

# JEZYKOZNAWSTWO OGÓLNE

Kognitywistyka UAM, rok II

## WYKŁAD 9: SKŁADNIA

JERZY POGONOWSKI

ZAKŁAD LOGIKI I KOGNITYWISTYKI UAM

### 1 Uwagi wstępne

#### 1.1 Osobista refleksja

1. Wedle informacji uzyskanych przez wykładowcę od Dyrekcji IP UAM, wykład *Językoznawstwo ogólne* został włączony do programu studiów na drugim roku na kognitywistyce na życzenie studentów. Pełni on rolę niejako służebną wobec wykładu *Językoznawstwo kognitywne* (trzeci rok).
2. Frekwencja podczas wykładu 28 kwietnia 2016 roku wyniosła 2,56 procenta.
3. Brak zainteresowania studentów wykładem nie ma żadnego znaczenia dla jego kontynuacji. Wykładowca wykona zlecenie, które otrzymał od Wydziału Nauk Społecznych UAM.
4. Ów brak zainteresowania wykładem obliguje jednak wykładowcę do przeprowadzenia egzaminu z tego przedmiotu ze szczególną uwagą, starannością, pieczołowitością i troską o wystawienie studentom ocen adekwatnie odpowiadających ich wiedzy oraz umiejętności jej wyrażenia. Egzamin jest przecie nagrodą: zdając każdy egzamin student potwierdza swoje aspiracje intelektualne, cieszy się z możliwości wykazania się wiedzą, udowadnia sobie oraz innym, że jest wartościową jednostką.

#### 1.2 Przedmiot składni

O przedmiocie składni tak pisze się w klasycznym opracowaniu Grochowski, Karolak, Topolińska 1984 (strona 11):

Przedmiotem składni jest opis relewantnych właściwości **wyrażeń złożonych** w poszczególnych językach naturalnych, tzn. własności istotnych z punktu widzenia sposobów ich tworzenia z wyrażeń prostych i ich funkcjonowania w aktach komunikacji. Składnia jest zatem teorią (częścią teorii języka), która formułuje reguły tworzenia wyrażeń złożonych (kompozycji) o różnym stopniu owej złożoności lub opisuje pewne właściwości regularnie występujące w wyrażeniach złożonych. Innymi słowy, składnia jest zbiorem reguł (prawideł, formuł) odzwierciedlających prawidłowości łączenia się (kombinatoryki) **wyrażeń prostych** w wyrażenia złożone.

Dodajmy do tej charakterystyki kilka ogólnych uwag:

1. Zdaniem piszącego te słowa geneza składni należy do największych zagadek lingwistycznych.
2. Rozumienie języka oraz sprawne posługiwanie się nim zakłada znajomość reguł składniowych. Nie jest przecież tak, że uczymy się *na pamięć* wszystkich zdań, których używamy w komunikacji. Tak nauczyć się można znaczenia nowo poznanego wyrazu lub morfemu. W przypadku zdań jest inaczej: to znajomość reguł składniowych pozwala nam tworzyć coraz to nowe wypowiedzi.
3. Poziom składniowy to pierwszy z omawianych na wykładzie poziomów językowych, na którym występuje *potencjalnie nieskończona* liczba jednostek – zdań prostych oraz złożonych. To, że jest potencjalnie nieskończenie wiele zdań prostych widać chociażby na przykładzie ogólnego schematu *Partia będzie rządziła przez n lat*, gdzie za  $n$  podstawić można dowolny liczebnik. Dowolne dwa zdania można połączyć np. koniunkcją, tworząc zdanie, a więc również zdań złożonych jest potencjalnie nieskończenie wiele.
4. Podstawowa kategoria składniowa (syntaktyczna) to *zdanie*. Charakterystyka tej jednostki językowej może brać pod uwagę: własności syntaktyczne, własności semantyczne, własności prozodyczne.

Na samym początku dzisiejszego wykładu konieczne jest wyraźne stwierdzenie, że podajemy tu niezwykle uproszczoną wizję składniową. Współczesne teorie składniowe (a jest ich niezliczone mnóstwo) to koncepcje niezwykle wyrafinowane pojęciowo. Ich przedstawienie wymagałoby omówienia szeregu niezbędnych preliminariów, na co nie możemy sobie pozwolić w wykładzie o charakterze pomocniczym. Podkreślamy zatem, że dzisiejsza prezentacja jest brutalnie uproszczonym

obrazem zależności składniowych, w dodatku dość tradycyjnym. Zdajemy sobie sprawę, że niektóre z wprowadzanych dalej terminów nie są opatrzone w precyzyjne definicje, ale są jedynie pośrednio charakteryzowane.

Zachęcamy słuchaczy do obejrzenia niedużej prezentacji Pana dra Konrada Juszczyka *Składnia – przegląd pojęć*:

<http://www.slideshare.net/Linguist/skadnia-przegld-poj2012>

W książce Grzegorzycowa 2007 na stronie 108 autorka jako podstawowe czynniki w tworzeniu wypowiedzi wymienia:

1. *Wybór określonego schematu zdania (będący konsekwencją wyboru głównego czasownika) lub wypowiedzi niezdaniowej.*
2. *Przyjęcie określonej postawy modalnej nadawcy i wybór odpowiedniego jej wykładnika.*
3. *Zastosowanie zasad rozwijania grup i podporządkowywania sobie członów składniowych, sygnalizowanego przede wszystkim za pomocą form fleksyjnych.*
4. *Różnorakie przekształcenia schematów.*
5. *Stosowanie zasad budowy zdań złożonych.*

## **2 Kategorie syntaktyczne**

Słuchacze wynieśli ze szkoły wiedzę na temat kategorii składniowych proponowanych dla języka polskiego. Wymieniano z pewnością:

1. *Części zdania.* Patrzymy na rysunek.
2. *Podmiot.* Patrzymy na rysunek.
3. *Orzeczenie.* Patrzymy na rysunek.
4. *Dopełnienie.* Patrzymy na rysunek.
5. *Okolicznik.* Patrzymy na rysunek.
6. *Przydawka.* Patrzymy na rysunek.
7. *Rodzaje zdań podrzędnie złożonych.* Patrzymy na rysunek.
8. *Zdania bezpodmiotowe.* Zmierzcha się.

### 3 Relacje syntaktyczne

W wyrażeniach złożonych pewne ich części mogą *określać* inne, np. wyraz *mały* w *mały pies* określa wyraz *pies*. Ta relacja określania pozwala na wydzielanie pewnych konstrukcji złożonych z wyrazów, ma także określone własności semantyczne. Proszę zauważyć, że *mały pies* jest psem, a jeśli ktoś *zdał egzamin na piątkę*, to oczywiście *zdał egzamin*. W pewnych przypadkach występują, powiedzmy, kolizje między składniowymi a semantycznymi własnościami tej relacji (np. *rzekomy ojciec* niekoniecznie jest *ojcem*, *były kolega* już nie jest *kolegą*).

1. *Walencja*. To liczba argumentów, które przyjmuje czasownik. Mając na myśli tę liczbę, mówi się także o własnościach *konotacyjnych* czasownika.
2. *Hipotaksa*. To, najogólniej mówiąc, *podrzędność* składniowa.
3. *Parataksa*. To, najogólniej mówiąc, *równorzędność* składniowa.
4. *Związek zgody*. Ta relacja (nazywana też *kongruencją*) zachodzi między dwoma elementami tworzącymi konstrukcję w zdaniu, gdy oba zgadzają się co do pewnych informacji gramatycznych (w tym sensie, że oba je posiadają). W wyrażeniu *mały pies* oba elementy zgadzają się co do: przypadku, liczby, rodzaju.
5. *Związek rzędu*. Ta relacja nazywana jest też *rekcją* i zachodzi między dwoma elementami zdania, gdy znaczenie gramatyczne jednego z nich wymusza na drugim przyjmowanie określonych znaczeń gramatycznych. Mówi się np., że czasowniki *rzędzą* wyrażeniami, pełniącymi funkcję dopełnień (np. *pacjens* zdania z czasownikiem przechodnim niesie w języku polskim informację biernika). Związkiem rzędu mogą też być połączone dwa rzeczowniki: w wyrażeniu *wycie skazańca* drugi człon w języku polskim niesie informację dopełniacza (i tak być musi w polskim).
6. *Związek przynależności*. Ta relacja, zwana też *adiunkcją leksykalną* zachodzi między tym elementami zdania, z których jeden (zwykle czasownik) jest nadrzędny składniowo wobec całej grupy, a drugi jest, mówiąc językiem szkolnym, nieodmienny. Dla przykładu: *Teraz zaśpiewam*.
7. *Akomodacja syntaktyczna*. Wszelkie wymagane sygnały zależności składniowej.
8. *Temat i remat*. Temat to to, o czym w danym zdaniu się mówi, zaś remat to to, co mówi się o temacie. Niektórzy zamiast *temat* i *remat* używają terminów *topic* oraz *comment*, odpowiednio. Zaznaczanie tematu w różnych języ-

kach może wyglądać różnie: używa się do tego szyku, intonacji, specjalnych partykuł, itd.

9. *Datum i novum*. Datum to informacja już znana, novum to informacja nowa.

## 4 Konstrukcje syntaktyczne

Konstrukcje składniowe mogą być *endocentryczne* (gdy cała konstrukcja pełni identyczną funkcję składniową jak dokładnie jeden z jej członów) lub *egzocentryczne* – gdy np. cała (powiedzmy, dwa wyrazy połączone spójnikiem) konstrukcja pełni tę samą funkcję składniową co każdy z jej członów (w tym przypadku: dwóch) oddzielnie. *Mały pies* jest konstrukcją endocentryczną, a *prawo i sprawiedliwość* jest konstrukcją egzocentryczną.

1. *Schemat składniowy zdania*. W charakterystyce tego pojęcia odwołam się do cytatu (Grzegorzyczkowa 2007, 112):

**Schemat zdania** to uogólniona postać zdania empirycznego, w której abstrahuje się od składu leksykalnego zdania, od budowy wewnętrznej grup, od porządku linearnego, od członów luźnych. Opiera się on na właściwościach konotacyjnych czasownika.

2. *Składniki bezpośrednie*. W zdaniach wyróżnić możemy ich komponenty, które z kolei możemy dalej rozczłonkować, aż do dotarcia do elementów syntaktycznie prostych. Podstawowy podział zdania to (w tym ujęciu) podział na grupę podmiotu oraz grupę orzeczenia. Dalsze składniki bezpośrednie są (z reguły) konstrukcjami endocentrycznymi. Zasady owego dzielenia wyznaczone są przez akomodację syntaktyczną. Patrzymy na rysunki.
3. *Drzewa składników bezpośrednich*. Zdaniu możemy przyporządkować jego *drzewo składników bezpośrednich* (takie, jakie ukazano na rysunkach).
4. *Drzewa zależności*. W tej koncepcji składniowej nadrzędnym elementem zdania jest jego czasownik. Traktowany jest on zatem tak, jak predykat w języku logiki pierwszego rzędu. Posiada zatem stosowną liczbę argumentów (określoną przez jego walencję). Patrzymy na rysunek.
5. *Inne pomysły*. Nie ma powodu, aby nie szukać innych jeszcze rozwiązań dla reprezentacji składniowej zdań. Można np. uznać, że elementem nadrzędnym w zdaniu jest jego temat. W przypadku każdego „pomysłuna-składnię” trzeba jednak pamiętać, że zachowana powinna zostać jakaś

forma zgodności składni z semantyką, że nowa propozycja powinna lepiej od starych koncepcji zdawać sprawę ze zjawisk składniowych w językach świata, że wreszcie powinna mieć, nazwijmy to tak, *potencjał logiczno-obliczeniowy*, czyli że z jej pomocą możliwe jest dokonywanie formalnych obliczeń na elementach poziomu składniowego oraz że nie stwarza ona jakiegoś zasadniczego konfliktu z ustaleniami logiki (w szczególności, prawami semantyki logicznej).

6. *Transformacje*. To operacje na strukturach składniowych, pozwalające na ich przekształcanie. Przykładem może być *pasywizacja*: przekształcanie zdania w stronie czynnej w zdanie w stronie biernej.

Twierdzi się, że w każdym języku ustalić można pewną liczbę (zwykle kilkadziesiąt) podstawowych schematów zdaniowych. W książce Grzegorzczkowska 2007 wymienia się kilka następujących przykładów dla języka polskiego (strona 113):

1. *Schematy blokujące mianownik, oparte na tzw. czasownikach niewłaściwych*. Świta. Warto się uczyć.
2. *Schematy konotujące jednego aktanta w mianowniku*. Prezes śpi. Prezydent czuwa.
3. *Schematy konotujące jednego aktanta w mianowniku i wyrażające predykat przymiotnikowo lub rzeczownikowo w połączeniu z czasownikiem posiłkowym być, stać się i in.* Maria jest niewinna. Marian staje się gangsterem.
4. *Schematy konotujące dwóch aktantów, oparte głównie na czasownikach oznaczających czynności skierowane na obiekt lub oznaczających doznania z wskazaniem na przyczynę*. Prezes lubi kota. Prezydent boi się kota.
5. *Schematy konotujące trzech aktantów, odnoszące się głównie do przekazu przedmiotów lub wartości (informacji)*. Zenon daje Krystynie gruszkę.
6. *Schematy konotujące czterech aktantów, odnoszące się głównie do sytuacji powodowania przemieszczania się obiektów*. Minister przenosi akta z willi generała do muzeum.
7. *Schematy konotujące uzupełnienia zdarzeniowe (wyrażane zdaniem lub bezokolicznikiem)*. Krystyna wie, że Wacław długo nie pociągnie.

Więcej przykładów (dla języka polskiego) znajdują słuchacze np. w pracach: Grochowski, Karolak, Topolińska 1984 lub Saloni, Świdziński 1998.

## 5 Reguły tworzenia zdań złożonych

W książce Grzegorzycykowa 2007 wymienia się trzy typy zdań złożonych języka polskiego (strona 116):

1. *Zdania łączone na zasadzie semantycznej.*
  - (a) *Parataktyczne (współrzędne).* Zakochał się i kompletnie zgłupiał.
  - (b) *Hipotaktyczne (podrzędne).* Zakochał się, ponieważ kompletnie zgłupiał.
2. *Określnik (zastępnik) grupy imiennej zdania nadrzędnego (zdania relatywne).* Kto myśli, ten jest rozumny.
3. *Konotowane przez predykat zdania nadrzędnego (zdania intensjonalne, dopełniające).* Widzę, że nadciąga katastrofa.

Dalej, autorka wymienia pięć typów zdań parataktycznych:

1. *Zestawienie zdarzeń (stanów rzeczy) na zasadzie przestrzennej i czasowej, lub też jakiegoś innego myślowego powiązania.* Sprzedał konia, a uzyskane pieniądze przepił.
2. *Przeciwstawianie: proste zauważanie różnic i przeciwstawianie stanów z ukrytym rozumowaniem.* Marian jest grzeczny, ale podstępny.
3. *Alternatywa, czyli informacja, że zachodzi jeden z dwóch pomyślanych stanów rzeczy.* Pacjent jest chory lub pijany.
4. *Wskaazywanie na skutek.* Agata zamawia pizzę z rabarbarem, bo lubi witaminy.
5. *Doprecyzowanie.* Prezydent jest prawnikiem, czyli nie łamie prawa.

Wedle autorki relacje semantyczne wyrażane hipotaktycznie sprowadzają się do trzech typów:

1. *Niektóre przypadki porównywania zdarzeń (stanów rzeczy).* Zachowuje się tak, jakby miał poparcie społeczne.
2. *Związki temporalne między zdarzeniami.* Wyszła za mąż, zanim ukończyła studia.

3. *Powiązania przyczynowo-skutkowe między zdarzeniami*. Ucieszyłby się, gdyby zapłacił wyższy podatek.

Wreszcie, ostatni z tych typów (czyli zdania przyczynowo skutkowe) obejmuje:

1. *Zdania przyczynowe zwykłe*. Umarła, ponieważ zjadła zapiekankę na dworcu w Kutnie.
2. *Zdania z niedostateczną przyczyną*. Chociaż zjadła zapiekankę na dworcu w Kutnie, to żyje do dziś.
3. *Zdania z przyczyną logiczną (hipotetyczną)*. Skoro nie podpisał nominacji, to widocznie miał powody.
4. *Zdania celowe*. Wyjechał, aby zacząć nowe życie.
5. *Zdania warunkowe*. Jeśli mnie nie zbałamucisz, pójdę do klasztoru.
6. *Zdania skutkowe*. Myślał tak intensywnie, że aż się spocił.

Na marginesie pozwolimy sobie dodać, że w językach świata spotyka się różne sposoby wyrażania połączeń zdań. W przypadku zdań złożonych współrzędnie dominujące są trzy typy (Majewicz 1989, 218):

1. sekwencja zdań prostych łączonych spójnikami (np. języki indoeuropejskie)
2. bezspójnikowa sekwencja zdań prostych (np. wiele języków izolujących)
3. sekwencja zdań prostych łączonych drogą afiksacji (np. japoński).

Natomiast zdania złożone podrzędnie mogą być wyrażane na następujące sposoby (Majewicz 1989, 219):

1. sekwencja zdań prostych połączonych zaimkami względnymi (np. polski)
2. atrybutyzacja zdań podrzędnych (np. japoński)
3. konstrukcja teleskopowa (np. chiński, tajski).



## 5.1 Przykład: zgony, ordery, siostry szarytki

Słuchacze zechcą zwrócić uwagę na różnice znaczeniowe występujące w podanych parach zdań złożonych:

1.	Umarł i dostał jakiś order.	Dostał jakiś order i umarł.
2.	Umarł, bo dostał jakiś order.	Dostał jakiś order, bo umarł.
3.	Umarł więc dostał jakiś order.	Dostał jakiś order więc umarł.
4.	Umarł, chociaż dostał jakiś order.	Dostał jakiś order chociaż umarł.
5.	Umarł, gdy dostał jakiś order.	Dostał jakiś order gdy umarł.
6.	Umarł, mimo że dostał jakiś order.	Dostał jakiś order, mimo że umarł.
7.	Nie dość, że umarł, to dostał jakiś order.	Nie dość, że dostał jakiś order, to umarł.
8.	Umarł, ale dostał jakiś order.	Dostał jakiś order, ale umarł.
9.	Umarł, zanim dostał jakiś order.	Dostał jakiś order zanim umarł.
10.	Umarł dzięki temu, że dostał jakiś order.	Dostał jakiś order dzięki temu, że umarł.

Słuchacze zechcą się również zastanowić, jaka jest *struktura logiczna* podanych par wypowiedzi. Jakim zdaniom języka rachunku predykatów pierwszego rzędu mogą odpowiadać powyższe pary zdań?

W książce Grochowski, Karolak, Topolińska 1984 podaje się charakterystyki około pięćdziesięciu relacji semantycznych, wyrażanych przez zdania złożone.

Przywołajmy też przykład podany przez Marka Tokarza (i trochę przez nas zmodyfikowany):

1. *Zmarła.*
2. *Zmarła po urodzeniu.*
3. *Zmarła zaraz po urodzeniu.*
4. *Zmarła zaraz po urodzeniu dziecka.*
5. *Zmarła zaraz po urodzeniu dziecka przez siostrę.*
6. *Zmarła zaraz po urodzeniu dziecka przez siostrę szarytkę.*
7. *Zmarła zaraz po urodzeniu dziecka przez siostrę szarytkę wydaloną ze zgromadzenia zakonnego.*

Słuchacze zechcą zastanowić się, co wiemy o zmarłej z każdego z podanych (coraz bardziej złożonych) zdań.

## 6 Typologie syntaktyczne

Języki świata możemy klasyfikować ze względu na występujące w nich kategorie oraz relacje i konstrukcje syntaktyczne. W podanych niżej przykładach stosujemy następujące, powszechne w lingwistyce oznaczenia:

1. Dt – *determinatum*, człon określany
2. Ds – *determinans*, człon określający
3. S – *subject*, podmiot
4. V – *verb*, orzeczenie
5. O – *object*, dopełnienie
6. A – *agens*, wykonawca czynności
7. P – *patiens*, odbiorca czynności

Niżej podamy przykłady tylko trzech typologii syntaktycznych:

1. *Typologia szyku w połączeniach elementu określającego i określanego.*
2. *Typologia morfosyntaktyczna Milewskiego.*
3. *Typologia szyku elementów S, V oraz O.*

### 6.1 Typologia szyku określające-określane

Możliwe są dwie sytuacje:

1. Dt poprzedza Ds (np.: wietnamski, tajski, suahili)
2. Ds poprzedza Dt (np.: japoński, chiński, angielski).

Może też być tak, że niektóre człony określające stoją przed, a inne za członami określanymi. W polskim np. przydawka przymiotnikowa (Ds) z reguły poprzedza Dt, natomiast przydawka dopełniaczowa (Ds) z reguły następuje po Dt, por.: *wielki prezes, ulubienica prezesa*. We francuskim krótkie przymiotniki (Ds) poprzedzają określaną rzeczownik, zaś przymiotniki długie (Ds) następują po określanym rzeczowniku.

## 6.2 Typologia morfosyntaktyczna

Według Milewskiego, języki świata podzielić można na sześć typów, w zależności od tego, jakie środki morfosyntaktyczne wykorzystywane są dla wyrażania następujących zależności składniowych:

1. *Podmiotu do orzeczenia: S – V.*
2. *Agensa do orzeczenia: A – V.*
3. *Patiensa do orzeczenia: P – V.*
4. *Członu określającego do określanego: Ds – Dt.*

Oczywiście każdy język używa wybranych środków morfosyntaktycznych dla wyrażenia tych zależności. Milewski twierdzi jednak, że żaden język nie wykorzystuje czterech różnych wykładników tych relacji, a jedynie dwa lub trzy ich rodzaje, a wszystkie języki świata można pogrupować w sześć typów. Oznaczmy te rodzaje wykładników przez a, b oraz c (mogą to zatem być, np.: szyk, afiksacja, lub inne jeszcze sposoby wyrażania znaczeń gramatycznych). Klasyfikację Milewskiego podaje tabela (Milewski 1967, Majewicz 1989):

Relacja/Typ	I	II	III	IV	V	VI
S – V	a	a	a	a	a	a
A – V	a	b	a	b	a	b
P – V	b	a	b	a	b	a
Ds – Dt	c	c	b	b	a	a

Zacytujmy komentarz do tej tabeli z książki Majewicz 1989 (strona 216):

Typy I i II obejmują języki dysponujące trzema różnymi wykładnikami formalnymi dla wyrażenia tych stosunków, przy czym w typie pierwszym, do którego należą – według Milewskiego – np. języki *indoeuropejskie*, *semito-chamickie*, *uralskie*, *ałtajskie*, *chińskotybetańskie*, agens jest formalnie równy podmiotowi, a w typie drugim (obejmującym, według Milewskiego, „większość języków *kaukaskich*”, *baskijski*, języki *paleoazjatyckie*, *sumeryjski*), nierozróżnialne formalnie są podmiot i pacjens; w językach obu tych typów występuje formalne rozróżnienie między zdaniem (z orzeczeniem) a grupą nominalną (Ds–Dt – z formalnym wykładnikiem genetywu). W językach typu trzeciego następuje identyczność agensa i podmiotu oraz

pacjensa z członem określającym, a w językach typu czwartego – pacjensa i podmiotu oraz agensa i członu określającego. W językach typu piątego formalnie identyczne winny być podmiot, agens i człon określający, natomiast w językach typu szóstego – podmiot, agens i człon określany. Języki typów III–VI mają zatem tylko po dwa wykładniki formalne wyrażające owe cztery wyżej wyliczone stosunki. Milewski wskazywał na szczególną rzadkość języków zaklasyfikowanych zwłaszcza do typów V i VI, a jako przykłady języków poszczególnych typów podał: *hopi*, języki *indonezyjskie* (typ III), języki *eskimoskie*, *saliskie*, *majańskie*, niektóre *kaukaskie* (typ IV), *niska* – dialekt tsimshiański znad rzeki Nass (typ V), oraz *tunica*, *guarani*, *tsimshian* (typ VI).

Słuchaczy zainteresowanych bardziej współczesnymi poglądami i ustaleniami dotyczącymi typologii morfosyntaktycznej zachęcamy do lektury nowszych opracowań. Można też zajrzeć np. na stronę:

[https://en.wikipedia.org/wiki/Morphosyntactic\\_alignment](https://en.wikipedia.org/wiki/Morphosyntactic_alignment)

### 6.3 Typologia szyku

Zachęcamy słuchaczy do obejrzenia krótkiej prezentacji Pana dra Konrada Juszczyka *Typologia języków według szyku wyrazów*:

<http://www.slideshare.net/Linguist/typologia-jzykw-wg-szyku-wyrazw>

W językach świata znajdujemy wszystkie sześć możliwych kombinacji elementów S, V oraz O (przy czym niektóre kombinacje występują znacząco częściej niż inne):

1. *Typ SOV*: japoński, koreański, tamil, tybetański, ...
2. *Typ SVO*: angielski, polski, chiński, niemiecki, igbo, jawański, ...
3. *Typ VSO*: semickie (arabski, hebrajski), celtyckie (szkocki, irlandzki, walijski), polinezyjskie (samoński, hawański), ...
4. *Typ VOS*: malgaski, fidżyjski, ...
5. *Typ OVS*: hixkaryana, apalai, arecuna, ...
6. *Typ OSV*: nadeb, hupda shavante (Amazonia).

Trzeba oczywiście pamiętać, że chodzi tu o szyk odbierany jako „najbardziej naturalny”. W wielu językach szyk wymienionych elementów pozostaje dość swobodny. Szyk ten jednak ma istotne znaczenie, jeśli to właśnie on jest wykładnikiem znaczeń gramatycznych.

Na koniec zauważmy, że w językach o szyku S – V oraz Dt – Ds zaciera się lub wręcz znika różnica między prostym zdaniem a frazą nominalną (np. w tajskim). W takim przypadku nie ma więc możliwości odróżnienia *Pies jest mały* od *Mały pies*.

## 7 Wesoła dygresja: kilka podstawowych pojęć lingwistyki matematycznej

### 7.1 Uwagi wstępne

We współczesnych teoriach składni logicznej zakłada się, że języki etniczne mogą zostać opisane w terminach (różnego rodzaju) *gramatyk kategorialnych*. Twórcą pierwszego takiego formalnego opisu był Kazimierz Ajdukiewicz (1935). Obecnie dysponujemy wieloma rodzajami gramatyk kategorialnych, znacznie subtelniejszych od owego pierwowzoru. Pół wieku temu rozpoczęto także intensywne badania nad *gramatykami formalnymi* w ogólności. Opisuje się *języki formalne*, różne typy gramatyk *generujących* takie języki oraz różne rodzaje *automatów* akceptujących (rozpoznających) języki formalne. Rozważania te dały początek *lingwistyce matematycznej*. To osobna dyscyplina matematyczna, mająca – z oczywistych powodów – liczne zastosowania w badaniach języków etnicznych. Ponadto, lingwistyka matematyczna powiązana jest również z *teorią obliczeń* oraz z *logiką matematyczną*.

W lingwistyce matematycznej rozważa się *języki* jako zbiory *słów* złożonych z elementów pewnego *alfabetu*. W szczególności, wszystkie języki logiki matematycznej są językami w tym rozumieniu. Natomiast tak pojmowane języki są z kolei pewnego rodzaju strukturami algebraicznymi: każdy język jest pewną algebrą *słów*, dla których określona jest funkcja *konkatenacji*.

Wyrażenia złożone języków formalnych nie są jedynie ciągami symboli: mają one także (jednoznaczną) strukturę *syntaktyczną*, odzwierciedlającą ich budowę (m.in. także proces ich generowania). Tak więc, na przykład termy i formuły języka logiki pierwszego rzędu są *drzewami*. Strukturę drzewa mają również np. *obliczenia* dokonywane w odpowiednim języku formalnym.

W tym punkcie przypomnimy niektóre wybrane podstawowe pojęcia dotyczące języków formalnych oraz konstrukcji, które pozwalają je *generować* i *rozpoznawać*, czyli – odpowiednio – *gramatyk formalnych* oraz *automatów*. Nie przywołujemy natomiast twierdzeń teorii gramatyk i automatów. Zainteresowany czytelnik zechce skorzystać z literatury przedmiotu.

## 7.2 Kategorialny opis języka

Ograniczymy się do następujących uwag, dotyczących samej idei *kategorialnego* opisu języka (sztucznego bądź naturalnego). W opisie takim ustala się pewien zbiór *kategorii*, które przypisywane są wyrażeniom badanego języka. Kategorie te mogą być *proste* (np.: nazwa, zdanie), bądź *złożone*. Kategorie złożone przypisuje się tym wyrażeniom, które – połączone z wyrażeniami określonych kategorii – dają w wyniku wyrażenia również ściśle określonej kategorii. Dla przykładu, wyrażenie, które w połączeniu z dwoma zdaniami daje w wyniku zdanie, otrzyma kategorię *funktora*, o (dwóch) argumentach zdaniowych i wartości, będącej zdaniem. Argumentami funktorów mogą być inne funktry – otrzymujemy w ten sposób całą hierarchię coraz to bardziej złożonych kategorii. W opisie kategorialnym przyjmuje się także odpowiednie *reguły skracania* – reguły ustalające, w jaki sposób ciąg występujących po sobie kategorii skrócić można do jednej kategorii. W podanym wyżej przykładzie, jeśli mamy ciąg kategorii: zdanie, funktor zdaniowy (dwuargumentowy) o argumentach zdaniowych, zdanie, to ciąg ten skrócimy do ciągu jednoelementowego: zdanie. Reguły skracania pozwalają ustalać, które ciągi kategorii dają (w wyniku konsekwentnego wykonania wszystkich skrótów) interesującą nas kategorię – np. kategorię zdania. W ten sposób zagadnienie *poprawności syntaktycznej* zostaje zredukowane do przypisania wyrażeniom kategorii oraz reguł skracania. W przypadku sztucznych języków logiki zagadnienie to zostało w pełni rozwiązane. W przypadku języków etnicznych sprawa jest o wiele bardziej skomplikowana: zarówno zestawy kategorii, jak i postać reguł skracania są tu o wiele bardziej złożone.

Kategorialny opis języka ma wielką zaletę – pozwala mianowicie na konsekwentne budowanie semantyki dla opisywanych języków, a przy tym semantyki w ściśle określonym sensie *zgodnej* ze składnią. Bez szczegółów natury technicznej powiedzmy jedynie, że denotacje dla wyrażeń prostych kategorii są jakoś z góry zadane, natomiast denotacje kategorii złożonych (czyli funktorów) są zawsze odpowiednimi *funkcjami*, o argumentach wyznaczonych przez kategorię argumentów rozważanego funkтора i o wartościach wyznaczonych przez kategorię wartości tego funkтора. Cała semantyka przyjmuje postać (teorio-mnogościowej) hierarchii, w której występują wyłącznie przedmioty określonych rodzajów oraz różnego typu funkcje.

## 7.3 Języki formalne

Języki formalne to zbiory ciągów symboli z ustalonego *alfabetu* (czasem używa się też terminu: *słownik*). Każdy taki ciąg wyposażony jest też w pewną strukturę *syntaktyczną*, gdy języki określane są poprzez *gramatyki* generujące ciągi lub przez

automaty rozpoznające ciągi.

Niech  $V$  będzie niepustym zbiorem skończonym, zwanym *alfabetem*. Skończone ciągi elementów zbioru  $V$  nazywamy *słowami* nad alfabetem  $V$ .

Zbiór wszystkich słów nad alfabetem  $V$  oznaczamy przez  $V^*$ .

Słowem nad każdym alfabetem jest słowo *puste*, nie zawierające żadnego symbolu i oznaczane przez  $\varepsilon$ .

Zbiór  $V^* - \{\varepsilon\}$  będziemy oznaczali  $V^+$ . Jest to zatem zbiór wszystkich niepustych słów nad alfabetem  $V$ .

Jeśli  $V = \{a_1, \dots, a_n\}$  oraz  $x = a_{i_1} \dots a_{i_k}$  i  $y = a_{j_1} \dots a_{j_m}$  są słowami nad  $V$ , to ich *złożeniem (konkatenacją)* nazywamy słowo  $xy = a_{i_1} \dots a_{i_k} a_{j_1} \dots a_{j_m}$ .

Operacja złożenia jest łączna, jej elementem neutralnym jest słowo puste. Tak więc, zbiór  $V^*$  wraz z operacją złożenia tworzy półgrupę (jest to półgrupa wolna, generowana przez  $V$ ).

*Długości* słowa nazywamy liczbę występujących w nim symboli. Długość słowa  $x$  jest oznaczana przez  $|x|$  i definiowana indukcyjnie (dla  $x \in V^*$  oraz  $a \in V$ ):

- $|\varepsilon| = 0$
- $|ax| = |x| + 1$ .

Przez indukcję określamy też operację *odbicia zwierciadlanego* (dla  $x \in V^*$  oraz  $a \in V$ ):

- $\varepsilon^R = \varepsilon$
- $(ax)^R = x^R a$ .

Operacje  $-_l$  *lewostronnej różnicy* oraz  $-_p$  *prawostronnej różnicy* słów określamy warunkami:

- $x -_l y = z$ , jeśli  $x = yz$
- $x -_l y = \varepsilon$ , jeśli nie istnieje  $z$  taki, że  $x = yz$
- $x -_p y = z$ , jeśli  $x = zy$
- $x -_p y = \varepsilon$ , jeśli nie istnieje  $z$  taki, że  $x = zy$ .

Mówimy, że  $x$  jest *podstawem*  $y$ , gdy istnieją  $z, w \in V^*$  takie, że  $y = zxw$ . Przy tym:

- Jeśli  $zw \neq \varepsilon$ , to  $x$  jest *podstawem właściwym*  $y$ .

- Jeśli  $z = \varepsilon$ , to  $x$  jest *podstawem początkowym*  $y$ .
- Jeśli  $w = \varepsilon$ , to  $x$  jest *podstawem końcowym*  $y$ .

Jeśli  $x$  jest podstawem  $y$ , to piszemy  $x \prec y$ .

Każdy podzbiór  $L$  zbioru  $V^*$  nazywamy *językiem* nad alfabetem  $V$ . Zauważmy, że istnieje continuum języków nad dowolnym niepustym skończonym alfabetem.

Zdefiniujemy kilka operacji na językach:

- Operacje teoriomnogościowe: sumy, iloczynu, różnicy, różnicy symetrycznej, iloczynu kartezjańskiego języków są definiowane w sposób oczywisty.
- *Odbiciem zwierciadlanym* języka  $L$  nazywamy język:  $L^R = \{x^R : x \in L\}$ .
- *Złożeniem* języków  $L_1$  i  $L_2$  nazywamy język:  $L_1L_2 = \{xy : x \in L_1 \wedge y \in L_2\}$ .
- *Potęę złożeniową* języka definiujemy indukcyjnie:  $L^0 = \{\varepsilon\}$ ,  $L^{n+1} = LL^n$ . Z kontekstu będzie zawsze jasno wynikać, czy mówimy o potędze złożeniowej, czy kartezjańskiej języka.
- *Domknięciem* języka  $L$  nazywamy język:  $L^* = \bigcup_n L^n$ .

Poszczególne klasy języków mogą być zamknięte na pewne operacje: sumy, iloczynu, itp. Dla przykładu, mówimy, że klasa  $\mathbb{K}$  języków jest *zamknięta ze względu na uzupełnianie*, gdy dla każdego alfabetu  $V$  i każdego języka  $L \subseteq V^*$ : jeśli  $L \in \mathbb{K}$ , to  $(V^* - L) \in \mathbb{K}$ . Podobnie dla innych operacji na językach.

Niech  $V$  będzie dowolnym alfabetem i założmy, że każdemu symbolowi  $a \in V$  przyporządkowano nowy, dowolny alfabet  $\Sigma_a$  oraz że podana została funkcja  $f : V \rightarrow \wp(\Sigma^*)$ , gdzie  $\Sigma = \bigcup_{a \in V} \Sigma_a$ . Funkcja  $f$  przyporządkowuje przy tym każdemu symbolowi  $a \in V$  pewien niepusty język  $f(a) \subseteq (\Sigma_a)^*$ . Rozszerzamy ją na cały zbiór  $V^*$  poprzez warunki:

- $f(\varepsilon) = \varepsilon$
- $f(ax) = f(a)f(x)$ , dla  $a \in V$  oraz  $x \in V^*$ .

Wtedy oczywiście  $f(a_1a_2 \dots a_n) = f(a_1)f(a_2) \dots f(a_n)$ . Definiujemy teraz funkcję  $f^* : \wp(V^*) \rightarrow \wp(\Sigma^*)$  przez warunek (dla każdego  $L \subseteq V^*$ ):

$$f^*(L) = \bigcup_{x \in L} f(x).$$



Język  $f^*(L)$  nazywamy wynikiem zastosowania *podstawienia*  $f$  do języka  $L$ . Szczególnym przypadkiem podstawienia jest homomorfizm: podstawienie  $f$  określone na  $V$  nazywamy *homomorfizmem* języków, gdy dla każdego  $a \in V$  język  $f(a)$  jest językiem jednoelementowym. Jeśli język jednoelementowy identyfikować z jego jedynym elementem, to homomorfizmem jest funkcja  $f$ , która każdemu elementowi  $a \in V$  przyporządkowuje słowo  $f(a) \in \Sigma_a^*$ . Wtedy  $f(a_1 a_2 \dots a_n) = f(a_1) f(a_2) \dots f(a_n)$ , gdzie składanie po prawej stronie równości rozumiane jest jako złożenie słów (a nie języków). Przy takim rozumieniu, homomorficzny obraz języka jest językiem. Różnowartościowe homomorfizmy określone na  $V^*$  nazywamy *izomorfizmami* języków.

## 7.4 Gramatyki formalne

*Gramatyką* nazywamy każdy układ  $G = (K, V, P, S)$  taki, że:

- $K$  jest skończonym alfabetem. Jego elementy to *symbole nieterminalne*.
- $V$  jest skończonym alfabetem. Jego elementy to *symbole terminalne*.
- Zbiory  $V$  oraz  $K$  są rozłączne.
- $P$  jest skończonym zbiorem par słów nad alfabetem  $K \cup V$  takich, że pierwszy element pary nie jest słowem pustym. A zatem  $P \subseteq (K \cup V)^+ \times (K \cup V)^*$ . Elementy zbioru  $P$  nazywamy *produkcjami* (albo *regułami przepisywania*).
- $S \in K$  jest wyróżnionym elementem zbioru  $K$ . Nazywamy go *symbolem początkowym*.

Jeśli  $(x, y) \in P$ , to będziemy pisali  $x \rightarrow y$  (z ewentualnym indeksem przy  $\rightarrow$ , odnoszącym się do rozważanej gramatyki).

Mówimy, że słowo  $y \in (K \cup V)^*$  *wyprowadzamy w jednym kroku* ze słowa  $x \in (K \cup V)^*$  w gramatyce  $G$ , jeśli istnieją słowa  $u_1, u_2, v_1, v_2 \in (K \cup V)^*$  takie, że:

- $x = u_1 v_1 u_2$
- $y = u_1 v_2 u_2$
- $(v_1, v_2) \in P$  (czyli  $v_1 \rightarrow_G v_2$ ).

Jeśli  $x$  wyprowadzamy w jednym kroku z  $y$ , to piszemy  $x \Rightarrow_G y$ . Niech  $\Rightarrow_G^*$  oznacza przechodnie domknięcie relacji  $\Rightarrow_G$ . Jeśli  $x \Rightarrow_G^* y$ , to mówimy, że  $x$  można wyprowadzić z  $y$  w gramatyce  $G$ .

Mówimy, że słowo  $x$  jest generowane przez gramatykę  $G$ , jeżeli  $S \Rightarrow_G^* x$ . Językiem generowanym przez gramatykę  $G$  nazywamy zbiór słów:

$$L(G) = \{x : S \Rightarrow_G^* x\}.$$

Mówimy, że gramatyka  $G$  jest:

- *typu 1*, albo *kontekstowa*, gdy każda jej reguła ma postać  $\alpha A \beta \rightarrow \alpha \psi \beta$ , gdzie  $A \in K$  oraz  $\psi \neq \varepsilon$ ;
- *typu 2*, albo *bezkontekstowa*, gdy każda jej reguła ma postać  $A \rightarrow \psi$ , gdzie  $A \in K$ ;
- *typu 3*, albo *prawostronnie liniowa*, gdy każda jej reguła ma postać  $A \rightarrow xB$  lub  $A \rightarrow x$ , gdzie  $A \in K$ .

Czasem o gramatykach, które nie mają żadnych ograniczeń na postać reguł przepisywania mówimy, że są *typu 0*.

Mówimy, że język  $L \subseteq V^*$  jest:

- *typu 0* (albo *rekurencyjnie przeliczalny*), gdy  $L$  jest generowany przez jakąś gramatykę  $G = (K, V, P, S)$ ;
- *typu 1* (albo *kontekstowy*), gdy  $L$  jest generowany przez jakąś gramatykę kontekstową  $G = (K, V, P, S)$ ;
- *typu 2* (albo *bezkontekstowy*), gdy  $L$  jest generowany przez jakąś gramatykę bezkontekstową  $G = (K, V, P, S)$ ;
- *typu 3* (albo *regularny*), gdy  $L$  jest generowany przez jakąś gramatykę prawostronnie liniową  $G = (K, V, P, S)$ .

Jeśli oznaczymy przez  $\mathcal{L}_i$  klasę wszystkich języków typu  $i$ , gdzie  $0 \leq i \leq 3$ , to zachodzą następujące inkluzje właściwe:

$$\mathcal{L}_3 \subset \mathcal{L}_2 \subset \mathcal{L}_1 \subset \mathcal{L}_0.$$

## 7.5 Automaty

Automatem skończonym nazywamy każdy układ  $M = (Q, \Sigma, \delta, q_0, F)$  taki, że:

- $Q$  jest skończonym zbiorem stanów
- $\Sigma$  jest alfabetem, nazywanym *alfabetem wejściowym*
- $\delta : Q \times \Sigma \rightarrow Q$  jest funkcją przejścia
- $q_0 \in Q$  jest wyróżnionym stanem początkowym
- $F \subseteq Q$  jest zbiorem stanów końcowych (albo: *akceptujących*).

Dla dowolnego automatu skończonego  $M = (Q, \Sigma, \delta, q_0, F)$  możemy jego funkcję przejścia  $\delta$  rozszerzyć do funkcji  $\bar{\delta} : Q \times \Sigma^* \rightarrow Q$  tak, aby zachodziły warunki (dla dowolnych  $q \in Q$ ,  $s \in \Sigma^*$  oraz  $a \in \Sigma$ ):

- $\bar{\delta}(q, \varepsilon) = q$
- $\bar{\delta}(q, sa) = \delta(\bar{\delta}(q, s), a)$ .

Mówimy, że słowo  $s \in \Sigma^*$  jest *akceptowane* przez automat skończony  $M$ , gdy  $\bar{\delta}(q_0, s) \in F$ . Zbiór  $L(M)$  wszystkich słów akceptowanych przez  $M$  nazywamy *językiem akceptowanym przez  $M$* . Mamy zatem:

$$L(M) = \{s \in \Sigma^* : \bar{\delta}(q_0, s) \in F\}.$$

Języki akceptowane przez automaty skończone nazywamy językami *regularnymi*. Nie czynimy tu zamieszania w terminologii, gdyż można dowieść, że języki regularne w tym sensie, to to samo, co języki regularne w sensie omówionym dla gramatyk.

Klasa wszystkich języków regularnych nad alfabetem  $\Sigma$  to najmniejsza klasa  $\mathbb{X}$  taka, że:

- $\emptyset \in \mathbb{X}$
- $\{\varepsilon\} \in \mathbb{X}$
- jeśli  $a \in \Sigma$ , to  $\{a\} \in \mathbb{X}$
- jeśli  $L_1, L_2 \in \mathbb{X}$ , to  $L_1 \cup L_2 \in \mathbb{X}$
- jeśli  $L_1, L_2 \in \mathbb{X}$ , to  $L_1 L_2 \in \mathbb{X}$

- jeśli  $L \in \mathbb{X}$ , to  $L^* \in \mathbb{X}$ .

Automatem skończonym ze stosem nazywamy każdy układ

$$M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$$

taki, że:

- $Q$  jest skończonym zbiorem stanów
- $\Sigma$  jest alfabetem, nazywanym *alfabetem wejściowym*
- $\Gamma$  jest skończonym zbiorem, nazywanym *alfabetem stosowym*
- $\delta : Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma \rightarrow \wp(Q \times \Gamma^*)$  jest *funkcją przejścia*
- $q_0 \in Q$  jest wyróżnionym *stanem początkowym*
- $Z_0 \in \Gamma$  jest *symbolem początkowym*
- $F \subseteq Q$  jest zbiorem stanów *końcowych* (albo: *akceptujących*).

Konfiguracją automatu  $M$  nazywamy trójkę  $(q, s, \gamma)$  taką, że  $q \in Q$ ,  $s \in \Sigma^*$  oraz  $\gamma \in \Gamma^*$ .

Relacją przejścia automatu  $M$  nazywamy dwuargumentową relację  $\Rightarrow$  między konfiguracjami automatu  $M$ , zdefiniowaną następująco:

$$\Rightarrow = \{((q, as, Z\gamma), (q', s, \gamma'\gamma)) : (q', \gamma') \in \delta(q, a, \gamma)\} \cup \{((q, s, Z\gamma), (q', s, \gamma'\gamma)) : (q', \gamma') \in \delta(q, \varepsilon, \gamma)\}.$$

Także przechodnie zwrotne domknięcie powyższej relacji nazywa się relacją przejścia automatu  $M$ .

Automaty ze stosem akceptują języki bezkontekstowe. Także na odwrót, każdy język bezkontekstowy jest akceptowany przez pewien automat ze stosem.

Klasa wszystkich języków bezkontekstowych:

- jest zamknięta ze względu na operacje sumy, złożenia i domknięcia;
- nie jest zamknięta ani ze względu na operacje iloczynu, ani ze względu na operację uzupełnienia.

Z kolei, klasa wszystkich języków kontekstowych jest zamknięta na następujące operacje:

- sumy

- przekroju
- złożenia
- domknięcia
- odbicia zwierciadlanego
- podstawiania języków kontekstowych nie zawierających słowa pustego.

O językach typu 0 (rekurencyjnie przeliczalnych) mówi się więcej w rozważaniach dotyczących matematycznych modeli obliczalności.

## 8 Uwaga redakcyjna

Niniejsza notatka stanowi jedynie hasłowy przewodnik po treściach omówionych na wykładzie. Treści poruszane na dzisiejszym wykładzie omawiane są w książce Grzegorzczukowa 2007, w rozdziale 8. Bardziej szczegółowo składnia języka polskiego omówiona została w Grochowski, Karolak, Topolińska 1984 oraz Saloni, Świdziński 1998. Kilka uwag o typologiach syntaktycznych znaleźć można w Majewicz 1989. O lingwistyce matematycznej poczytać można w Blikle 1971 lub w wielu nowszych opracowaniach.

## Odnośniki bibliograficzne

- Blikle, A. 1971. *Automaty i gramatyki. Wstęp do lingwistyki matematycznej*. Państwowe Wydawnictwo Naukowe, Warszawa.
- Grochowski, M., Karolak, S., Topolińska, Z. 1984. *Składnia*. Państwowe Wydawnictwo Naukowe, Warszawa
- Grzegorzczukowa, R. 2007. *Wstęp do językoznawstwa*. Wydawnictwo Naukowe PWN, Warszawa.
- Majewicz, A.F. 1989. *Języki świata i ich klasyfikowanie*. Państwowe Wydawnictwo Naukowe, Warszawa.
- Milewski, T. 1967. *Językoznawstwo*. Państwowe Wydawnictwo Naukowe, Warszawa.
- Saloni, Z., Świdziński, M. 1998. *Składnia współczesnego języka polskiego*. Warszawa.