



AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE

Wprowadzenie

Teoria automatów i języków formalnych

Dr inż. Janusz Majewski
Katedra Informatyki



Literatura (1)

1. Aho A. V., Sethi R., Ullman J. D.: *Compilers. Principles, Techniques and Tools*, Addison-Wesley, 1986 (tłumaczenie polskie: *Kompilatory. Reguły, metody i narzędzia*, WNT 2002).
2. Aho A. V., Ullman J. D.: *The Theory of Parsing, Translation and Compiling, vol. 1*, Prentice-Hall, 1972 (tłumaczenie rosyjskie: Ахо А., Ульман Дж.: *Теория синтаксического анализа, перевода и компиляции*, Издательство „Мир”, 1978)
3. Blikle A.: *Automaty i gramatyki*, PWN, 1971
4. Blikle A.: *Wybrane zagadnienia lingwistyki matematycznej*, w: *Problemy przetwarzania informacji, tom 2, praca zbiorowa pod red. A. Mazurkiewiczza*, WNT 1974.
5. Gries D.: *Compiler Construction for Digital Computers* (jest tłumaczenie polskie).
6. Homenda W.: *Elementy lingwistyki matematycznej i teorii automatów*, Oficyna Wydawnicza Politechniki Warszawskiej 2005.



Literatura (2)

7. Hopcroft J. E., Motwani R., Ullman J. D.: Wprowadzenie do teorii automatów, języków i obliczeń, PWN, 2005.
8. Hopcroft J. E., Ullman J. D.: Wprowadzenie do teorii automatów, języków i obliczeń, PWN, 1994
9. Kowalski St., Mostowski A. Wł.: Teoria automatów i lingwistyka matematyczna, PWN, 1979.
10. Ross K. A., Wright C. R. B.: Matematyka dyskretna, PWN, 1996.
11. Sipser M.: Wprowadzenie do teorii obliczeń, WNT 2009.
12. Waite W. M., Goos G.: Konstrukcja kompilatorów, WNT, 1989.



Lingwistyka, języki

Dziedzina wiedzy:

Lingwistyka matematyczna

Czym będziemy się zajmować:

Opisywać języki

- Opis języka (semiotyka)
- (a) syntaktyka (opis składni)
- (b) semantyka (znaczenie)
- (c) pragmatyka (użyteczność)

Jaki aspekt języka będziemy opisywać?

Składnia – zasady budowy poprawnych wypowiedzi w języku (tylko poprawnych; nie bierze się pod uwagę aspektu znaczeniowego i pragmatycznego, nie analizuje się znaczenia wypowiedzi, a tym bardziej prawdziwości lub fałszu wypowiedzi)

Jakie języki będziemy opisywać?

Języki:

- (a) naturalne (np. polski, angielski, chiński,...)
 - (b) formalne** (sztuczne, np. Pascal, C, Algol,...)
- Zajmujemy się wyłącznie opisem języków formalnych.



Przykład (1)

Przykład: (notacja BNF)

```
<zdanie> ::= <grupa_podmiotu> <grupa_orzeczenia>
<grupa_podmiotu> ::= <fraza_rzeczownikowa>
<grupa_orzeczenia> ::= <fraza_czasownikowa>
<fraza_rzeczownikowa> ::= <przymiotnik> <fraza_rzeczownikowa>
| <rzeczownik>
<fraza_czasownikowa> ::= <czasownik> <fraza_przysłówkowa>
| <czasownik> <fraza_rzeczownikowa> | <czasownik>
<fraza_przysłówkowa> ::= <przyimek> <fraza_rzeczownikowa>
| <przysłówek>
| <przysłówek> <przyimek> <fraza_rzeczownikowa>
<przymiotnik> ::= szybki | rudy | omszały | głęboki | mądry
<rzeczownik> ::= lis | pień | dół
<przysłówek> ::= zgrabnie | bardzo
<przyimek> ::= przez
<czasownik> ::= przeskoczył | polubił
```



Przykład (2)

Poprawne zdania:

szybki rudy lis przeskoczył przez omszały pień

mądry lis polubił głęboki omszały dół

głęboki lis przeskoczył przez mądry dół

szybki głęboki pień polubił przez mądry rudy dół

lis przeskoczył zgrabnie przez głęboki dół

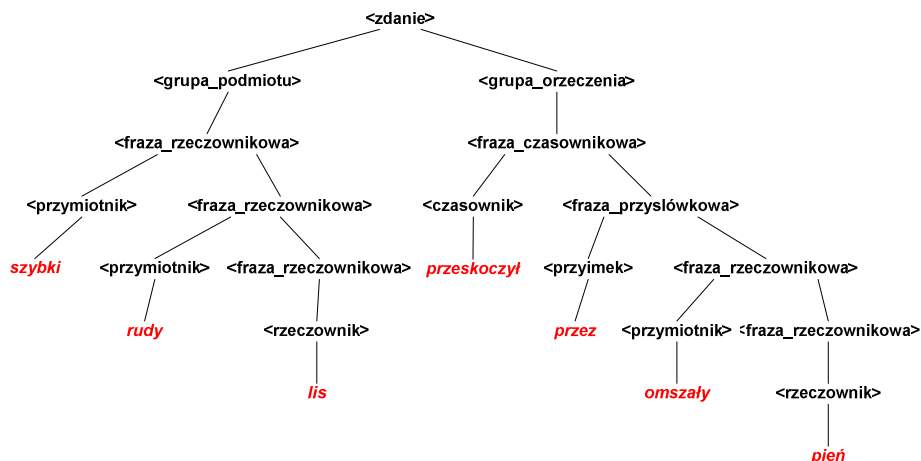
Niepoprawne zdania:

lis przeskoczył bardzo zgrabnie przez głęboki dół (nie można użyć dwóch przysłówków w jednym zdaniu)

szybki rudy wilk przeskoczył przez omszały pień (rzeczownik *wilk* nie występuje w definicji języka)

polubił głęboki omszały dół (w zdaniu musi wystąpić <grupa_podmiotu>)

Przykład (3)



BNF

Do opisu języka trzeba używać *metajęzyka*. Powyższa przykładowa definicja języka została stworzona w tzw. *notacji Backusa-Naura*. W zapisie tym elementami metajęzyka są:

<zdanie>, <grupa_podmiotu>, ... — zwane *symbolami nieterminalnymi*, *symbolami pomocniczymi* lub *zmiennymi językowymi*,

::= — *operator podstawienia metajęzykowego*,

| — symbol metajęzykowy „lub”.

Elementami języka są zaś tzw. *symbole terminalne (końcowe)* jak np.:
szybki, rudy, lis.

Zapis typu: <...> ::= ... nazywamy *produkcją* lub *regułą syntaktyczną*.

Zapis typu:

<AAA> := aaa | bbb

jest równoważny dwóm produkcjom

<AAA> := aaa

<AAA> := bbb



Gramatyka języka formalnego

Do pełnej definicji języka musimy określić:

V – zbiór symboli nieterminalnych, np.: $V = \{ \langle \text{zdanie} \rangle, \langle \text{grupa_podmiotu} \rangle, \dots \}$

Σ – zbiór symboli terminalnych, np.: $\Sigma = \{ \text{szybki, rudy, lis,} \dots \}$

P – zbiór produkcji np.: $P = \{ \langle \text{zdanie} \rangle ::= \langle \text{grupa_podmiotu} \rangle \langle \text{grupa_orzeczenia} \rangle, \dots \}$

$S \in N$ – symbol początkowy od którego zawsze rozpoczynamy wyprowadzanie napisów języka, u nas $S = \langle \text{zdanie} \rangle$.

Te cztery kategorie łącznie noszą nazwę *gramatyki* języka formalnego.



Antynomia Grelinga (1)

Dlaczego do opisu języka musimy używać jakiegoś innego języka (metajęzyka)? Problem ten ilustruje tzw. antynomia Grelinga, która stanowi „dowód” na to, że trzeba pojęciowo odróżniać język od metajęzyka.

Przypuśćmy, że język L (zbiór pewnych słów) sam się opisuje, tzn. że można w nim opisać jego słowa. Dokonajmy podziału słów tego języka na dwie klasy:

słowa autosemantyczne, które orzekają o sobie

$a \in L$ jest autosemantyczne $\Leftrightarrow a$ jest a

słowa heterosemantyczne, które nie orzekają o sobie

$a \in L$ jest heterosemantyczne $\Leftrightarrow \neg (a$ jest $a)$

Przykłady słów autosematycznych:

„rzeczownik” jest rzeczownikiem

„osiemnastoliterowe” jest osiemnastoliterowe

Przykłady słów heterosemantycznych:

„czasownik” nie jest czasownikiem (bo jest rzeczownikiem),

„pięcioliterowe” nie jest pięcioliterowe (bo jest czternastoliterowe)



Antynomia Grelinga (2)

Niech $h \in L$, $h =$ „heterosemantyczny”

Do której z dwóch klas należy słowo h ?

Założmy, że h jest heterosemantyczne:

(h jest heterosemantyczne) \Rightarrow

(„heterosemantyczne” jest heterosemantyczne) \Rightarrow

(h orzeka o sobie) \Rightarrow (h jest autosemantyczne)

(SPRZECZNOŚĆ)

Założmy, że h jest autosemantyczne:

(h jest autosemantyczne) \Rightarrow

(„heterosemantyczne” jest autosemantyczne) \Rightarrow

(h nie orzeka o sobie) \Rightarrow (h jest heterosemantyczne)

(SPRZECZNOŚĆ)

Tak więc napotykamy istotne trudności chcąc dany język opisać jego własnymi słowami.