

SEMIOTYCZNE ASPEKTY GENETYKI MOLEKULARNEJ¹

JERZY POGONOWSKI
Zakład Logiki Stosowanej UAM
www.logic.amu.edu.pl

*Jako przestanie, kod jest listem, pisanym przez Nikogo
i wystanym do Nikogo; dopiero teraz, utworzywszy sobie informatykę,
zaczynacie pojmować, że coś takiego, jak listy opatrzone sensem,
których nikt nie układał rozmyślnie, aczkolwiek powstały i istnieją,
jak również uporządkowane odbieranie treści owych listów
jest możliwe pod nieobecność jakichkolwiek Istot i Rozumów.*

(Wykład inauguracyjny Golema; St. Lem „Golem XIV”)

Uwagi wstępne

Niniejszy artykuł poświęcony jest dyskusji semiotycznych aspektów niektórych procesów genetycznych. Po rozpatrzeniu argumentów na rzecz tezy, iż język genetyczny jest specyficznym systemem semiotycznym, przedstawia się pewne propozycje dotyczące możliwości zastosowania lingwistycznej aparatury pojęciowej w badaniach procesów genetycznych.

Język genetyczny

Przez język genetyczny rozumiemy określony system jednostek genetycznych powiązanych wzajemnymi relacjami. Jednostki genetyczne to przede wszystkim

¹Opublikowane w: *Kosmos* 3, 1983, 425–431. Pierwotna wersja tego artykułu wygłoszona została na I Ogólnopolskiej Sesji z Biologii Teoretycznej (Uniwersytet im. Adama Mickiewicza w Poznaniu, październik 1981). Obecny tekst opracowany został dzięki finansowej pomocy Alexander von Humboldt Stiftung. Dopisek z 2008 roku: ten tekst był pisany ponad ćwierć wieku temu, co uprasza się brać pod uwagę przy jego ewentualnej ocenie.

określone segmenty nieregularnych polimerów — DNA, mRNA oraz białek. Jednostkami planu wyrażania języka genetycznego są np.:

- *nukleotydy*: zasady, których sekwencje tworzą cząsteczki DNA oraz mRNA (adenina, guanina, tymina, cytozyna i uracyl),
- *kodony*: 64 trójki zasad, z których każda koduje określony aminokwas,
- *cistrony*: sekwencje zasad kodujące pojedyncze łańcuchy polipeptydowe,
- *transkryptyony*: sekwencje zasad kodujące cząsteczki mRNA,
- *replikony*: sekwencje cistronów uczestniczące jako całość w procesie replikacji,
- *segregony*: ciągi zasad uczestniczące jako całość w procesie segregacji,
- *genotypy*: całe zbiory replikonów komórki,
- *znaki interpunkcji*: ciągi zasad kodujące rozpoczęcie i zakończenie transkrypcji, translacji, replikacji, segregacji, etc.

Jednostkami planu treści języka genetycznego (tj. znaczeniami i składnikami znaczeń) są dla przykładu:

- *aminokwasy*: cząsteczki, których sekwencje tworzą łańcuchy polipeptydowe,
- *białka i ich funkcjonalnie relewantne segmenty*: np. pojedyncze łańcuchy polipeptydowe, ośrodki aktywne, etc.
- *funkcje biochemiczne*: funkcje pełnione przez jednostki planu wyrażania języka genetycznego (np. funkcja matrycowa) oraz funkcje spełniane przez kodowane przez te jednostki segmenty polipeptydów (np. funkcja katalityczna, metaboliczna, etc.).

Relacje genetyczne rozumiemy ekstensjonalnie, tzn. jako zbioru układów (par, trójek, etc.) jednostek genetycznych powiązanych określonymi zależnościami. Należą do relacji genetycznych zatem:

- *proces translacji (kod genetyczny)*: zależność między kodonami a wyznaczonymi przez nie aminokwasami,

- *proces transkrypcji*: zależność łącząca transkryptyony z odpowiadającymi im cząsteczkami mRNA,
- *relacja komplementarności*: zależność łącząca komplementarne zasady,
- *proces replikacji*: zależność łącząca komplementarne replikony,
- *synonimia kodonów*: zależność łącząca kodony kodujące ten sam aminokwas,
- *pogrupowania jednostek genetycznych w klasy (relacje jednoargumentowe)* ze względu na wspólne własności tych jednostek (np. klasy cistronów kodujących określone rodzaje polipeptydów) etc.

Język genetyczny jako system semiotyczny

Ustalenie, że pewien zespół jednostek i relacji tworzy system semiotyczny polega na wykazaniu, że zespół ten posiada cechy definiujące systemy znakowe. W szczególności, aby dany system znakowy był językiem, posiadać powinien cechy wspólne wszystkim językom — tzw. uniwersalia językowe. Do podstawowych uniwersaliów językowych należą takie cechy strukturalne jak: semantyczność, arbitralność, dwuklasowość, produktywność (otwartość), dyskretność.² Argumentację za tym, iż język genetyczny wszystkie te cechy posiada, podaliśmy w innym miejscu.³ Ograniczymy się tu zatem do nadania powyższym cechom semiotycznym interpretacji genetycznej.

Semantyczność systemu semiotycznego polega na tym, że wyróżnić można w nim dwa plany: wyrażania i treści. Jednostki planu wyrażania (wyrażenia językowe) powiązane są relacjami semantycznymi (relacja oznaczania) z odpowiadającymi im jednostkami planu treści (znaczeniami). W języku genetycznym za plan wyrażania uważać można zbiór wszystkich (relewantnych genetycznie) segmentów polinukleotydowych, a za plan treści — zbiór wszystkich segmentów polipeptydowych. Związki semantyczne między jednostkami tych planów wyznacza proces translacji (kod genetyczny).

²Por. Charles F. Hockett, *Zagadnienie uniwersaliów w języku*, w tomie: *Językoznawstwo strukturalne*, PWN, Warszawa 1979, s. 209–229.

³Jerzy Pogonowski, *Semiotyczna charakterystyka języka genetycznego*. Poznańskie Studia z Filozofii Nauki, nr 7, PWN, Warszawa – Poznań 1982, s. 53–62.

Arbitralność systemu semiotycznego polega na tym, że jego znaki nie są ikonicznymi odpowiednikami nazywanych przedmiotów. Arbitralne w tym sensie są wszystkie języki ludzkie — nie ma żadnego koniecznego związku między formą dźwiękową wyrazu a nazywanym przezeń przedmiotem. Cecha arbitralności występuje również w języku genetycznym: proces translacji bazuje na określonym powinowactwie chemicznym (poszczególne aminokwasy nie są w żadnym sensie ikonicznymi odpowiednikami kodonów).

Dyskretność przysługuje tym systemom znakowym, których komunikaty nie dają się w nieskończoność dzielić na coraz mniejsze jednostki. Zatem w dyskretnych systemach semiotycznych każdy komunikat jest kombinacją określonej ilości jednostek elementarnych. Język genetyczny jest oczywiście dyskretny — wszystkie komunikaty genetyczne można traktować jako kombinacje czterech nukleotydów (adenina, guanina, tymina i cytozyna w DNA oraz adenina, guanina, cytozyna i uracyl w mRNA).

O **dwuklasowości** systemu znakowego mówimy wtedy, gdy dają się w nim wyróżnić dwa podsystemy jednostek: system cenemów i system pleremów. Cenemę są jednostkami nie niosącymi samodzielnie żadnego znaczenia, stanowią one jedynie materiał budulcowy jednostek znaczących, tzn. pleremów. W językach ludzkich system cenemów odpowiada systemowi dźwiękowemu. Pojedyncze dźwięki nie niosą żadnego znaczenia, dopiero ich kombinacje (morfy, słowa, wyrazy, frazy, zdania, etc.) tworzą jednostki znaczące. Istnienie w danym systemie zespołu cenemów i pleremów znacznie zwiększa ekonomię budowy tego systemu (możliwe są systemy znakowe złożone wyłącznie z pleremów, wtedy jednak ilość minimalnych pleremów w takich systemach musi być automatycznie stosunkowo duża). Język genetyczny uznać należy za dwuklasowy. Jeżeli przyjąć, że minimalnymi składnikami znaczeń genetycznych są poszczególne aminokwasy, to za system cenemów uznać należy zbiór kilku nukleotydów tworzących łańcuchy DNA i mRNA. Pamiętając o tym, że pozycja zasady w kodonie ma wpływ na rodzaj kodowanego aminokwasu, można też uznać, że poszczególne nukleotydy są minimalnymi jednostkami znaczącymi (minimalnymi pleremami), a za system cenemów uznać atomy tworzące cząsteczki nukleotydów.

Z **produktywnością** (otwartością) systemu znakowego mamy do czynienia wtedy, gdy istnieje w nim potencjalnie nieskończona ilość znaków złożonych. Języki ludzkie posiadają tę cechę: w każdym z nich istnieje potencjalnie nieskończona ilość zdań. Ma ją również język genetyczny. Jeżeli nie istnieją żadne naturalne (fizykochemiczne) ograniczenia liczby potencjalnie istniejących białek, to również liczba istniejących potencjalnie cistronów jest nieskończona.

Interesujące wydaje się występowanie w języku genetycznym, co prawda w

ograniczonym zakresie, jeszcze jednej cechy charakterystycznej dla języków ludzkich, a mianowicie *metajęzykowości*. Cecha metajęzykowości polega na tym, że możliwe jest w danym systemie tworzenie komunikatów o komunikatach. Właśność ta przysługuje m.in. wszystkim językom ludzkim, a nie występuje np. w takim systemie znakowym, jak taniec pszczół (nie można w tańcu pszczół przekazać żadnej informacji na temat samego tańca). Wydaje się, że dobrą ilustracją metajęzykowości języka genetycznego są procesy regulacji syntezy enzymów, a więc np. działanie operonu laktozowego u *Escherichia coli* lub operonu histydynowego u *Salmonella typhimurium*. Operon laktozowy u *Escherichia coli* składa się z promotora, operatora cistronu regulatorowego (kodującego białko zwane represorem) oraz cistronów strukturalnych (kodujących acetylazę, permeazę oraz β -galaktozydazę). Kontakt represora z operatorem hamuje transkrypcję cistronów strukturalnych. Obecność laktozy powoduje utratę powinowactwa represora do operatora — następuje wtedy odblokowanie operatora i transkrypcja cistronów strukturalnych. Można zatem uważać, że regulator spełnia funkcję metajęzykową w odniesieniu do reszty tekstu genetycznego operonu.

Występowanie powyższych cech w określonej grupie procesów molekularno-genetycznych usprawiedliwia więc używanie terminu „język genetyczny”, a także uprawomocnia stosowanie w badaniach języka genetycznego procedur semiotycznych i lingwistycznych. Co więcej, fakt, że językowi genetycznemu przysługują wymienione wyżej uniwersalia ukazuje, że twierdzenia niektórych lingwistów dotyczące łącznego występowania tych cech jedynie w językach ludzkich są nie-
trafne.⁴

Podając wyżej przykłady jednostek języka genetycznego dzieliliśmy je na jednostki planu wyrażania i planu treści. Jednostki planu wyrażania (komunikaty, wyrażenia, etc. genetyczne) to segmenty cząsteczek polinukleotydowych. Natomiast segmenty cząsteczek polipeptydowych traktujemy jako składniki znaczeń genetycznych, a mianowicie jako denotacje: denotacją danego wyrażenia genetycznego jest kodowana przez to wyrażenie jednostka polipeptydowa. Oprócz denotacji, składnikami znaczeń genetycznych są również sensory: sensem danego wyrażenia genetycznego jest ogół funkcji biochemicznych spełnianych przez denotację tego wyrażenia. Zatem denotacją np. pojedynczego kodonu jest kodowany przezeń aminokwas, denotacją cistronu — odpowiedni łańcuch polipeptydowy, denotacją transkryptonu — łańcuch mRNA, etc. Sensem cistronu są funkcje biochemiczne kodowanego przez ten cistron białka; sensem replikonu jest funkcja matrycowa, etc.

⁴Por. John Lyons, *Semantics*, Cambridge University Press, Cambridge 1977, s. 79.

Istnieją liczne analogie między planem wyrażania języka genetycznego a planami wyrażania języków naturalnych (ludzkich). Do najważniejszych należy hierarchiczna struktura obu systemów. Podobnie jak w językach naturalnych, w planie wyrażania języka genetycznego istnieją poziomy, grupujące jednostki tego samego rodzaju. Tak więc kodony, cistrony, transkryptyony, replikony, etc. tworzą poszczególne poziomy języka genetycznego. Poziomy te są hierarchicznie uporządkowane: jednostki poziomu wyższego są określonymi kombinacjami (najczęściej, choć nie zawsze, liniowymi) jednostek poziomu niższego. Począwszy od poziomu cistronów, poziomy języka genetycznego wyposażone są w swoiste znaki interpunkcji; istnieją również specyficzne zależności między jednostkami a znakami interpunkcji różnych poziomów. Za kryterium wydzielenia jednostek genetycznych służą ich funkcje biochemiczne (a więc ich powiązania relacjami genetycznymi z innymi jednostkami) — podobne kryterium stosuje się w lingwistyce przy ustalaniu jednostek powiązanych relacjami syntagmatycznymi i paradygmatycznymi. Przez relacje syntagmatyczne rozumie się w językoznawstwie zależności łączące jednostki tego samego poziomu w wyższe hierarchicznie całości. Jest więc relacją syntagmatyczną np. związek zgody — zależność polegająca na zgodności wykładników informacji gramatycznych (takich jak czas, rodzaj, liczba, przypadek, osoba, etc.) w wyrazach tworzących zdania poprawne. Relacje paradygmatyczne to te relacje między jednostkami tego samego poziomu językowego, które nie są syntagmatyczne (a więc np. synonimia lub podrzędność znaczeniowa wyrazów). W języku genetycznym relacją paradygmatyczną jest np. synonimia kodonów, zaś przykładem relacji syntagmatycznej jest chociażby łączliwość poszczególnych cistronów operonu, wyznaczona przez współdziałanie w procesach biochemicznych kodowanych przez te cistrony związków.

Oprócz podobieństwa w hierarchicznym uporządkowaniu poziomów językowych, pomiędzy językiem genetycznym a językami naturalnymi istnieją również pewne podobieństwa natury syntaktycznej i semantycznej. W planie wyrażania języka genetycznego istnieją minimalne jednostki swobodne posiadające samodzielne znaczenie (denotację) — kodony. Zatem kodony spełniają podobną funkcję jak wyrazy. Kompletnie konstrukcje składniowe (niezależnie od kontekstu) języka genetycznego to cistrony — stąd uzasadnione ich porównanie do zdań. Komunikaty genetyczne większe od cistronów (transkryptyony, replikony, segregony, etc.) pełnią zatem funkcję tekstów genetycznych. Części składowe tekstów genetycznych powiązane są zależnościami semantycznymi (spójność tekstowa) — np. poszczególne cistrony strukturalne operonów kodują enzymy współdziałające w stosownych procesach genetycznych.

Warto zwrócić uwagę, że niektóre jednostki planu treści języka genetycznego,

podobnie jak jednostki jego planu wyrażania, są również wewnątrznie ustrukturowane — chodzi nam w tym momencie o pierwszo-, drugo-, trzecio- i czwartorzędowe struktury białek, ich centra aktywne, etc. Znaczenia komunikatów genetycznych składają się zarówno z obiektów konkretnych (segmenty białek jako denotacje) jak i abstrakcyjnych (funkcje biochemiczne jako sensory). Zarówno plan treści jak i plan wyrażania języka genetycznego ma tego samego rodzaju nośnik materialny — mianowicie segmenty nieregularnych polimerów (inaczej niż w językach naturalnych, w których jednostki planu wyrażania — segmenty wypowiedzi — są obiektami innego rodzaju niż jednostki planu treści — znaczenia, pojęcia, sądy, etc.).

Zasadniczą cechą odróżniającą język genetyczny od języków naturalnych, a także od wielu innych systemów znakowych jest jego bezpodmiotowość: w języku genetycznym brak jest nadawcy i odbiorcy komunikatów. W związku z tym, nie jest sensowne dyskutowanie posiadania przez język genetyczny tzw. uniwersaliów pragmatycznych, tj. cech opisujących związki między znakami systemu a użytkownikami znaków. Fakt bezpodmiotowości języka genetycznego ma istotne znaczenie dla opisu jego struktury syntaktycznej — mianowicie najtrajniejsze wydaje się rozpatrywanie języka genetycznego jako języka rozkazów: każdy cistron (zdanie genetyczne) jest poleceniem dokonania syntezy odpowiedniego łańcucha polipeptydowego. Zatem zdania genetyczne mają prawdopodobnie strukturę syntaktyczną podobną do struktur grup orzeczenia w językach naturalnych (opis procesu wraz z warunkami jego urzeczywistnienia).⁵ Być może, owa cecha składniowa języka genetycznego wpływa na trudności w interpretacji w języku genetycznym niektórych podstawowych pojęć semantycznych (np. pojęcia prawdy).

Fakt istnienia systemu, któremu przysługują wszystkie podstawowe strukturalne uniwersalia językowe, a który jest jednocześnie systemem bezpodmiotowym, ma w naszej opinii ważne implikacje dla postaci definicji pojęcia systemu znakowego. Uważamy, że system znakowy scharakteryzowany powinien być przez cechy strukturalne (uniwersalia językowe). Natomiast klasa systemów podmiotowych (tj. takich, w których występują nadawcy i odbiorcy) powinna być scharakteryzowana jako pewna podklasa klasy wszystkich systemów semiotycznych.

⁵Por. W.A. Ratner, *Molekularno-geneticzeskije sistemy uprawlenija*. Nauka, Nowosybirsk, 1975, s. 81–101.

Możliwości opisu języka genetycznego za pomocą lingwistycznej aparatury pojęciowej

Jak już stwierdziliśmy powyżej, fakt iż pewne procesy genetyczne mają charakter semiotyczny otwiera możliwości stosowania do ich opisu aparatury pojęciowej semiotyki i lingwistyki. Pewne środki lingwistyczne zostały już przez nas wykorzystane: dzieliliśmy jednostki genetyczne na jednostki planu wyrażania (formy) i jednostki planu treści (znaczenia), rozróżnialiśmy składniki znaczeń (denotacje i sensory), mówiliśmy o poziomach językowych i zależnościach między jednostkami genetycznymi (relacje genetyczne, wśród których można wyróżnić, podobnie jak w lingwistyce, relacje syntagmatyczne i paradygmatyczne). W literaturze przedmiotu spotkać można dalsze zastosowania pojęć lingwistycznych do opisu języka genetycznego — od najprostszych, jak np. badania dotyczące ilości informacji zawartej w kodzie genetycznym, do bardziej zaawansowanych teoretycznie, jak np. problem znalezienia gramatyki języka genetycznego. Problem ten, dotyczący znalezienia reguł budowy składniowej komunikatów genetycznych ma fundamentalne znaczenie dla genetyki molekularnej. Można proponować różnorakie podejścia do tego zagadnienia. Dla przykładu, poszukiwana gramatyka może być tzw. gramatyką generatywną, podającą formalne reguły budowy składniowej komunikatów genetycznych (reguły te muszą oczywiście brać pod uwagę również aspekty semantyczne języka genetycznego).⁶ Być może, ze względu na bezpodmiotowy charakter języka genetycznego, użyteczne byłoby poszukiwanie analogii między językiem genetycznym a językami programowania.⁷ Inną możliwością zastosowania lingwistycznej aparatury pojęciowej może być opis języka genetycznego za pomocą tzw. gramatycznych sposobów wyrażania. Punktem wyjścia w takim podejściu są trzy zbiory obiektów: segmentów, znaczeń i sposobów wyrażania. Segmenty to jednostki planu wyrażania, znaczenia zaś to jednostki planu treści (zob. wyżej). Sposoby wyrażania traktujemy jako zależności łączące segmenty i znaczenia (sposoby, w jakie dany segment wyraża określone znaczenie: np. sposób, w jaki denotacja wyrażenia zależy od jego struktury, sposób, w jaki sens wyrażenia zależy od struktury jego denotacji). Do sposobów wyrażania (w języku genetycznym) możemy zaliczyć: pierwszo-, drugo-, trzecio- i czwartorzędowe struktury białek, tworzenie α -spiral, β -struktur,

⁶Prosty przykład gramatyki formalnej dla kodu genetycznego podaje Zdzisław Pawlak w popularnej pracy *Gramatyka i matematyka*. PZWS, Warszawa 1965, s. 75–94.

⁷Por. Ratner, op. cit. s. 100.

mostków dwusiarczkowych, mostków wodorowych etc. Denotacje komunikatów genetycznych wyrażane są przez uporządkowanie zasad w kodonach i kodonów w cistronach. Sekwencje zasad w segmentach polinukleotydowych wyznaczają bezpośrednio pierwszorzędowe struktury białek, a w ostatecznym rozrachunku również ich drugo-, trzecio- i czwartorzędowe struktury. Niektóre z wymienionych wyżej sposobów wyrażania traktować możemy jako sposoby urzeczywistniania sensów genetycznych (a więc funkcji biochemicznych pełnionych przez jednostki genetyczne). Znaki interpunkcji wyrażają określone informacje gramatyczne: rozpoczęcie i zakończenie stosownych procesów (transkrypcji, translacji, replikacji, segregacji, etc.). Idea opisu języka genetycznego za pomocą segmentów, znaczeń i sposobów wyrażania jest na razie dopiero wstępnie opracowywana i wiele zagadnień z nią związanych wymaga dalszych badań. Dla przykładu, jednostki planu treści (denotacje i sensy) powinny zostać poklasyfikowane w odpowiednie kategorie, podobnie jak jednostki planu wyrażania podzielone są np. w poziomy językowe. Podział taki byłby odpowiednikiem dzielenia znaczeń językowych na leksykalne, gramatyczne, etc. (znaczenie leksykalne wskazuje na odpowiednik pozajęzykowy danego wyrażenia, zaś znaczenia gramatyczne — jak np. czas, rodzaj, liczba, przypadek — traktować można jako modyfikatory znaczeń leksykalnych). Istotne w omawianym podejściu byłoby też znalezienie reguł, według których znaczenie wyrażenia złożonego zależy od znaczeń jego składników (tzn. opis sposobu realizacji zasady Fregego w języku genetycznym). Reguły takie mogłyby stanowić podstawę opisu języka genetycznego za pomocą aparatury pojęciowej teorii gramatyk formalnych.

Ideę opisu przez segmenty, znaczenia i sposoby wyrażania uważamy za interesującą m.in. z dwóch powodów: po pierwsze, podejście takie jest uniwersalne — da się zastosować do wszelkich systemów językowych⁸; po drugie, badanie związków między segmentami, znaczeniami i sposobami wyrażania stanowi może podstawę do uniwersalnej typologii systemów semiotycznych. Być może, również niektóre problemy specyficznie genetyczne mogłyby być łatwiej rozwiązane przy użyciu środków lingwistycznych w ich formułowaniu.

⁸Idea takiego opisu w odniesieniu do języków naturalnych przedstawiona została w pracy A.F. Majewicz, J. Pogonowski, *On categorial marking in natural languages*, *Lingua Posnaniensis* XXVI, Poznań 1983, 55–68.