kurs lingwistyki matematycznej - dodatkowo twierdzenia o postaciach normalnych poszczególnych gramatyk. Będziemy także mówić o zagadnieniach determinizmu i zupełności tak gramatyk, jak i automatów.

Skupimy się tu mianowicie głównie na opisie	Zagadnienia te wchodzą w zakres odpowiednio	Konkretnie omawiać tu będziemy
koderów (którymi będą dla nas gramatyki formalne) oraz generowanych przez nie języków	lingwistyki matematycznej (rozdz. I – V)	poszczególne gramatyki i generowane przez nie języki (regularne, liniowe, kontekstowe i bezkontekstowe oraz struktur frazowych), jak również ich własności i wzajemne zależności
jak również akceptujących je dekodeerów (którymi będą dla nas automaty)	teorii automatów (rozdz. VI - VIII)	automaty skończone i ze stosem oraz maszyny Turinga, jak i ich związek z w/w językami

Tab. 0.1.

Zawarty w opracowaniu tym materiał wzbogacony został wieloma przykładami (w tym także odnoszącymi się do języka naturalnego), co ma służyć ułatwieniu jego zrozumienia oraz przyswajania.

W książce stosuje się ogólnie przyjętą symbolikę matematyczną. Zwróćmy tylko uwagę na fakt, że zakończenie zadania, przykładu, koniec dowodu twierdzenia, czy też koniec twierdzenia z pominiętym dowodem oznaczamy symbolem „□”, a często stosowany skrót „witw” należy czytać „wtedy i tylko wtedy”. Ponadto symbol „⊆” oznacza inkluzję (zawieranie się) zbiorów, a symbol „⊂” właściwe zawieranie się zbiorów. Symbole „[” i „]” należy traktować odpowiednio jako otwierający i zamykający nawias obejmujący swym zasięgiem poszczególne formuły, w przypadkach których wskazujemy ich przynależność do pewnego zbioru. Ponadto przyjmujemy zawsze, że zbiór liczb naturalnych $N = \{1, 2, 3, \dots\}$, a $N_0 = N \cup \{0\}$.

W czasie kursu czytelnik będzie spotykać się z pojęciami „problemy rozstrzygalne” i „problemy efektywnie rozstrzygalne”. Przez problemy rozstrzygalne rozumiemy te problemy, które jesteśmy w stanie rozstrzygnąć w przeliczalnej liczbie kroków, a przez problemy efektywnie rozstrzygalne - te z nich, które jesteśmy w stanie rozstrzygnąć w skończonej liczbie kroków.

W każdym z rozdziałów przyjęto niezależną numerację twierdzeń, faktów, przykładów i zadań, a także tabel, rysunków i diagramów. Numeracja wniosków jest niezależna dla każdego z twierdzeń, którego dotyczy.