

Siergiej I. Gindin

Przykłady analizy struktury tekstu za pomocą słowników semantycznych

Pamiętnik Literacki : czasopismo kwartalne poświęcone historii i krytyce literatury polskiej 66/1, 283-325

1975

Artykuł został zdigitalizowany i opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

SIERGIEJ I. GINDIN

PRZYKŁADY ANALIZY STRUKTURY TEKSTU ZA POMOCĄ SŁOWNIKÓW SEMANTYCZNYCH

Pamięci Grigorija D. Dreidena

Wprowadzenie

Przez strukturę tekstu rozumie się zwykle całokształt wyróżnionych w określony sposób części tekstu (subtekstów), które nie nakładają się na siebie¹ i wyczerpują całość tekstu, oraz całokształt związków pomiędzy tymi częściami. Istnieje obecnie dość dużo rozmaitych koncepcji teoretycznych odnoszących się do struktury tekstu i metod jej wykrywania. Wiele z nich opiera się na założeniu, że jednym z zasadniczych środków powiązań między subtekstami są powtórzenia leksykalne (a także ich analogony funkcjonalne w rodzaju zaimków — substytutów)². W odróżnieniu od pionierskiej pracy Żyrmunskiego [2], w której powtórzenia leksykalne ujmowano jako dodatkowy środek stylistyczny, jako oznakę tekstów określonego stylu, w propozycjach z lat pięćdziesiątych i sześćdziesiątych (zob. przede wszystkim [3—5]) rozpatrywano je już poza wszelkimi związkami ze stylistyką, jako środek gramatyczny, czysto języ-

[Siergiej Josifowicz Gindin — pracownik naukowy Moskiewskiego Instytutu Pedagogicznego Języków Obcych im. M. Thoreza, w Laboratorium przekładu maszynowego; doktorat (pracę kandydacką) obronił w 1972 r. Ważniejsze prace w „Bibliografii” na końcu artykułu.

Przekład według wyd.: С. И. Гиндин, *Опыты анализа текста с помощью семантических словарей. Статья I*. „Машинный перевод и прикладная лингвистика” 16, 1972, s. 42—112. Wszystkie skróty zastosowane w wersji polskiej pochodzą od autora.]

¹ Nieprzecinanie się występuje w wypadku pewnego zarejestrowanego rozczłonkowania tekstu na subteksty, natomiast subteksty wyróżniane przy rozmaitych podziałach tekstu mogą w zasadzie nakładać się na siebie.

² Por. krytykę podejścia opierającego się na wykrywaniu powtórzeń, w [1]. Jej autor jednakże opiera swoją koncepcję również na pojęciu powtórzenia, tyle tylko, że nie leksemu, lecz kategorii gramatycznej.

kowy. Gdy pokazano, że powtórzenie stanowi uniwersalną zasadę budowy tekstu linearnego, wpływającą z samego faktu jednościi tekstu i jego jednoczesnego rozczłonkowania na subteksty (zob. [6, rozdz. 1] oraz nieco przestarzałe [7]), takie ujęcie powtórzeń leksykalnych uzyskało uzasadnienie teoretyczne. Zarazem jednak stało się oczywiste i potwierdzone przez praktykę, że powtórzenia leksykalne jako środek powiązania subtekstów nie mają charakteru uniwersalnego. Świadczy o tym istnienie takich tekstów, w których strukturze powtórzenia leksykalne nie mogą odgrywać istotniejszej roli z uwagi na ich małą liczbę lub nawet nieobecność. Jak wykazano w [6—9], powtórzenie leksemów stanowi zaledwie jeden spośród przewidywanych przez język środków realizacji powtórzeń, jedną z licznych odmian powtórzenia semantycznego, określonego jako powtórzenie pewnego znaczenia, pewnego zespołu sensów, któremu nie towarzyszy powtórzenie żadnego ze sposobów wyrażania tego znaczenia. Tamże wysunięto hipotezę, że metodyka analizy struktury tekstu, polegająca na wykrywaniu w tekście powtórzeń nie tylko leksykalnych, ale również dowolnych powtórzeń semantycznych, odznaczałaby się i większą mocą rozstrzygającą, i znacznie szerszym zakresem jej stosowalności.

Niniejszy artykuł jest próbą praktycznego udowodnienia tej hipotezy na materiale dwóch tekstów rosyjskich; analiza ich struktury oparta na samych powtórzeniach leksykalnych okazuje się praktycznie bezowocna.

Drugie zadanie tego artykułu polega na udowodnieniu wysuniętej przeze mnie tezy [7—8], że procedura wykrywania powtórzeń semantycznych może być nie mniej obiektywna od procedury wykrywania powtórzeń leksykalnych, pod warunkiem, że będziemy się posługiwali pewnym słownikiem semantycznym (lub systemem słowników semantycznych) i jako człony powtórzenia będziemy rozpatrywali tylko te słowa tekstu, które wchodzą w tym słowniku (lub słownikach) do wystarczającej liczby ogólnych klas semantycznych. Dlatego właśnie, choć nieleksykalne powtórzenia semantyczne³ są w analizowanych tekstach „widoczne gołym okiem” i przez to mogłyby być wyróżnione na podstawie intuicji językowej, wyodrębnienia powtórzeń dokonaliśmy tu automatycznie — na podstawie świadectw wybranych słowników semantycznych⁴. [...]

³ Zaproponowany w [7] termin „powtórzenie pola” [полевой повтор], a wprowadzony w celu oznaczenia nieleksykalnych powtórzeń semantycznych, wydaje się dziś nieudany.

⁴ Harris, który jako pierwszy wypracował uogólnienie pojęcia powtórzenia leksykalnego, tj. ekwiwalencję dystrybucyjną, już w 1957 r. zauważył, że w zasadzie możliwe jest uogólnienie alternatywne — „zgrupowanie słów zgodnie z polami słownikowymi, do których one wchodzą” [4]. Zaproponowana przez nas metodyka z jej wykorzystywaniem słowników semantycznych realizuje w istocie tę właśnie marginesową uwagę Harrisa. Porównanie rezultatów analizy konkretnych

1. Tekst: język ? tekst: słownik, czyli o prawomocności zastosowania słownika semantycznego w celu wykrywania powtórzeń semantycznych

Ale czy słuszne jest posługiwanie się słownikiem semantycznym jako narzędziem obiektywizacji wyodrębniania powtórzeń semantycznych? Czy w ten sposób wyodrębnione powtórzenia będą charakteryzowały nie tyle tekst i język, co sam słownik?

W podanych wątpliwościach istotne jest nie to, że za pomocą słownika można wyodrębnić nie wszystkie istotne dla tekstu powtórzenia⁵. Jasne jest bowiem, że w dowolnym tekście i na dowolnych jego poziomach możliwe są zjawiska przypadkowe, które nie wchodzą nie tylko do słownika, ale i do języka, a które dają się wyprowadzić i wytłumaczyć wyłącznie z samego tekstu. O wiele poważniejsza jest druga część wątpliwości — niekompletność słownika wobec języka. Dowolny słownik semantyczny stanowi zaledwie „makiętę” [14], model leksykalno-semantycznego systemu językowego, rozumianego jako całość realnie istniejących w ontologii języka (i określających tożsamość członów powtórzeń) rozróżnień lub zbieżności składu leksykalnego na klasy elementów pod jakimś względem jednakowych. Dlatego najdoskonalszy nawet słownik, jak i każdy model, będzie niepełnym i nieadekwatnym odzwierciedleniem tego systemu, i dlatego również rezultaty wyodrębniania powtórzeń za pomocą słownika będą w jakimś sensie niepełne i nieadekwatne względem języka. Wada ta (właściwa zresztą wszelkim badaniom, w których system lub normę językową traktuje się — zgodnie z nakre-

tekstów, dokonanej na podstawie ekwiwalencji dystrybutywnej oraz na podstawie powtórzenia semantycznego, zasługuje, jak się zdaje, na specjalne studia. Wobec tekstów, w których nie ma powtórzeń leksykalnych, pojęcie ekwiwalencji dystrybutywnej, w odróżnieniu od powtórzenia semantycznego, w ogóle nie jest przykładowe (ponieważ w jego rekursywnym określeniu na pierwszym miejscu występuje powtórzenie leksykalne), ale im dłuższy jest tekst, tym mniejsze podobieństwo, iż nie wystąpią w nim powtórzenia leksykalne. Jeżeli zatem przyjmiemy, że znane hipotezy Szajkiewicza [10] i Szreidera [11], według których istnieje pewna zgodność między odległością dzielącą słowa w tekście a odległością dzielącą je w słowniku, sprawdzają się tym lepiej, im dłuższy jest tekst, to możemy oczekiwać, że przy narastaniu tekstu do nieskończoności rezultaty analizy opartej na powtórzeniu semantycznym i na ekwiwalencji dystrybutywnej będą coraz bardziej zbliżały się do siebie.

⁵ Zresztą, jak wykazuje nasza próba statystycznej rekonstrukcji semantyki Tiutczewa, związki semantyczne, podlegające właściwej rekonstrukcji tylko na podstawie samych tekstów, mogą dokładnie tak samo nie wystarczać do analizy ich struktury, jak i związki ustanawiane na podstawie samego języka (słownika). Jak się zdaje, pełne wymodelowanie odbioru tekstu powinno łączyć w sobie oba te podejścia i obejmować zarówno analizę tekstu w oparciu o słownik, jak i pewną korektę słownika w oparciu o tekst. O analogicznym problemie na gruncie matematycznym zob. [11] (por. zastosowanie pojęcia „tezaurus” w opisach odbiornika informacji w teorii informacji [13, s. 161]).

ślonym zadaniem — jako coś danego i utrwalonego) okazuje się zarazem i zaletą: będąc skończonym, niepełnym i niejako z konieczności nieadekwatnym słownik semantyczny pozostaje w zasadzie wobec tekstu w tym samym stosunku, co i nosiciele danego języka. Żaden z nosicieli języka również bowiem nie włada całością semantycznego systemu językowego i w jakimś sensie jego przyzwyczajenia semantyczne nieuchronnie odbiegają od norm użycia, niemniej jednak nie przeszkadza mu to w rozpoznawaniu struktury tekstów. Wszelkie rozumienie (a rozpoznawanie struktury tekstu jest zarazem częścią rozumienia tego tekstu) jest rozumieniem opartym na pewnym, jakkolwiek dopuszczającym stopniowe wzbogacanie, jednakże bezwarunkowo ograniczonym zasobie wiadomości apriorycznych. Dlatego procedura wykrywania powtórzeń za pomocą słownika może służyć jako pewien „wypośrodkowany pod względem uzusu” model rozpoznawania tych powtórzeń (a zatem — również struktury tekstu) przez nosicieli języka⁶ oraz jako część ogólnego modelu rozumienia tekstu w warunkach pewnego utrwalonego poziomu władania językiem, modelu będącego odwrotnością tradycyjnego deskryptywnego modelu rekonstrukcji języka na podstawie tekstu. To wszystko pozwala sądzić, że badania struktury tekstów za pomocą dostatecznie sprawdzonych i autorytatywnych słowników semantycznych nie tylko dają określone korzyści praktyczne, ale są również wystarczająco poprawne pod względem metodologicznym.

2. Przyjęte założenia oraz metodyka analizy

Siewbo [3] i Harris [4] poprzedzają bezpośrednią analizę struktury tekstu sprowadzeniem składających się na ten tekst zdań do odpowiedniej postaci standardowej — najbardziej elementarnej u Siewbo i naj-

⁶ Tym samym więc przyjmujemy, że słownik semantyczny (jeżeli oczywiście spełnia minimum wymagań naukowych) odzwierciedla nie tylko semantyczny system języka, ale też i semantyczne przyzwyczajenia nosicieli tego języka. Trzeba i powinno się mówić o indywidualnych zwyczajach semantycznych mówiących, o odchyleniach, często bardzo poważnych, od norm uzualnych (por. dychotomię bliższych i dalszych znaczeń oraz jej korelację wobec dychotomii „narodowe” — „osobiste” w [16, s. 146—147]. Jednakże twierdzenie, że związki, jakie „zachodzą między słowami w psychice mówiących”, są „zupełnie innego rodzaju” niż związki językowe i że „różnego rodzaju ‘pola semantyczne’, ‘grupy semantyczne’ itd. powstające na podstawie świadomego porównywania znaczeń słownikowych, są irrelevantne wobec mechanizmów mowy” [15, s. 3—4], jest znacznie mocniejsze i oznacza albo tyle tylko, że dany słownik semantyczny nie odzwierciedla semantycznego systemu danego języka, albo że mówiący nie znają semantyki danego języka. Nieprzypadkiem więc jako przykład rażącego odchylenia empirycznie uzyskanych semantycznych grup skojarzeniowych od zarejestrowanych w słownikach semantycznych przytacza się (zob. [17]) rezultaty badań małych dzieci, które nie są jeszcze właściwie nosicielami danego języka.

bardziej typowej dla danego tekstu u Harrisa. W niniejszej pracy zdecydowaliśmy się analizować tekst w jego postaci pierwotnej — lingwistyczne korzyści takiego podejścia są oczywiste, pewne zaś nieuniknione straty będziemy odnotowywali w trakcie analizy. Ponieważ rezygnujemy z transformacji tekstu, cała reszta analizy struktury tekstu za pomocą słowników semantycznych sama przez się rozpada się na trzy etapy:

1. Rozczłonkowanie tekstu na jednostki terminalne, tj. na takie minimalne subteksty, z których składają się subteksty bardziej złożone;

2. Wyodrębnienie powtórzeń semantycznych za pomocą wybranego słownika;

3. Zbudowanie struktury tekstu, odzwierciedlającego związki jednostek, na które został rozcłunkowany tekst w pierwszym etapie, jak również subtekstów bardziej złożonych, za pomocą powtórzeń wyodrębnionych w drugim etapie. W tej właśnie kolejności będziemy rozpatrywali dalej metodykę analizy.

2.1. Założenia dotyczące minimalnych subtekstów

Wyobrażenie o strukturze tekstu z konieczności zakłada wyobrażenie o jednostkach terminalnych, elementach tej struktury, „minimalnych subtekstach”, których wyodrębnieniem analiza się kończy, a których zespolenie podczas syntezy daje większe jednostki (subteksty) i w końcu — sam tekst. W istniejących pracach jako subteksty minimalne traktuje się zazwyczaj zdania. [...] Tymczasem, jak podpowiada intuicja, ponieważ tekst interesuje nas jako formacja językowa, za terminalne jednostki jego struktury powinniśmy uważać jednostki właśnie językowe, nie zaś po prostu „powycinane” z tego tekstu „odcinki” o jednakowej lub jakoś inaczej utrwalonej długości. Bynajmniej nie jest rzeczą oczywistą, że nie można zbudować opisu struktury tekstu (choćby wyjątkowo skomplikowanego), którego jednostkami terminalnymi byłyby np. morfemy lub fonemy, a zdania nie powstawałyby nawet jako subteksty pośrednie.

Przyjęcie za jednostkę terminalną właśnie zdania zakłada przyjęcie przynajmniej jednego spośród następujących dwóch postulatów:

a) postulat strukturalnej prostoty opisu: opis powinien być maksymalnie prosty pod względem strukturalnym, tj. nie tylko najkrótszy, ale i zbudowany z niewielkiej liczby możliwie niezależnych bloków;

b) postulat ontologicznej doniosłości zdania: pomimo to, iż sens tekstu może się rozmaicie rozkładać między zdaniami, sam fakt rozcłunkowania tekstu na zdania nie jest rzeczą przypadku, lecz konieczną właściwością jego struktury (zdanie jest „kwantem” tekstu) i dlatego powinien znaleźć odpowiednie odzwierciedlenie w opisie tej struktury.

Ponieważ istnieje opis struktury zdania, przyjęcie pierwszego postulatu pozwala udowodnić, że wybór zdania w charakterze jednostki ter-

minalnej znacznie upraszcza opis struktury tekstu i dlatego jest pożyteczny, ale wcale nie pozwala udowodnić, iż wybór ten jest najkorzystniejszy i dlatego nieodzowny.

Natomiast przyjęcie także drugiego postulatu pozwala wykazać, że wybór taki jest nieunikniony: ponieważ podział na zdania stanowi nieodłączną właściwość struktury tekstu, podział ten powinien być odzwierciedlony również w opisie tej struktury. A więc zdanie musi figurować w tym opisie przynajmniej jako subtekst przejściowy. Dlatego jednocześnie operowanie regułami kombinatoryki jednostek terminalnych (lub rozczłonkowania na jednostki terminalne), pomijającej zdania, prowadziłoby do zdublowania, skomplikowania opisu. I dlatego właśnie operowanie powinno być pominięte. W związku z tym, że struktura zdania jest już opisana, resztę opisu całkiem naturalnie można podzielić na dwie części (dwa etapy): opis struktury zdania w terminach jednostek mniejszych i ich kombinacji oraz opis struktury tekstu w terminach zdań i ich kombinacji.

Pierwszy postulat jest ogólnie przyjętym postulatem naukowym, drugi zaś, jak się wydaje, w wystarczającym stopniu opiera się na znanych z praktyki faktach, że dowolny, choć trochę sensowny tekst składa się ze zdań (lub segmentów „zdaniopodobnych”, por. *Moja twoja nie rozumieć*) i że jednostki poziomów „przedzdaniowych” mogą tworzyć samodzielny tekst tylko wówczas, gdy jednocześnie tworzą także zdanie (por. łacińskie *i* i rosyjskie *и*). Dlatego, przyjmując oba postulaty, w charakterze minimalnego subtekstu my również chcielibyśmy rozpatrywać zdanie.

Trudności tkwią jednakże w niedostatecznej ścisłości samego terminu „zdanie”. [...] Dlatego musimy go zastąpić pewną roboczą analogią, zachowującą wszystkie istotne właściwości zdania, ale posiadającą ściśle określony zakres i treść. W charakterze takiej analogii wykorzystaliśmy pojęcie „maksyntemy”, wprowadzone przez następującą definicję indukcyjną:

1. Zdanie proste jest maksyntemą.
2. Zdanie złożone podrzędnie jest maksyntemą.
3. Każda ze współrzędnych części zdania złożonego współrzędnie jest maksyntemą.
4. Innych maksyntem nie ma ⁷.

⁷ Terminy „zdanie proste”, „podrzędnie złożone” oraz „współrzędna część zdania współrzędnie złożonego” są już wystarczająco precyzyjne; można więc uważać, że ich definicja operatywna określana jest przez listę ich możliwych struktur w odpowiednich rozdziałach zwykłej składni „wewnątrzzdaniowej”. Sam termin „maksyntema” został utworzony jako skrót terminu „maksymalna jednostka syntaktyczna”, zaproponowanego w [18, s. 29], jednakże w odróżnieniu od Dolininy nie stosujemy tego terminu wobec zdań podrzędnych.

Pojęcie maksyntemy i jej intuicyjnego eksplikandum — zdania — jest przykładowe do tekstów i prozatorskich, i wierszowanych. Specyficzna osobliwość struktury tekstów wierszowanych polega na jej „podwojeniu” — obok występującego również w tekstach prozatorskich rozczłonkowania kompozycyjno-znaczeniowego (przekształcającego się w obrębie zdania w rozczłonkowanie syntaktyczne) mamy tu równoległe i całkiem autonomiczne rozczłonkowanie wierszowe, którego jednostkę terminalną stanowi już nie maksyntema (zdanie), lecz linijka (wers). Pomiedzy elementami tego drugiego rozczłonkowania istnieją specyficzne powiązania (rymowe i rytmiczne), ale rozciągają się na nie także związki, które zachodzą między elementami rozczłonkowania pierwszego typu — powtórzenia leksykalne i semantyczne, nawiązania syntaktyczne itp. Dlatego, nawet pomijając owe specyficzne związki rymowe i rytmiczne, pełne odzwierciedlenie struktury tekstu wierszowanego wymaga rozpatrzenia jednych i tych samych związków w dwóch zbiorach — maksyntem i wersów — oraz zbudowania odpowiednio dwóch schematów struktury tekstu (przy uwzględnieniu specyficznych związków wierszowych liczba schematów wzrosłaby do 4). Ponieważ w pracy niniejszej interesują nas problemy analizy dowolnych tekstów, dlatego analizując także teksty wierszowane, będziemy tworzyli tylko schematy związków między maksyntemami. Schematy, których jednostkami terminalnymi są wersy, czytelnik z łatwością zbuduje na podstawie przytoczonych wykazów powiązań. Pewne najbardziej interesujące relacje pomiędzy schematami obu tych rozczłonkowań postaramy się jednakże wydobyć.

2.2. Metodyka automatycznego rozpoznawania powtórzeń semantycznych za pomocą słowników semantycznych

2.2.1. Schemat ogólny. Po rozczłonkowaniu tekstu na maksyntemy przystępujemy do wyodrębnienia powtórzeń. Procedura wyodrębniania powtórzeń za pomocą słownika semantycznego dzieli się na dwa jakościowo różne etapy: etap analizy semantycznej, na którym słowa tekstu są zestawiane z ich „obrazami semantycznymi”, tj. z objaśnieniami przypisywanymi tym słowom w wykorzystywanym słowniku semantycznym⁸, oraz etap właściwego wyodrębniania powtórzeń, na którym

⁸ Każdy słownik semantyczny narzuca grupowanie leksyki odpowiedniego języka według klas semantycznych — albo w postaci jawnej, jak w słownikach synonimicznych i ideograficznych lub słowotwórczych [19], albo w postaci utajonej, jak zwykle słowniki objaśniające lub słowotwórcze typu [20], w których w celu uzyskania klas trzeba odnaleźć słowa o jednakowych lub nakładających się ob-

spośród wszystkich przeanalizowanych na pierwszym etapie słów wybiera się jako tworzące powtórzenie te słowa, których obrazy semantyczne odznaczają się wystarczająco mocnym przecięciem. Podstawowe problemy pierwszego etapu to α i β , gdzie α polega na odnalezieniu obrazów dla słów homonimicznych, którym w danym słowniku odpowiada kilka jednoimiennych wejść (jak i w innych zadaniach opracowania tekstu problem ten, niemal nie istniejący przy ręcznej, „ludzkiej” analizie, staje się wyjątkowo istotny i trudny podczas automatyzacji procedury); β zaś stanowi wybór poziomu, którego klasy powinny tworzyć obrazy semantyczne; problem ten powstaje, jeżeli słownik charakteryzuje się organizacją hierarchiczną (jak np. w [27]. Zasadniczy problem drugiego etapu — γ — polega natomiast na wyborze właściwego kryterium podobieństwa znaczeniowego, tj. na ustaleniu, jaki stopień przecięcia się obrazów semantycznych dwóch słów można uważać za wystarczający dla stwierdzenia powtórzenia semantycznego. Rozwiązanie każdego z tych problemów oraz konkretna metodyka realizacji obu etapów procedury zależy od tego, jakie mianowicie słowniki zostały wybrane do danej analizy.

2. 2. 2. Metodyka wykrywania powtórzeń za pomocą wybranych słowników „przedideograficznych”. Jak już mówiliśmy w [6—8], rozmaite poszczególne odmiany powtórzeń semantycznych wymagają przy ich wykrywaniu różnych szczegółowych słowników semantycznych. W niniejszej pracy wykorzystaliśmy: do wykrywania rdzeniowych i afiksalnych powtórzeń semantycznych — słownik Potichy [20]; przy wykrywaniu powtórzeń synonimicznych — słowniki Aleksandrowa [21] i *Słownik synonimów języka rosyjskiego* [22, t. 1] ([19] i [22, t. 2] dotarły do nas już po ukończeniu tej pracy). Do wykrywania częściowych powtórzeń semantycznych korzystaliśmy także ze słownika Wwiedenskiej [23] oraz z aparatu teorii funkcji leksykalnych (parametrów semantycznych): jakkolwiek z opartego na tej teorii słownika „objaśniająco-kombinatorycznego” języka rosyjskiego opublikowano jak dotąd zaledwie kilkadziesiąt haseł słownikowych, jednakże dostatecznie wyraźne sprecyzowanie większości funkcji (zob. [24] i [25] oraz ostatni chwilowo wykaz funkcji w [26]) pozwala dość obiektywnie ustalać parametryczne związki między słowami.

Wszystkie te słowniki odzwierciedlają jeden poziom. Metodykę wykrywania powtórzeń za pomocą tych słowników określają przede wszystkim umieszczone w nich Zasady posługiwania się słownikiem i przez to dość dobrze poddaje się ona algorytmizacji. W wypadku [20] każdy ko-

jaśnieniach. Dlatego bez ograniczenia wspólnoty można uważać, że obraz semantyczny, odnajdywany dla słowa tekstu na pierwszym etapie, stanowi szereg klas, do których słowo to wchodzi zgodnie z danym słownikiem.

lejnny wyraz tekstu podlega rozczłonkowaniu według tego słownika na morfemy, a następnie uzyskany podział porównuje się do takich samych podziałów wszystkich następnnych wyrazów tekstu. Homonimia wejść w [20] nie występuje, za kryterium podobieństwa znaczeniowego przyjmuje się tu (przy wyróżnieniu powtórzeń rdzeniowych) występowanie przynajmniej jednego wspólnego morfemu rdzeniowego. W wypadku [21—22] homonimia wejść okazuje się natomiast zjawiskiem powszechnym, albowiem pokaźna część słów wchodzi do więcej niż jeden szeregów synonimicznych. Za semantyczny obraz wyrazu homonimicznego uznawaliśmy wszystkie szeregi synonimiczne, do których wyraz ten wchodzi w [21] (odpowiednio w [22, t. 1]; tym samym za obraz uznawaliśmy teoretyczne mnogościowe zespolenie obrazów wszystkich znaczeń słowa, zanotowanych w danym słowniku). Kryterium podobieństwa znaczeniowego było minimalne: wyrazy uznawano za powtórzenie synonimiczne, jeżeli wchodziły one przynajmniej do jednego wspólnego szeregu synonimicznego przynajmniej w jednym z tych słowników (przy tym, korzystając z [21], szeregi „pokrewne”, powiązane ze sobą odsyłaczem „por.”, łączyliśmy w jeden wspólny szereg). W sposób analogiczny traktowaliśmy homonimy i kryteria podobieństwa również w wypadku [23]. Wykorzystywaliśmy także łącznie dwa słowniki, co umożliwiło wyodrębnienie powtórzeń bardziej skomplikowanych, będących superpozycją dwóch elementarnych, np. kiedy jeden z wyrazów tekstu jest synonimem nie samego innego wyrazu tekstu, lecz jego nieobecnego w tekście antonimu.

Podczas wyróżniania powtórzeń za pomocą aparatu funkcji leksykalnych dwa słowa uważaliśmy za przynależne do wspólnej „klasy parametrycznej” wtedy i tylko wtedy, jeżeli jedno z nich jest znaczeniem pewnego, choćby nawet bardzo skomplikowanego, parametru drugiego słowa.

2. 2. 3. W y b ó r s ł o w n i k a i d e o g r a f i c z n e g o. Słownik ideograficzny (według terminologii Szczerby — „ideologiczny”; dalej określamy skrótem SI) stanowi najbardziej pełny i uniwersalny model leksykalno-semantycznego systemu języka [14] i w zasadzie może być środkiem wykrywania różnorodnych odmian powtórzenia semantycznego. Badania nasze napotkały jednakże poważne trudności z tej prostej przyczyny, że dla języka rosyjskiego SI na razie nie istnieje⁹. Dlatego musieliśmy posłużyć się przekładem: słowa tekstu tłumaczyliśmy na język obcy, posługując się słownikami dwujęzycznymi, następnie dokonywaliśmy analizy semantycznej uzyskanych ekwiwalentów według istniejącego SI tego języka i dopiero potem w określony sposób zestawialiśmy re-

⁹ Nieliczne rosyjsko-obcojęzyczne słowniki tematyczne są stanowczo niezadowolające (zob. [14, s. 40—42]).

zultaty tej analizy („obrazy semantyczne”) z wyjściowymi wyrazami rosyjskimi. Sam fakt przekładu nie powinien nas niepokoić — ostatecznie, jak wykazała semantyka teoretyczna, objaśnianie znaczeń słów zawsze (nawet w zwyczajnym słowniku jednojęzycznym) polega na ich przekładzie na pewien inny język („metajęzyk” autora słownika), co więcej — przekład na język obcy już nie raz stosowano w pracach z zakresu semantyki. Dlatego, wybierając do analizy struktury tekstu obcojęzyczny SI, powinniśmy zadbać o to, aby: a) wybrany słownik dawał dokładny i szczegółowy obraz semantyki swego języka; b) metodyka przekładu dawała wystarczająco adekwatne ekwiwalenty i przy tym była na tyle sformalizowana, by każdy dokonujący tłumaczenia zawsze uzyskiwał dla każdego wyrazu rosyjskiego jedno i te same ekwiwalenty obcojęzyczne¹⁰; i wreszcie c) aby metodyka przenoszenia rezultatów analizy semantycznej tych ekwiwalentów na słowa wyjściowe odznaczała się należytą ostrożnością i również była sformalizowana.

Jako podstawową „wykładnię” semantyczną wybraliśmy znany angielski *Thesaurus* Rogeta (w wariancie R. A. Dutcha [27]), ponieważ stuletnia praktyka jego funkcjonowania, uzupełnienia oraz jego wpływ na strukturę wszystkich późniejszych słowników ideograficznych stanowią gwarancję jego pełności i dokładności, natomiast logiczne zasady jego budowy redukują do możliwego minimum niebezpieczeństwo przeniesienia do tekstu rosyjskiego obcych mu relacji języka angielskiego¹¹.

2. 2. 4. Metodyka automatycznego wykrywania powtórzeń semantycznych za pomocą obcojęzycznego SI. Przy zastosowaniu słownika obcojęzycznego procedura analizy rozpada się już nie na dwa, lecz na trzy etapy: etap tłumaczenia, na którym dla wyrazów rosyjskich tekstu wyjściowego odnajdujemy ich ekwiwalenty obcojęzyczne; etap analizy semantycznej, który obejmuje analizę ekwiwalentów uzyskanych za pomocą SI oraz przeniesienie tych rezultatów na słowa tekstu wyjściowego; etap, na którym spośród prze-

¹⁰ Jeżeli analiza struktury tekstu za pomocą słownika semantycznego danego języka stanowi, jak już mówiliśmy w § 1, pewnego rodzaju model odbioru struktury tekstu przez nosicieli danego języka, to analiza za pomocą obcojęzycznego słownika ideograficznego (pod warunkiem trafnego wyboru metodyki tłumaczenia) może służyć jako wypośredkowany model odbioru struktury tekstu przez kogoś, dla kogo język tekstu nie jest językiem ojczystym. [...]

¹¹ [...] Innym środkiem, pozwalającym nieco złagodzić zasygnalizowane niebezpieczeństwo, może być kontrolne ponowienie analizy według SI jakiegoś trzeciego języka oraz zestawienie uzyskanych wyników. Analizy naszych tekstów ponowiliśmy według SI języka francuskiego [28], albowiem jego „grupy analogiczne” zebrane wokół wyrazu centralnego można przyrównać do haseł Rogeta. Jak się okazało, rezultaty analizy według Maqueta nie są sprzeczne z rezultatami analizy według Rogeta, choć są znacznie uboższe (zob. [6, rozdz. III, §§ 2, 6]).

analizowanych słów wykrywamy słowa realizujące relację powtórzenia semantycznego. W tej kolejności podajemy niżej szczegóły zastosowanej metodyki.

2. 2. 4. 1. Etap tłumaczenia. W celu algorytmizacji procedury tłumaczenia można całkiem naturalnie (jak czyni to każdy człowiek, który tłumaczy jakiś tekst) wykorzystać słowniki dwujęzyczne. Uwzględniając wspomnianą wyżej modelującą funkcję analizy struktury tekstu za pomocą obcojęzycznego słownika ideograficznego, staraliśmy się tak wybrać przeciwległe słowniki translatorskie [29, 30], aby nie za bardzo się różniły zarówno pod względem ich orientacji, jak i pod względem czasu powstania, oraz — o ile jest to możliwe — aby należały nawet do tego samego autora.

Natomiast zasadniczy warunek, którego spełnienia powinno się wymagać od rezultatu adekwatnego tłumaczenia, to *odwracalność* tego przekładu: jako ekwiwalenty obcojęzyczne wyjściowego wyrazu rosyjskiego powinno się wybierać tylko takie warianty proponowane przez słownik rosyjsko-obcojęzyczny, przy których tłumaczeniu słownik przeciwległy — obcojęzyczno-rosyjski — proponuje z kolei wśród innych wariantów także wyjściowy wyraz rosyjski. Ponadto zakładając, że przekłady w słownikach dwujęzycznych podawane są nie w dowolnej kolejności, lecz na ogół w kolejności odpowiadającej mniej więcej zmniejszaniu się typowości (akceptacji, powszechności, prawdopodobieństwa) ich użycia jako ekwiwalentów słowa tłumaczonego, dla wyjściowego słowa rosyjskiego, wybieraliśmy pierwszy ekwiwalent spośród jego odwracalnych przekładów (tj. nazwany na pierwszym miejscu w słowniku rosyjsko-obcojęzycznym)¹². Na przykład dla rosyjskiego wyrazu *злость* 'złość, troska, utrapienie' [29] proponuje następujący szereg przekładów: *wickedness, spite, malice, fury* oraz *bitterness*. Wyraz *wickedness*, którego w [30] w ogóle nie ma, nie jest odwracalny, pierwszym zaś przekładem odwracalnym jest *spite*, i ten właśnie wyraz zostanie wybrany jako ekwiwalent rosyjskiego *злость*. Jednakże pierwszy odwracalny ekwiwalent będzie odrzucony, jeżeli słownik rosyjsko-obcojęzyczny podaje zakaz użycia go w tym kontekście, w jakim słowa wyjściowego użyto w badanym tekście, lub też jeżeli zaleca jego stosowanie tylko przy użyciu wyrazu wyjściowego w kontekstach, które się różnią od kontekstu

¹² W § 6 [w niniejszym przekładzie paragraf ten został pominięty] na zasadzie eksperymentu wypróbowaliśmy inną — „mnożściową” — metodykę przekładu, według której dla wyrazu wyjściowego wybieraliśmy nie jeden, lecz wszystkie jego przekłady odwracalne. Ponieważ dla automatycznego wykrywania (ale nie w celu potwierdzenia wykrytych intuicyjnie) powtórzeń metodyka ta okazała się mniej efektywna, jako zasadniczą traktujemy w niniejszej pracy pierwszą z nich — unarną.

badanego tekstu¹³. Wówczas wybieramy najbliższy z pozostałych przekładów odwracalnych. W wypadku, jeżeli wyjściowy wyraz rosyjski słownik tłumaczy przez parę słów, jak np. *слетаться* — ‘zlatywać się’ — *fly together*, wybieramy dla niego — oprócz wypadków, kiedy jednym z członów tej pary jest czasownik posiłkowy — ekwiwalent złożony z tych dwu słów. W razie homonimii wyrazu wyjściowego, kiedy to w słowniku rosyjsko-obcojęzycznym odpowiadają mu dwa lub więcej haseł (np. wyraz *править*, ‘rządzić, kierować, powozić; poprawiać, ostrzyć, prostować’) i kiedy hasła te nie zawierają żadnych wskazówek na temat kontekstu, który pozwoliłby uniknąć homonimii w danym tekście, wtedy dla wyrazu wyjściowego wybieramy dwa lub więcej ekwiwalentów.

2. 2. 4. 2. Etap analizy semantycznej. Jeżeli wyraz wyjściowy nie wchodzi do słownika rosyjsko-obcojęzycznego (np. *вопявшие* ‘wołające o pomstę do nieba’ w [29]), lub też nie ma ani jednego przekładu odwracalnego (takim jest np. w [29—30] wyraz *гады*, ‘gadzi-ny’), to w dalszej analizie wyraz ten nie jest już uwzględniany. Wszystkie pozostałe słowa podlegają analizie semantycznej za pomocą SI tego języka, na który dokonano tłumaczenia. Dla każdego przetłumaczonego ekwiwalentu wyjściowego wyrazu tekstu odnajdujemy jego obraz semantyczny — szereg numerów lub nazw tych klas semantycznych, do których ekwiwalent ten jest zaliczany przez dany SI. I ten właśnie szereg uważany jest dalej za „obraz semantyczny” wyrazu wyjściowego, będącego praobrazem danego ekwiwalentu w analizowanym tekście¹⁴. Jeżeli zaś SI jest zbudowany hierarchicznie (tj. stanowi typ tezaurusu) i przez

¹³ Uwagi słownikowe, zalecające któryś z wariantów tłumaczenia przy użyciu wyjściowego wyrazu rosyjskiego właśnie w tym otoczeniu (połączeniu), w jakim występuje on w danym tekście, braliśmy pod uwagę wówczas jedynie, jeżeli wariant zalecany był odwracalny.

¹⁴ Takie nieskomplikowane przejście od semantycznego obrazu ekwiwalentu do semantycznego obrazu słowa wyjściowego możliwe jest tylko w wypadku unarnej metodyki tłumaczenia, ale nawet przy tej metodyce przejście to nie sprawdza się przy homonimach, ponieważ dla nich wybieramy po kilka ekwiwalentów, a wspólny obraz semantyczny słowa wyjściowego musi się kształtować z obrazów ekwiwalentów. W wypadku homonimów całkiem uzasadniony byłby wybór albo zbioru, albo przecięcia się obrazów wszystkich ekwiwalentów. Obie te propozycje mają także i niewątpliwe wady. W tekstach, które rozpatrujemy w niniejszym artykule, takich homonimów nie ma, dlatego nie ma tu sposobności do eksperymentalnego porównania obu sposobów analizy homonimów, jednakże pewne wnioski podpowiadają badania analogicznych sytuacji na etapie analizy semantycznej po zastosowaniu mnogościowej metodyki tłumaczenia [autor rozważa ten problem w § 6, pominiętym w naszym przekładzie], kiedy to omawiana sytuacja występuje już nie tylko w zakresie homonimii, ale też w zakresie wszystkich słów, które mają więcej niż jeden przekład odwracalny.

to przypisuje słowom semantyczne kategorie różnych poziomów, wówczas trzeba odnotować najpierw poziom, na którym będzie się dokonywało takiej analizy. Tak np. w [27] hasła (*heads*), w które grupuje się słowa, są z kolei łączone w podrozdziały i podpodrozdziały¹⁵, podrozdziały łączą się w rozdziały (*sections*), rozdziały, zaś w podklasy (*divisions*) lub bezpośrednio w klasy (*classes*). Analizy można dokonywać na każdym z tych poziomów, dbając tylko o to, by nie myliły się różne poziomy. Jednakże im wyższy jest poziom, tym mniejsza jest zaświadczana przezeń znaczeniowa bliskość obejmowanych wyrazów, dlatego poziomy wyższe od poziomu haseł wykorzystywaliśmy nie do całkowitej analizy wszystkich słów, lecz tylko do dodatkowego sprawdzenia takich par wyrazów, które nie poddawały się powiązaniu na poziomie haseł, ale intuicyjnie wydawały się powiązane, jak również do klasyfikacji mocy powtórzeń, wyodrębnionych na poziomie haseł. Tak więc rezultaty analizy na poziomie haseł wykorzystujemy przy automatycznym wykrywaniu powtórzeń, natomiast analizę na wyższych poziomach — jedynie jako środek potwierdzający powtórzenia wyróżnione intuicyjnie. Niewykluczone, że bardziej systematyczne wykorzystanie rozmaitych poziomów *Thesaurusa* ujawni nowe perspektywy analizy struktury tekstu i pozwoli w przyszłości na podjęcie problematyki związanej z hierarchią poziomów redukcji [свертывания] tekstu.

Konkretna procedura analizy za pomocą [27] wygląda następująco: dla wyrazu wyjściowego wybieramy, korzystając z alfabetycznego Indeks do [27], szereg numerów (lub nazw)¹⁶ tych haseł semantycznych, do których wchodzi w *Thesaurusie* obcojęzyczny ekwiwalent tego wyrazu. Tak np. *spite* wchodzi do hasła 709 „*Dissension*”, 735 „*Severity*”, 881 „*Enmity*”, 888 „*Hatred*”, 891 „*Resentment, Anger*”, 898 „*Malevolence*”, 910 „*Revenge*”, 912 „*Envy*”, 926 „*Detraction*”. A zatem dla wyrazu rosyjskiego *злoбa*, z którym na pierwszym etapie łączyliśmy ekwiwalent *spite*, wybieramy cały szereg tych haseł¹⁷. Jeżeli wyjściowy

¹⁵ Tak nazywamy poziomy rozczłonkowania, pośrednie pomiędzy *head* a *section*, podane w tabeli kategorii [27, s. XXXIX—XLVIII] kursywą lub prostymi dużymi literami z oznaczeniami 1°, 2° itp., ale nie mające odrębnej nazwy.

¹⁶ Indeks przy numerach haseł podaje nazwy podhaseł (*key words*), dlatego nazwy haseł wybieramy z zasadniczego tekstu *Thesaurusa*. Wyodrębnianie powtórzeń na poziomie samych *key words* wydaje się niezbyt skuteczne.

¹⁷ W wypadku, jeżeli w *Indeksie Thesaurusa* przy ekwiwalencie słowa wyjściowego figuruje jeszcze odsyłacz do wyrazu (angielskiego) o tym samym rdzeniu, ale należącego do innej części mowy, wówczas wszystkie hasła, do których wchodzi ten drugi wyraz, również włączamy do obrazu semantycznego słowa wyjściowego. Takie „pogłębianie” analizy przewidują *Reguły postępowania się Thesaurusem*: „Czasownik *deteriorate* ma tylko cztery odsyłacze, rzeczownik *deterioration* — czternaście. Oznacza to, że *deteriorate* ma cztery odsyłacze do haseł, w których *deterioration*

wyraz rosyjski posiada ekwiwalent złożony, będący parą słów (*слетаться* — *fly together*), to wybieramy dla niego cały zespół szeregów tych haseł semantycznych, do których wchodzi w *Thesaurusie* każdy składowy człon przekładu (z wyjątkiem takich wypadków, kiedy jednym z członów jest czasownik posiłkowy typu *be, get, itp.*, którego obraz semantyczny po prostu opuszczamy).

2. 2. 4. 3. Etap wykrywania powtórzeń. Rezultaty przekładu i analizy semantycznej układamy w tabelę¹⁸: w jej pierwszym dziale podajemy wyraz wyjściowy, w drugim — ekwiwalent obcojęzyczny tego wyrazu, a w trzecim — obraz semantyczny odnaleziony według danego SI. Tabela ta stanowi podstawę etapu końcowego, na którym wykrywa się powtórzenia semantyczne na drodze porównywania uzyskanych obrazów słów wyjściowych. Kryterium podobieństwa znaczeniowego, które tu stosowaliśmy, było minimalne: jeżeli obrazy dwóch wyrazów zawierają przynajmniej jeden wspólny numer (nazwę) hasła, to słowa takie uznajemy za semantycznie powiązane, a mianowicie — za człony powtórzenia tego hasła, do którego wchodzi oba dane wyrazy. Wyżej widzieliśmy np., że słowu *злоба* odpowiada według [27] szereg haseł 709, 735, 881, 888, 891, 898, 910, 912, 926. W sposób analogiczny dla słowa *враг* 'wróg' z jego ekwiwalentem *enemy* odnajdujemy szereg 663, 705, 881, 888. Przecięcie się tych szeregów, zawierające hasła 881 „*Enmity*” oraz 888 „*Hatred*” nie jest puste, i dlatego wyrazy *злоба* *враг* są członami powtórzenia semantycznego.

Na poziomie haseł [27] kryterium takie, jak zobaczymy niżej, sprawdza się całkiem dobrze. Dlatego występowanie w miejscach przecinania się obrazów semantycznych więcej niż jednego wspólnego elementu można wykorzystać nie tylko w celu stwierdzenia obecności powiązania znaczeniowego, ale również w celu określenia mniejszej lub większej mocy tego powiązania¹⁹.

nie występuje, i że pozostałe czternaście haseł pokrywają się z sobą” [27, s. XLIX]. W niektórych wypadkach ekwiwalenty w ogóle nie występują w [27] w postaci należącej do tej części mowy, w jakiej zostały uzyskane na etapie tłumaczenia (są to np. wszystkie przysłowki na *-ly*). W takich wypadkach braliśmy hasła odpowiadające formie o tym samym rdzeniu tej części mowy, od której bezpośrednio została utworzona forma nieobecna; tak np. zamiast form przysłówek na *-ly* braliśmy hasła przypisane do form przymiotnikowych, a uzyskiwanych poprzez odrzucenie *-ly*: zamiast *valiantly* — *valiant*. [...]

¹⁸ [Dodatek, zawierający odpowiednie tabele dla obu rozpatrywanych tekstów, w przekładzie został pominięty.]

¹⁹ Gdybyśmy dokonywali automatycznego wykrywania powtórzeń za pomocą wyższych poziomów *Thesaurusa*, wówczas — w związku z mniejszą mocą diagnostyczną wchodzenia do znacznie większych zespołów semantycznych — musielibyśmy zaproponować o wiele mocniejsze i bardziej wyważone kryterium występowania związków znaczeniowych między słowami. Na przykład: ponieważ wyższych kate-

2. 2. 4. 4. Przyjęty zapis struktury tekstu. Do zapisu struktury tekstu proponowano wykorzystać zarówno znane w zwykłej składni grafy (ale już nie drzewa) zależności pomiędzy subtekstami [31—32; 33, §§ 8—9; 34] i grafy składowych bezpośrednich [35, s. 61], jak i w specjalny sposób zbudowane schematy Siewbo [3, rozdz. III]. Każdy z tych typów schematów posiada pewne zalety i braki, każdy z nich odzwierciedla własny aspekt struktury tekstu. Schemat zależności stanowi najbardziej bezpośrednie odzwierciedlenie wyników analizy, łatwo też go zbudować w sposób automatyczny według wykazu wykrytych powiązań. Jednakże wówczas schemat zależności odznacza się sztywną zależnością od wybranego jednego rozczłonkowania tekstu na subteksty: jeżeli zbudujemy np. schemat zależności między maksyntemami, to istnienie w tekście bardziej złożonych subtekstów, powstających wskutek zespolenia kilku maksyntem (np. dzięki pewnym „przenikającym” je powtórzeniom lub specjalnym wiązaniom), odbije się w takim schemacie w najlepszym razie pośrednio (w postaci nierównomiernego rozkładu połączeń), ale jako samodzielne elementy tekstu subteksty te w schemacie zależności nie wystąpią. Obecność tych bardziej złożonych subtekstów można wprawdzie odzwierciedlić za pomocą innych schematów zależności między subtekstami „wyższego poziomu”, wówczas jednak każdy kolejny poziom będzie wymagał nowego schematu, a ponadto ilość tych poziomów (i schematów) w różnych częściach tekstu musiałaby być różna.

Rodzi się więc konieczność odzwierciedlenia hierarchicznego charakteru struktury tekstu w jakimś wspólnym schemacie. Za schemat taki może posłużyć graf składowych bezpośrednich (SB), który odpowiadałby lepiej również i takiej — tradycyjnej przy operowaniu tekstem w rozmaitych dziedzinach wiedzy i praktyki społecznej — procedurze, jak układanie planu tekstu, zawierające w sobie stopniowe rozbicie tekstu na nieprzecinające się części. Ale tutaj napotykamy dobrze znany m. in. w literaturoznawstwie problem niejedyności takiego rozbicia, i jeżeli terminalne, elementarne składowe — maksyntemy — narzucone nam są przez przyjęte wyżej założenia, to określenie granic bardziej złożonych składowych nie jest, jak dalej zobaczymy, sprawą zbyt łatwą.

Pewne braki cechują także schematy Siewbo (dalej będziemy je nazywali dynamicznymi). Jak i schematy zależności, nie wykazują one wyraźnie hierarchicznej struktury tekstu. Układanie schematów dyna-

gorii semantycznych (klas) jest w *Thesaurisie* tylko sześć, to prawdopodobieństwo, że dla każdego z jakichś dwu słów zupełnie przypadkowo odpowiada jedno hasło jakiejś wspólnej klasy, będzie całkiem duże. Dlatego też, aby stwierdzić związek semantyczny tych dwu słów na poziomie klas, trzeba będzie uwzględnić ponadto i ilość haseł trafiających do tej wspólnej klasy, i stosunek tej liczby do ogólnej liczby haseł odpowiadających tym dwom słowom.

micznych dla tekstów, zawierających ilość zdań jednoczłonowych lub niepełnych dwuczłonowych, utrudnia fakt, że nie jest jeszcze rozwiązany problem „głównego członu” takich zdań. Schematy dynamiczne nie odzwierciedlają związków typu nawiązania [зацепления] [7, s. 120—121] — wszystkie łączniki pozostają poza tymi schematami. Ale pod względem odzwierciedlania powtórzeń schematy te są znacznie bogatsze, aniżeli grafy zależności z zakreślonymi łukami, ponieważ uwzględniają one nie tylko agens powtórzenia i numer łączonych subtekstów, ale również strukturalną lokalizację członów powtórzenia wewnątrz subtekstu. I wreszcie schematy dynamiczne najlepiej odzwierciedlają sam fakt jedności tekstu oraz dynamikę tekstową, rozwój, ponieważ tekst przedstawiają one nie tylko jako ciąg różnych obiektów (subtekstów), jak w grafach zależności, i nie jako zakończoną i statycznie podzieloną całość, jak w grafach SB, lecz jako szereg następujących po sobie stanów (w tym również wprowadzania i usuwania) jednych i tych samych obiektów, a mianowicie — agensów występujących w tym tekście powtórzeń²⁰, niezależnie od tego, czy są to leksemy, pola semantyczne czy też kategorie gramatyczne. Każdemu takiemu agensowi powtórzenia odpowiada jego „linia zachowania”, a całość tekstu opisuje się przez zbiór takich linii jako pewnego rodzaju system dynamiczny (zob. [36, rozdz. 2]), którego dyskretny czas mierzy się numerami zdań.

Gdyby uznać za celowe rozgraniczenie terminów „struktura tekstu” i „kompozycja tekstu”, to najprawdopodobniej właśnie schemat dynamiczny z jego następczym „rozkładem” i przemieszczaniem jednych i tych samych obiektów znaczeniowych według gotowych form syntaktycznych należałoby uznać za sposób przedstawiania kompozycji tekstu, natomiast grafy zależności oraz SB — za przedstawienie struktury tekstu, odpowiednio w przekroju poziomym i pionowym. W świetle tych rozważań ograniczenie się do schematów jakiegoś jednego rodzaju wydaje się niemożliwe i dlatego jest sens przy analizie tekstów budować wszystkie trzy schematy. Będziemy przestrzegali następujących założeń²¹.

W charakterze grafu zależności (niezorientowanych) wszędzie będziemy budowali „odpowiadający danemu tekstowi graf minimalny” [37, s. 20—21], tj. dwa subteksty (w naszym wypadku — maksyntemy), które są powiązane z sobą pewnym powtórzeniem, łączymy krawędzią grafu wtedy i tylko wtedy, jeżeli żaden z subtekstów pośrednich nie jest z nimi powiązany powtórzeniem tego samego agensu.

²⁰ Przez agens powtórzenia rozumiemy w niniejszej pracy pewną podstawę, na jakiej powtórzenie to zostało wyodrębnione, tj. taką wspólną część dwóch lub więcej elementów, która sprawia, że są one wszystkie członami jednego powtórzenia.

²¹ [Założenia dotyczące schematów dynamicznych, jak również zbudowany na tej podstawie schemat tekstu „B” (w oryginale rys. 5) w przekładzie opuszczono.]

Graf SB buduje się według wykazu związków znacznie trudniej, aniżeli graf zależności pomiędzy maksyntemami. Podstawowa reguła, której przestrzegaliśmy przy przechodzeniu od grafu zależności do grafu SB, polega na tym, że za sygnał styku dwóch subtekstów służy urwanie się łańcucha powtórzenia (np. *aaab...*) lub odwrotnie — ponowienie pewnego wcześniej odnotowanego, ale następnie urwanego powtórzenia (np. *aa/bab/a...*). W związku z dużą liczbą powtórzeń oraz niepokrywaniem się zakresu ich działania wyróżnienie składowych napotyka jednakże istotne trudności. Algorytmizacja takiego wyróżniania byłaby względnie łatwa, gdybyśmy kierowali się jakąś ustaloną regułą preferencji powtórzeń: można się umówić np. (jak w [38]), by za granicę kolejnej składowej uznać takie powtórzenie spośród jeszcze nie wykorzystanych, a rozpoczynających się od lewej granicy składowej, wyróżnionej na poprzednim etapie (warunek ten uniemożliwia wyróżnienie składowych przerywanych), które odznaczają się największym zasięgiem (tj. zwiększone o 1 liczbę subtekstów oddzielających pierwszy i ostatni człon danego powtórzenia). Ponieważ jednak mieliśmy do czynienia nie z leksykalnymi, lecz dowolnymi powtórzeniami semantycznymi, uznaliśmy więc za niewłaściwe *a priori* uznać zasięg powtórzenia za jedyny czynnik istotny przy określaniu mocy powtórzenia i jego znaczenia w strukturze tekstu. Odpowiednio do tego również wyróżnienia składowych dokonywaliśmy na razie nie algorytmicznie, lecz w oparciu o merytoryczną ocenę względnego znaczenia rozmaitych powtórzeń. Jednocześnie dążyliśmy do całkowitej eksplicytności kryteriów owej oceny merytorycznej, z tym jednak, aby następnie w miarę narastania wyników analizy struktury różnych tekstów można było wysnuć wystarczająco pewne wnioski co do relacji rozmaitych czynników, które wywierają określony wpływ na rolę powtórzeń w strukturze tekstu oraz określają zasadę ich preferencji przy konstruowaniu schematu SB według listy powtórzeń. W charakterze takich czynników uwzględnialiśmy, prócz zasięgu powtórzenia, typ powtórzenia (na podstawie typologii, która po raz pierwszy została sformułowana w [7] i uściślona w [6, rozdz. 1; 8]); rangę powtórzenia, która równa się maksymalnej randze elementów tego powtórzenia jako członów swych zdań²²; ilość członów powtórzenia; stopień podobieństwa znaczeniowego między nimi oraz stopień charakterystyczności powtórze-

²² Zgodnie z nawiązującą do Jespersena i Tesnière'a klasyfikacją Dolininy [18, s. 32], do I rangi (poziomu, hierarchii) należy orzeczenie, do II rangi — człony zdania bezpośrednio zależne od orzeczenia (podmiot, dopełnienie ...), do III zaś rangi — człony zdania bezpośrednio zależne od członów zdania II rangi oraz wszystkie inne. Możliwe, że uwzględniając większą przeciętną długość konkretnych zdań, miałyby sens wprowadzenie większej ilości rang — uznając za człony zdania *n*-tej rangi wszystkie słowa, które bezpośrednio zależą od członów zdania (*n*-1)-tej rangi.

nia dla wybranego tekstu, tj. fakt, czy w tekście tym występują jakieś inne powtórzenia, których agensy są zbliżone (w naszym wypadku — zbliżone semantycznie) do agensu danego powtórzenia (o dwóch ostatnich cechach, w sposób istotny zależnych od struktury wykorzystywanego słownika, zob. dokładniej § 5.2.2.). Sam stopień zgodności ocen powtórzeń opartych na tych różnorodnych czynnikach, a związany z jednoznacznością-niejednoznacznością rozczłonkowania tekstu na SB stanowi istotną cechę jego struktury. Tak więc wysoki stopień zgodności wskazuje na obecność w tekście pewnej wyraźnie zmanifestowanej dominanty strukturalnej (w tym sensie szczególnie wymowny jest tekst Tiutczewa *Jest w morskich falach śpiewność zgodna...* zanalizowany w [6, rozdz. II, § 5.3], w którym dominuje powtórzenie hasła 410 „Мелодия: согласие” ponawiane w całym szeregu innych „muzycznych” haseł).

Przeprowadzone analizy pozwalają przypuszczać, że w warunkach jednorodności [однотипности] i mniej lub więcej jednakowej mocy konkurencyjnych powtórzeń semantycznych szczególne znaczenie zyskuje taki czynnik, jak ilość powtórzeń urywających się jednocześnie. Gdyby schemat SB był budowany nie na podstawie samej listy powtórzeń, lecz na podstawie odpowiadającego jej schematu zależności, wówczas algorytm, uwzględniający przede wszystkim ten właśnie czynnik, wymagałby każdorazowego wyznaczania pionu, który kolejne dwie SB rozgraniczałby w taki sposób, że przecinałby najmniejszą ilość krawędzi grafu zależności. Przypuszczalnie algorytm taki działałby lepiej niż algorytm uwzględniający tylko zasięg powtórzeń, jednakże już sam fakt, że uniemożliwia on wyodrębnianie przerywanych SB (w rodzaju wyróżnianych w § 5.2), podpowiada, iż wymaga on i poważnego sprawdzenia merytorycznego, i wprowadzenia innych jeszcze zasad. [...]

Przy oznaczaniu agensów powtórzenia w grafach zależności wykorzystaliśmy: dla oznaczenia powtórzenia synonimicznego — hasłowy człon szeregu synonimicznego; dla powtórzenia elementów tej klasy — nazwę semantycznej klasy SI. Powtórzenia antonimiczne, jak również powtórzenia będące superpozycją antonimicznych, oznaczyliśmy w postaci zwykłego zapisu przez łącznik członów tego powtórzenia. Natomiast oznaczenia wszystkich nieleksykalnych powtórzeń semantycznych ujmujemy w cudzysłowy.

3. Analizowane teksty oraz wyróżniane w ich strukturze związki leksykalno-semantyczne

3.1. Dlaczego analizujemy teksty właśnie poetyckie?

Mogłoby się wydawać, że automatyzacja i algorytmizacja analizy w wypadku tekstów poetyckich jest najmniej osiągalna. Nieleksykalne

powtórzenia semantyczne również nie stanowią wyróżnika organizacji tekstów poetyckich — nawet nieliczne istniejące dane porównawczo-stylistyczne wskazują, iż rola ich jest dostatecznie duża także w budowie tekstów naukowo-popularnych (zob. podliczenia ilości „związków nieformalnych” w [39], z uwzględnieniem uwagi sformułowanej w [37, s. 15]) i przynajmniej niektórych odmian prozy artystycznej (zob. przykład z Oleszy w [40, s. 27], który pozwala sądzić, że zasadnicza różnica między dwoma wyróżnionymi w [40] odmianami prozy polega nie na obecności lub nieobecności związków międzydaniowych, lecz na stosunku ilościowym leksykalnych i nieleksykalnych powtórzeń semantycznych; zob. ponadto w [41] podliczenia ilości wypadków rozwijania się tematu następnego wypowiedzenia nie bezpośrednio z tematu lub rematu wypowiedzeń poprzedzających, lecz jedynie „z tego samego pola semantycznego”). Z drugiej zaś strony, także w organizacji tekstów poetyckich dość często przeważają powtórzenia leksykalne: tak np. wybrane przez nas dwa teksty stanowią wyjątek w całości zbioru [42] pod tym względem, że ich spójność (definicję spójności zob. w [37, s. 15—17]) nie sposób stwierdzić na podstawie wyłącznie powtórzeń leksykalnych.

Względy, które skłoniły nas do wyboru właśnie tekstów poetyckich, były trojakiego rodzaju. Po pierwsze — kwestia wygody i technicznego zagwarantowania analizy: a) Ich wierszowany charakter pociąga za sobą duży stopień przejrzystości organizacji syntaktycznej, w nich łatwiej bowiem wyodrębnić takie jednostki syntaktyczne, które można rozpatrywać jako minimalne subteksty. b) Właściwe tekstom poetyckim zjawiska tzw. „aktualizacji” [43, s. 29] oraz „zwartości szeregu wierszowego” [44] zmuszają do uwzględniania wszystkich występujących w nich związków wraz ze związkami między elementami stosunkowo peryferyjnymi oraz pozwalają na przynajmniej tymczasowe pominięcie kwestii hierarchizowania tych związków. c) W tekstach naukowych poważna część związków (powtórzeń) zachodzi pomiędzy terminami specjalistycznymi, które praktycznie nie mają odpowiedniego odzwierciedlenia w ogólnych słownikach semantycznych i dlatego wymagają przy ich analizie wyspecjalizowanych słowników semantycznych obejmujących terminologię z określonych dziedzin. W tekstach poetyckich natomiast kompozycyjna rola „poetyzmów” (zjawiska, które w języku poetyckim stanowi odpowiednik terminologii), ich udział w powtórzeniach jest zazwyczaj mniej intensywny, i dlatego właśnie wydaje się, że ogólne słowniki semantyczne, uwzględniające ponadto znaczną część poetyzmów dzięki swemu tradycyjnemu ukierunkowaniu na literaturę piękną, okażą się wystarczająco efektywne dla analizy struktury tekstów poetyckich ²³.

²³ Do wyjątków będą zapewne należały teksty szkół poetyckich, które dążyły do możliwie największej umowności języka poetyckiego i odizolowania się od języka ogólnego.

Po drugie — kwestia trudności i lingwistycznych perspektyw takiej analizy. Jak wiadomo, teksty poetyckie odznaczają się wielką złożonością swej organizacji, implicytnym charakterem związków semantycznych, i na tej właśnie podstawie w stosunku do tekstów poetyckich mówi się o „zasadniczej niemożności” wypracowania uniwersalnych, a tym bardziej — sformalizowanych — zabiegów analitycznych (zob. np. [45, s. 111, punkt 5]). Nie kwestionując potrzeby indywidualnego podejścia do każdego tekstu, chcielibyśmy jednakże wykazać, że uniwersalne zabiegi analityczne także nie są bezowocne. Wolno zatem przypuszczać, że jeżeli zabiegi takie dadzą jakieś pozytywne rezultaty przy analizie tak precyzyjnych i indywidualnych konstrukcji tekstowych, to tym bardziej mogą się one okazać przydatne w analizie do tekstów o większej standaryzacji w innych stylach lub odmianach mowy. Co więcej: znany pogląd Tomaszewskiego na temat podobieństwa praw kompozycyjnych czystej liryki i wywodu logicznego [46] pod względem ich wspólnego przeciwstawienia strukturom narracyjnym pozwala żywić nadzieję, iż badania struktury tekstów lirycznych mogą być użyteczne bezpośrednio przy badaniach przynajmniej niektórych odmian tekstów naukowych, w szczególności zaś filozoficznych.

I wreszcie kierowaliśmy się względami ciągłości tradycji naukowej: pierwszą próbę semantycznej analizy struktury tekstu przeprowadzić należało właśnie na materiale, na którym po raz pierwszy rozwinął analizę Żyrmunski drogą wykrywania powtórzeń leksykalnych [2] i dla którego Szengeli, skrytykowany przez Żyrmunskiego za „błąkanie się po powierzchni”, zaproponował najbliższy odpowiednik powtórzenia semantycznego — pojęcie „tematu” [47] jako takiego aspektu znaczenia słowa lub wyrażenia, który podlega wyakcentowaniu dzięki współodniesieniu zdania, w którego skład wchodzi dany wyraz lub wyrażenie, do innych zdań.

3. 2. Rozczłonkowanie wybranych tekstów na jednostki terminalne

Przedmiot naszej analizy stanowią dwa teksty Niekrasowa: *В столицах шум, гремят витии...* (tekst „A”) oraz *Смогли честные, доблестно павшие* (tekst „B”). Cytujemy je w całości (według [42]), zaznaczając rozczłonkowanie na maksyntemy: porządkowy numer maksyntemy oznaczamy rzymską cyfrą w nawiasie usytuowaną na początku odpowiedniej maksyntemy.

- „A”: (I) В столицах шум, (II) гремят витии,
 (III) Кипит словесная война,
 (IV) А там, во глубине России, —
 Там вековая тишина.

- (V) Лишь ветер не дает покою
Вершинам придорожных ив,
(VI) И выгибаются дугою,
Целуясь с матерью-землею,
Колосья бесконечных нив...
„B”: (I) Смолкли честные, доблестно павшие,
(II) Смолкли их голоса одинокие,
За несчастный народ вопиявшие,
(III) Но разнuzданы страсти жестокие.
(IV) Вихорь злобы и бешенства носится
Над тобою, страна безответная.
(V) Все живое, все доброе косится...
(VI) Слышно только, о ночь безрассветная,
Среди мрака тобою разлитого,
Как враги, торжествуя, скликаются,
Как на труп великана убитого
Кровожадные птицы слетаются,
Ядовитые гады сползаются!

3. 3. Leksykalne i syntaktyczne związki między daniowe

Nim przystąpimy do analizy struktury tych tekstów za pomocą słowników semantycznych, zobaczymy najpierw, jakie w nich występują leksykalne i syntaktyczne środki nawiązania. Oba teksty cechuje całkowity lub prawie całkowity brak powtórzeń leksykalnych: w „A” nie ma ich w ogóle²⁴, w „B” tylko maksyntemy I i II są połączone powtórzeniem leksykalnym *смолкли-смолкли* oraz odpowiadającym powtórzeniu nawiązaniem leksykalnym *честные-их*. Spójki (nawiązania syntaktyczne) odgrywają poważną rolę jedynie w tekście „A”, w którym wszystkie trzy spójniki — *а, лишь, и* — wskutek swego rozkładu mogłyby zapewnić spójność całego ciągu 6 maksyntem nawet wówczas, gdyby nie zachodziły inne powiązania. W tekście „B” występuje zaledwie jedna spójka syntaktyczna *но*, którą łączy III z zespolonymi dzięki powtórzeniu leksykalnemu I i II, jednakże z następnym tekstem czterowiersz ten, odgraniczony jedynym w wierszu odstępem autorskim (o roli takiego odstepu jako specjalnego sygnału zaznaczającego styk części kompozy-

²⁴ A ściślej mówiąc, w tekście „A” mamy jedno powtórzenie leksykalne — *в столицах*, które jest członem okolicznikowym nie tylko maksyntemy I, ale i dwu następnych. Jednakże wykryć to powtórzenie można dopiero po przekształceniu tekstu, odwołując się do specjalnej „reguły uzupełnienia brakującego okolicznika”, której zastosowanie w dużym stopniu będzie zapewne zależało od tego, jak są rozgraniczone odpowiednie maksyntemy — kropką czy też przecinkiem (por. „W Londynie ogłoszono kwarantannę, krążą rozmaite pogłoski o jej przyczynach” oraz „W Londynie ogłoszono kwarantannę. Krążą rozmaite pogłoski o jej przyczynach”) oraz od tego, czy w tekście późniejszym występują jakieś inne „konkurencyjne” okoliczniki.

cyjnych w tekstach poetyckich zob. [48, 49]), nie jest powiązany ani powtórzeniem leksykalnym, ani spójkami syntaktycznymi.

Istotniejszą rolę odgrywają w tych tekstach powtórzenia gramatyczne. Wprawdzie wyraźnego powtórzenia całej konstrukcji syntaktycznej (syntaktycznego paralelizmu) tutaj nie ma: w „A” zarysowuje się paralelizm maksytem I i IV, który pozostaje jednak nie zrealizowany wskutek wtargnięcia do IV dwukrotnego *там* oraz zdecydowanej różnicy co do wielkości maksytem; w „B” paralelizm wskazuje związek dwóch wersów (12 i 13) w obrębie tej samej maksytemy. Powtórzenie czasów czasowników jest tu jednak dość wyraźne. W „A” są to formy osobowe czasu teraźniejszego orzeczenia czasownikowego w maksytemach II, III, V i VI oraz „implicytny” czas teraźniejszy w I i IV. W „B” maksytemy IV, V, VI powiązane są powtórzeniem znaczenia czasu teraźniejszego: *носятся — косится — слышно*. To samo powtórzenie łączy wszystkie trzy maksytemy z III (*разнузданы*) i przez to — całość pierwszego czterowiersza z całością pozostałych dziewięciu wersów. Maksytema III znajduje się na pograniczu obu tych subtekstów i dlatego w jej orzeczeniu *разнузданы* akcentuje się jego ambiwalencję temporalną: ponieważ został tu opuszczony czasownik posiłkowy, *разнузданы* można potraktować jako czas przeszły i teraźniejszy, choć ostatnie znaczenie wydaje się nam nieco wyraźniejsze²⁵. Powtórzenie czasu teraźniejszego jest dlatego tak wyraziste, że wszędzie występuje w tej samej pozycji funkcjonalnej jako czas orzeczenia. I odwrotnie: jedną z przyczyn stosunkowo niewielkiej roli strukturalnej powtórzenia czasu przeszłego (*perfectum*) *смогли — убитого* stanowi, obok oddzielającej dużej odległości od siebie, zróżnicowanie pozycji funkcjonalnych oraz rang członów tego powtórzenia w odpowiednich zdaniach (orzeczenie i przydawka).

Rozpatrując powtórzenia czasu czasownikowego, zaliczanego do „semantycznie pełnych” [50, s. 71] kategorii gramatycznych, tj. kategorii, które przenoszą informację o rzeczywistości pozajęzykowej i jednocześnie nie są narzucane przez sytuację pozajęzykową, wkroczyliśmy już na obszar powtórzeń oznaczanego, niekoniecznie sprzęgniętych z powtórzeniem odpowiedniego oznaczającego. Przejdźmy obecnie do ustalenia nieleksykalnych powtórzeń semantycznych.

4. Wyniki wykrywania powtórzeń semantycznych za pomocą słowników przedideograficznych

4. 1. T e k s t „A”

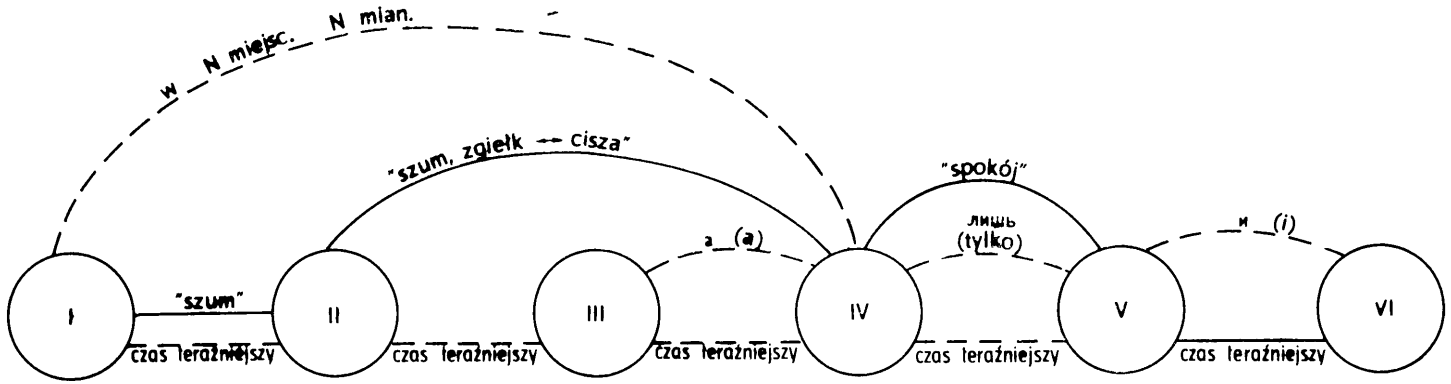
W tekście tym wyróżniamy: powtórzenie synonimiczne *тишина — покой* ‘cisza — spokój’ (szereg „*спокойствие*” ‘spokój’ według [21]);

²⁵ Zauważmy, że gdyby odstęp ten znalazł się nie po czwartym wersie, lecz przed nim, wówczas powtórzenie czasu teraźniejszego zespałoby zaledwie dziesięć wersów 4—13, potencjalna perfektywność wyrazu *разнузданы* byłaby odczuwana

antonimiczne powtórzenie шум — тишина 'szum, zgiełk — cisza' (według [23]) oraz złożone powtórzenia шум — греметь 'szum, zgiełk — grzmieć, rozbrzmiewać' będące superpozycją synonimicznego (według [21]) i rdzeniowego (według [20]) powtórzenia греметь = Syn (шуметь) = Syn (v (шум)) oraz тишина = Ant (гром 'grzmot') = Ant (N (греметь) gdzie v oznacza „derywat” czasownikowy, a N — „derywat” substancywny (terminy według [24, s. 201]). Te cztery powtórzenia faktycznie wyczerpują listę powtórzeń odczuwanych intuicyjnie, z wyjątkiem wprowadzie takiego istotnego powtórzenia antonimicznego jak столица — глубина 'stolica — głębokość'. Takie pełne wykrycie powtórzeń semantycznych na poziomie przedideograficznym pozwala nam na zbudowanie schematu struktury tekstu „A”, nie czekając na wyniki zastosowania słownika ideograficznego. Schemat zależności, uwzględniający powtórzenia semantyczne wykryte za pomocą słowników, a także wymienione poprzednio spójki i powtórzenia gramatyczne, został przedstawiony na rys. 1 (krawędzie, które odpowiadają powtórzeniom semantycznym, podane zostały linią ciągłą, natomiast odpowiadające związkom gramatycznym — linią przerywaną).

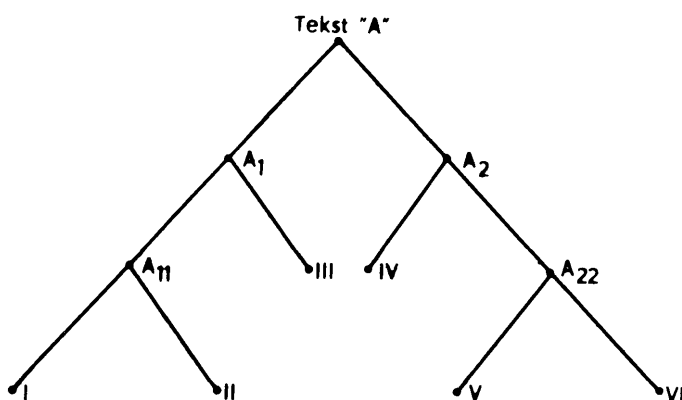
Przejdźmy obecnie do rozcłonkowania tekstu „A” na SB. Na pierwszy rzut oka granicę między SB całości tekstu należałoby przeprowadzić po IV, ponieważ tutaj kończy się powtórzenie шум — тишина odznaczające się największym zasięgiem spośród wykrytych powtórzeń semantycznych, a zaczynających się od maksyntemy I. Po tym powtórzeniu następuje jednakże powtórzenie synonimiczne тишина — покой i zgodnie z prawem naturalnym $y = \text{Syn}(\text{Ant}(x)) \rightarrow y = \text{Ant}(x)$ wspólne powtórzenie obejmuje faktycznie V, a przez to również całość tekstu, albowiem VI jest niewątpliwie zespolone z V w jeden subtekst dzięki spójce *u*²⁶. Klasa semantyczna, będąca agensem tego powtórzenia (umownie można go nazwać „cisza i jej zakłócenie”), może więc być wybrana, tak samo jak i powtórzenie czasu teraźniejszego, w charakterze „adnotacji” całości tekstu (szczegółowo zabieg ten przeprowadzamy na tekście „B” w § 5.2.5), ale nie może być zastosowana przy rozcłonkowaniu tekstu na SB. W wypadkach, kiedy powtórzenie antonimiczne obejmuje cały tekst, granicę SB tekstu najlepiej jest przeprowadzić nie znacznie słabiej, natomiast środkiem powiązania tych dziesięciu wersów z wersami 1—3 byłaby tylko spójka *но*.

²⁶ Zakres działania łączników międzyzdaniowych, charakter ich współoddziaływania i stosunek do powtórzeń zmuszeni jesteśmy określać wyłącznie intuicyjnie. Sformułowanie reguł konstruktywnych takiego określenia wydaje się najpilniejszym problemem tego kierunku badań jednostek ponadzdaniowych, który podąża od zdania do tekstu. Bez rozwiązania tego problemu rozpowszechnione obecnie pojęcie złożonej całości syntaktycznej (jednostki ponadzdaniowej) nadal pozbawione będzie sensu.



Rys. 1

po maksyntemie, w której po raz pierwszy występuje człon antonimiczny, lecz przed nią, czyli — w naszym wypadku — przed IV, co podtrzymuje zarówno spójka *a*, jak i zasygnalizowane już wyżej ukryte powtórzenie *в столицах* (zauważmy także, że obszar oddziaływania nieujawnionego powtórzenia antonimicznego *столицах — глубине* podtrzymanego przez wskazujące *там* pokrywa się z obszarem oddziaływania *шум — тишина*. Wyróżnione składowe samego tekstu $A_1 = \{I-III\}$ oraz $A_2 = \{IV-VI\}$ wydają się zbudowane identycznie: mają po trzy maksyntemy, a w każdej trójce dwie pierwsze maksyntemy są połączone powtórzeniem albo synonimiczno-rdzeniowym, albo synonimicznym. Mogłoby się więc wydawać, że rozgraniczyć je należy w oparciu o urwanie się tych powtórzeń także jednakowo: na I—II i III oraz na IV—V i VI. Dla A_1 rozkład taki rzeczywiście wydaje się najbardziej naturalny i znajduje nawet dodatkowe uzasadnienie w zjawisku „jedności szeregu wierszowego” [44, s. 94—108]: I i II ujęte w ramy jednego wersu są powiązane mocniej niż II i III (lub tym bardziej I i III) usytuowane w różnych wersach. Tymczasem w A_2 rozkład taki wydaje się niewłaściwy: dwie spójki — wewnętrzna *u* oraz zewnętrzna *лишь* — związek V z VI czynią bardziej zwartym od związku V z IV. Dla A_2 wybieramy zatem rozkład $A_2 = A_{21} \cup A_{22}$, gdzie $A_{21} = IV$ oraz $A_{22} = V-VI$, bez względu na fakt, że w tym celu musimy najpierw wyróżnić SB nie przylegającą do lewego boku swej bezpośredniej zawierającej²⁷. W wyniku drugiego etapu rozcłonkowania na SB uzyskujemy człon czteroelementowy $A = (A_{11} \cup A_{12}) \cup (A_{21} \cup A_{22}) \cdot A_{12} = III$ oraz $A_{21} = IV$ są pojedynczymi maksyntemami i już same przez się stanowią jednostki terminalne, natomiast A_{11} oraz A_{22} rozpadają się na jednostki terminalne w jedyny sposób. Ostateczny schemat SB przedstawia rys. 2.



Rys. 2

²⁷ Jednostka X jest bezpośrednią zawierającą (BZ) dla jednostek X_1, X_2, \dots, X_n wtedy i tylko wtedy, jeżeli X_1, \dots, X_n są bezpośrednimi składowymi X . [...]

Istotną i charakterystyczną cechą tego schematu hierarchicznej struktury tekstu „A” stanowi ścisła symetria lustrzana [52, wykład I]. Schemat SB, którego jednostkami terminalnymi są już nie maksyntemy, lecz wersy (otrzymujemy go ze schematu na rys. 2 łącząc wierzchołki I i II oraz wprowadzając rozgałęzienie wierzchołków IV i VI), właściwości tych już nie posiada: w rozkładzie jednostek syntaktycznych według wersów symetria zostaje zachwiana²⁸.

4. 2. T e k s t „B”

Na przedideograficznym poziomie analizy struktura tekstu „B” ujawnia się nieco słabiej: nie występują tu w ogóle powtórzenia rdzeniowe, za powtórzenie synonimiczne można by uznać chyba jedynie parę *нае шие — убитого* ‘polegli — zabitego’; za pomocą [21] nie sposób go stwierdzić z powodu braku wyrazu *наших* ‘poległy’, jednakże strata nie wydaje się zbyt wielka, albowiem rola tego powtórzenia w organizacji tekstu nie jest duża: a) z powodu wielkiej odległości dzielącej oba te człony w tekście oraz b) z powodu ich peryferyjnej pozycji w funkcjonalnej strukturze odpowiednich maksyntem (są bowiem członami zdania trzeciej rangi).

O wiele istotniejszą rolę odgrywają w tym tekście takie powtórzenia, które mogą być opisane jako powtórzenia elementów jednej klasy parametrycznej. Faktycznie, cały wiersz został oparty na wrażeniach słuchowych: w I i II mówi się, że ktoś zamilkł, tj. nie można go więcej usłyszeć; w IV mówi się o nieodpowiadaniu [безответственности] kraju, a zatem — kraj milczy, nie słyhać go. I w końcu w VI czytamy, że coś jednak słyhać. I właśnie parametry semantyczne pozwalają opisać tę konstrukcję opartą na „wrażeniach słuchowych”: *смокнуть* = Fin_1 (звук ‘dźwięk’) = Fin_2 (звучать ‘brzmieć’); *звук* ‘dźwięk’ = Gener (голос ‘głos’) = Gener (S_2 (слушать ‘słyszeć’)). Dlatego *смокли* — *голос-*

²⁸ Gdybyśmy rozczłonkowali A_1 nie na $A_{11} = \{I-II\}$ oraz $A_{12} = III$, lecz na $A'_{11} = \{I\}$ oraz $A'_{12} = \{II-III\}$, wówczas schemat SB, którego jednostkami terminalnymi są maksyntemy, odznaczałby się już nie lokalną, lecz przenośną symetrią [52, wykład 2] SB samego tekstu, tj. poddrzew o rdzeniu A'_1 i A'_2 . Przekształcenie tego schematu w schemat, którego jednostkami terminalnymi są wersy, wymagałoby już nie tylko rozgałęzienia i łączenia wierzchołków końcowych, ale także albo poprzedzającego łączenie przewieszania gałęzi II (a to w zasadzie odpowiadałoby przejściu do rozczłonkowania na A_{11} i A_{12}), albo też zakłócenia postaci drzewa uzyskiwanego grafu z powodu podporządkowania jednostki terminalnej 1 (pierwszego wersu) naraz dwóm wierzchołkom poprzedzającego poziomu A'_{11} i A'_{12} . Co więcej, wszystkie wymienione tu przekształcenia można opisać jako kombinację minimalnych przekształceń, całkiem analogicznych do zaproponowanych dla drzewa zależności w [53, s. 44—47].

— *слышно* stanowi powtórzenie semantyczne, jest właśnie powtórzeniem elementów klasy parametrycznej, którą umownie określić można jako „dźwięk (i jego brak)”²⁹.

Znacznie trudniej wykazać, że elementem tego powtórzenia jest również wyraz *безответная* ‘nieodpowiadający (kraj)’, ponieważ nieobecność odpowiedzi nie musi być powiązana z brakiem właśnie dźwięku. Jednakże, jeżeli zastosujemy zawężony wywód logiczny X — *безответный* — X , m. in. nie mówi w odpowiedzi, to wówczas otrzymamy wyraz *говорить* ‘mówić’, który już należy do parametrycznej klasy „dźwięk”. Już sama długość przeprowadzonej operacji świadczy o stosunkowo niskim stopniu pokrewieństwa znaczeniowego wyrazu *безответная* z innymi członami tego powtórzenia. Udział *безответная* w ogólnej doniosłości powtórzenia zmniejsza się także przez to, iż wyraz ten w najlepszym razie należy do trzeciej rangi członów zdania, gdy tymczasem *смогли* i *слышно* — do pierwszej, a *голоса* — do drugiej.

Drugim powtórzeniem parametrycznym jest tu para *народ — страна* ‘naród — kraj’: *страна* = S_{loc} (*народ*), tj. kraj — to typowe „miejsce zamieszkania” narodu. Jednakże funkcjonalna waga tego powtórzenia jest stosunkowo niewielka, oba wyrazy, będąc członami zdania trzeciej rangi, zajmują bowiem w swych maksyntemach pozycję raczej peryferyjną.

Parametry nie wyczerpują możliwych powiązań semantycznych między wyrazami i obejmują tylko najbardziej regularne. Tak np. w rozpatrywanym tekście intuicyjnie dostrzegamy pewne pokrewieństwo semantyczne wyrazów *жестокий* ‘okrutny’, *разнузданы*, ‘rozpasane’, *злоба* ‘złość’, *бешенство* ‘wściekłość’, *враги* ‘wrogowie’, *кровожадные* ‘krwiożercze’, *ядовитые* ‘jadowite’, pokrewieństwo, które nie podlega żadnym usystematyzowanym relacjom typu parametrów. Zaproponowane w [24, s. 216—218] pojęcie „indywidualnej funkcji leksykalnej” tutaj się nie przydaje, ponieważ z samej swej istoty jest sprzeczne i stanowi jedynie

²⁹ Abstrahując od faktu, że w tym wypadku połączyliśmy w klasę słowa związane nie tylko stosunkami synonimicznymi i antonimicznymi, ale też dowolnymi stosunkami parametrycznymi, nie sposób nie zauważyć, że jest to właściwie ta sama klasa, którą wyżej zaproponowaliśmy dla tekstu „A”, określając ją jako „cisza i jej zakłócenie”. Bardzo wymowna jest jednakże różnica semantycznego wyakcentowywania w tych synonimicznych nazwach, przeciwległość ich „punktów wyjściowych”, ich „aktualnego rozczłonkowania”, przeciwległość, która stanowi odbicie przeciwstawnego rozwoju tych dwu tekstów w obrębie wspólnego pola semantycznego i odzwierciedlonej w nich hierarchii wartości poetyckich: punkt wyjściowy tekstu „A” stanowi otaczający poetę „szum, zgiełk”, któremu przeciwstawia się następnie kojąca ciszę; inaczej w tekście „B” — tutaj panuje przygniatająca cisza, będąca przeciwstawieniem poprzednich „głosów”, a podkreślana przez obecne odgłosy triumfu wroga.

uznanie naturalnego niepełnego stanu systemu, nie zaś środek przecięcia tego niepełnego stanu. Środkiem obiektywnego notowania relacji tego typu może się stać jedynie SI, który opisuje wielorakość związków indywidualnych uwzględnionych w nim wyrazów.

5. Wykrywanie powtórzeń semantycznych za pomocą „Thesaurusa” Rogeta. Budowanie schematów struktury tekstu „B”

5. 1. Tekst „A”

Jakkolwiek strukturę tego tekstu przeanalizowaliśmy już dostatecznie, niemniej dokonaliśmy także wykrywania powtórzeń za pomocą [27] w tym celu, aby sprawdzić uprzednio obcojęzyczny SI na materiale, który został już przebadany za pomocą słowników rosyjskich³⁰.

Na poziomie haseł [27] udaje się odnotować następujące powtórzenia: *столицях — вершинам* ‘stolicach — szczytom’ (wspólne hasła 34 „*Superiority*. Starszeństwo, przewaga”; 213 „*Summit*. Szczyt, góra (wierzch)”); *шум — гремят* ‘szum — rozbrzmiewają, grzmią’ (400 „*Loudness*. Głośność”; 528 „*Publication*. Ogłoszenie, publikacja”); *кипит — вершинам* ‘wrze — szczytom’ (355 „*Bubble. Cloud: air and water mixed*”³¹. Bańki Obłok: przemieszanie powietrza i wody”); *кипит — дугою* ‘wrze — łukiem’ (379 „*Heat*. Gorąco”); *глубине — бесконечных* ‘głębokości — bezkresnych’ (32 „*Greatness*. Wielkość; w przeciwieństwie do „małości” — S. G.)”); *тишина — покою* ‘cisza — spokoju’ (266 „*Quiescence*. Spokój, bezruch”; 399 „*Silence*. Cisza”; 677 „*Inaction*. Bezczywność”); *ветер — вершина* ‘wiatr — szczytom’ (315 „*Rotation*. Krążenie”).

Jak widać, oba powtórzenia synonimiczne *шум — гремят* oraz *тишина — покою* zostały wyjawione również na poziomie haseł *Thesaurusa*, przy czym, jeżeli bezpośrednio synonimy *тишина* i *покой* mają trzy wspólne hasła, to bardziej odległe *шум* i *греметь* (jak pamiętamy, ich powtórzenie stanowi superpozycję dwóch powtórzeń) mają w przekroju tylko jedno hasło wspólne. Taki zaś fakt, że na poziomie haseł nie wystąpiło antonimiczne powtórzenie *шум — тишина*, znaj-

³⁰ [Wyniki analizy semantycznej podano w *Dodatku* (w tabeli I), pominiętym w niniejszym przekładzie.]

³¹ Nazwy haseł [27] mają własną składnię: słowa podawane przez kropkę są równoprawne i tworzą charakterystykę nazywanej klasy tylko zespołowo, gdy tymczasem słowa następujące po dwukropku i pisane małą literą stanowią objaśnienie tych słów podstawowych, znajdujących się przed dwukropkiem. W *Planie klasyfikacji* i w *Synoptycznej tabeli kategorii* [27] te pomocnicze słowa na ogół są pomijane.

duje swe naturalne wytłumaczenie w ogólnej strukturze *Thesaurusu* Rogeta: antonimy do słów danego hasła zawsze się w nim przenosi do hasła następnego (tak np. hasło 266 „*Quiescence*. Spokój. Bezruch” następuje po 265 „*Motion: successive change of place*. Ruch: stopniowa zmiana miejsca”). Dlatego właśnie antonimy z zasady nie mogą być wykryte na poziomie hasel i dlatego w celu ich wykrycia należałoby wprowadzić specjalny poziom pośredni, którego semantyczne „A-klasy” powstają dzięki połączeniu dwóch kolejnych hasel antonimicznych (hasła zaś, które nie mają odpowiednich hasel antonimicznych, stanowią samodzielne A-klasy). Tak np. powtórzenie *шум — тишина* ujawnia się jako powtórzenie A-klasy i jest utworzone dzięki połączeniu hasel 398 „*Sound*. Dźwięk”, wchodzącej do obrazu semantycznego słowa *шум* oraz 399 „*Silence*. Cisza, milczenie”, wchodzącej do obrazu słowa *тишина*. Natomiast powtórzenie *столицах — глубине*, którego nie zdołaliśmy odnotować za pomocą [23], nie może być zanotowane także za pomocą A-klas, albowiem obrazy tych słów nie zawierają hasel antonimicznych (lecz zaledwie „zbliżone” 213 „*Summit*” oraz 211 „*Depth*”).

Spośród innych powtórzeń *ветер — вершина* pozostaje środkiem powiązania wersów w obrębie jednej maksyntemy, a *столицах — вершинам* oraz *глубине — бесконечных* nie odgrywają w organizacji tekstu szczególnej roli z powodu peryferyjnej pozycji ich członów (dość od siebie odległych) w strukturze maksyntem, a zwłaszcza z powodu różnic tych pozycji³². Za czysto szumowe, powstałe jako skutek przekładu i analizy, można, jak się wydaje, uznać jedynie powtórzenia *купит — вершина* oraz *купит — дугою*; a zatem szum wytwarzany przez naszą metodykę stanowi w tym wypadku $2/11 = 18\%$ (za jednostkę obliczeniową tu i dalej przyjmujemy parę słów połączonych wspólnym hasłem; parę słów, których obrazy semantyczne mają n wspólnych hasel, traktujemy jako n jednostek), czyli nie przekracza pułapu dopuszczalnego we współczesnych systemach automatycznego sporządzania indeksów i poszukiwania informacji.

Przekonując się, że zastosowanie [27] jest skuteczne i daje wyniki dość zgodne z wynikami uzyskanymi przy zastosowaniu słowników rosyjskich, wykorzystajmy ten słownik przy analizie tekstu „B”, dla którego słowniki rosyjskie okazały się prawie nieprzydatne.

³² Drugie z tych powtórzeń, którego członki mają jednakową liczbę sylab nieakcentowanych w części przedakcentowej i zajmują jednakowe miejsce w rytmicznej strukturze swych wersów, jest jednak bardziej widoczne od pierwszego i tworzy swoistą ramę maksyntem IV—VI, a to dodatkowo świadczy na korzyść naszego wyróżnienia tych maksyntem — zob. § 4.1 — jako składowej bezpośredniej całości tekstu.

5. 2. Tekst „B”

5. 2. 1. Wykryte powtórzenia, graf zależności oraz schemat dynamiczny. Analiza na poziomie haseł [27] ujawnia w „B” następujące powtórzenia semantyczne (zapisujemy od razu pełny szereg wszystkich słowoform połączonych powtórzeniem zaznaczonego w nawiasie hasła:

смогли — смогли — навшие ‘zamilkli — zamilkli — polegli’ (405 „*Non-resonance. Niebrzmienie*”); *смогли — смогли — одинокие* ‘zamilkli — zamilkli — samotni’ (883 „*Unsociability. Seclusion. Nietowarzystwość (nierozmowność). Odosobnienie*”); *смогли — смогли — сползаются* ‘zamilkli — zamilkli — spełzają się’ (525 „*Concealment. Ukrywanie się*”); *смогли — смогли* ‘zamilkli — zamilkli’ (145 „*Cessation: change from action to rest. Zaprzestanie: przejście od działania do spokoju*”; 161 „*Impotence. Bezsilność*”; 266 „*Quiescence. Spokój, bezruch*”; 399 „*Silence. Cisza, milczenie*”; 479 „*Confutation. Obalenie*”, 578 „*Aphony. Utrata głosu*”; 582 „*Taciturnity. Milkliwość (nierozmowność)*”; 334 „*Lubrication. Smar (smarowanie)*”; 924 „*Disapprobation. Dezaprobata*”);

честные — доброе ‘uczciwi — dobre’ (931 „*Disinterestedness. Bezinteresowność*”);

навшие — несчастный ‘polegli — nieszczęsny’ (731 „*Adversity. Klęska*”); *навшие — ночь — мрака* ‘polegli — noc — mroku’ (129 „*Evening. Autumn. Winter. Wieczór. Jesień. Zima*”);

голоса — скликаются ‘głosy — z(na)wołują się’ (528 „*Publication. Ogłoszenie*”);

одинокие — разнузданы ‘samotni — rozpasane’ (744 „*Freedom. Wolność*”); *несчастный — страсти* ‘nieszczęsny — namiętności’ (825 „*Suffering. Cierpienie*”); *несчастный — жестокие — ядовитые* ‘nieszczęsny — okrutny — jadowne’ (827 „*Painfulness. Umęczenie*”);

народ — страна ‘naród — kraj’ (192 „*Abode: place of habitation or resort. Mieszkanie: miejsce zamieszkania*”);

разнузданный — страсти — жестокие — бешенства — летаются ‘rozpasany — namiętności — okrutne — wściekłości — zlatują się’ (176 „*Violence. Przemoc*”);

страсти — бешенства ‘namiętności — wściekłości’ (822 „*Excitability. Podniecenie, zapalczywość*”); *страсти — бешенства — злости* ‘namiętności — wściekłości — złości’ (891 „*Resentment. Anger. Oburzenie. Gniew*”); *страсти — скликаются* ‘namiętności — z(na)wołują się’ (859 „*Desire. Pragnienie*”);

жестокие — злости ‘okrutne — złości’ (735 „*Severity. Okrucieństwo*”);

жестокие — злости — кровожадные — ядовитые ‘okrutne — złości —

krwiożercze — jadowite' (898 „*Malevolence*. Zła wola, złośliwość”); *жестокие — кровожадные — ядовитые* ‘okrutne — krwiożercze — jadowite’ (645 „*Badness*. Nieprzydatność. Szkodliwość”); *жестокие — кровожадные* ‘okrutne — krwiożercze’ (718 „*War*. Wojna”);

вихорь — слетаются ‘wicher — zlatują się’ (318 „*Agitation: irregular motion*. Niepokój: ruch chaotyczny”);

злобы — враги ‘złości — wrogowie’ (881 „*Enmity*. Wrogość”; 888 „*Hatred*. Nienawiść”);

бешенства — великана ‘wściekłości — olbrzymia’ (970 „*Fairy*. Baśniowość”, podhasło „*demon*”);

носятся — косится ‘unosi się — kosi się’ (258 „*Smoothness*. Gładkość”); *носятся — птицы — слетаются* ‘unosi się (lata) — ptaki — zlatują się’ (271 „*Aeronautics*. Aeronautyka”); *носятся — слетаются* ‘unosi się — zlatują się’ (274 „*Vehicle*. Wehikuł”);

живое — слетаются ‘żywe — zlatują się’ (498 „*Intelligence*. Wisdom. Inteligencja. Mądrość”);

слышно — скликаются ‘słyszać — z(na)wołują się’ (400 „*Loudness*. Głośność”);

ночь — мрака ‘noc — mroku’ (418 „*Darkness*. Ciemności”); *ночь — безрассветная* ‘noc — bez świtu’ (419 „*Dimness*. Przyćmienie”);

мрака — сползаются ‘mroku — spełzają się’ (523 „*Latency*. Stan ukryty”); *торжествуя — скликаются* ‘triumfując — z(na)wołują się’ (408 „*Human cry*. Wołanie człowieka (ludzkie)”);

скликаются — слетаются ‘z(na)wołują się — zlatują się’ (547 „*Indication*. Wskazywanie”; 38 „*Addition*. Zwiększanie (dodawanie”); 74 „*Assemblage*. Zbieranie się razem”; 89 „*Accompaniment*. Towarzyszenie”);

убитого — кровожадные ‘zabitego — krwiożercze’ (362 „*Killing: destruction of life*. Zabójstwo: niszczenie życia”);

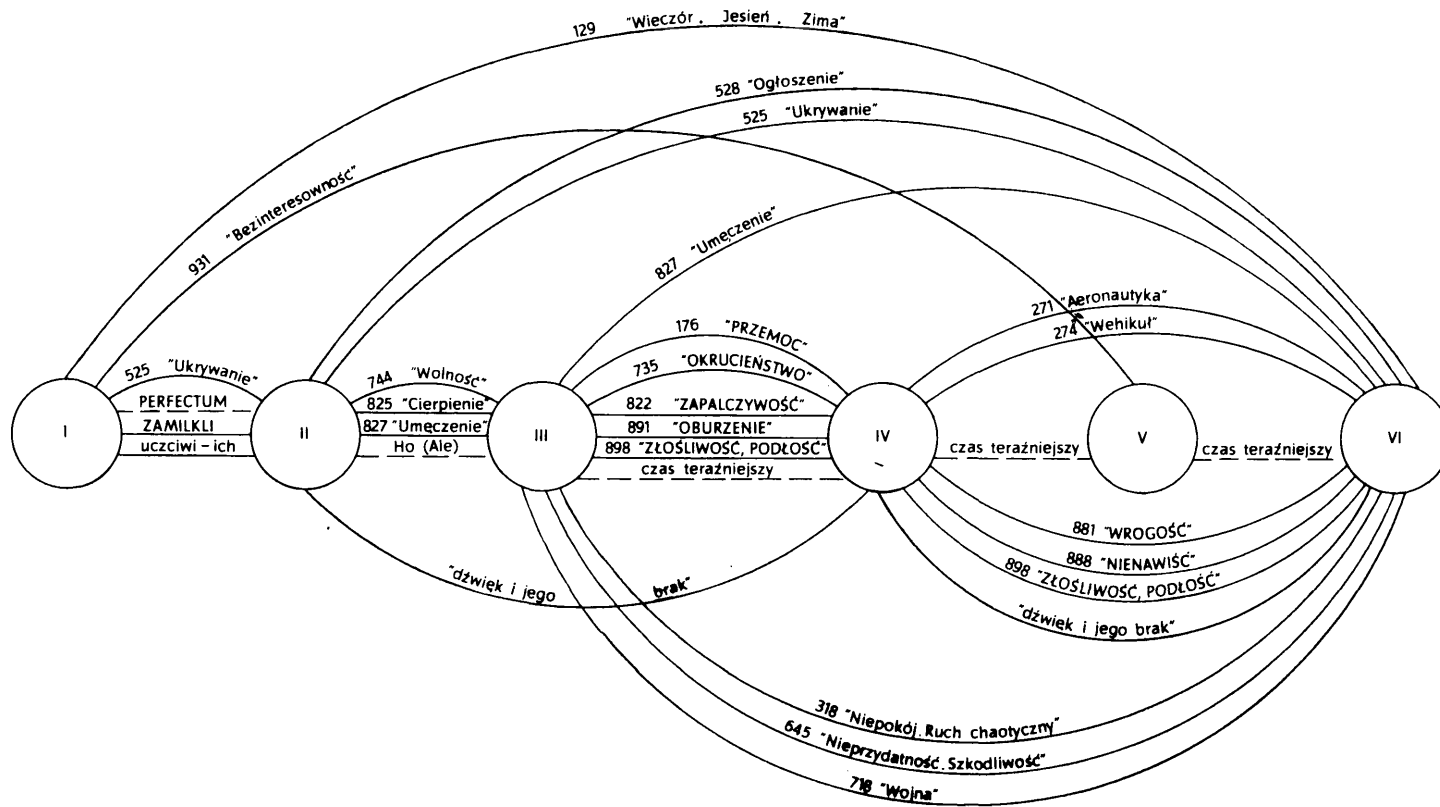
птицы — слетаются ‘ptaki — zlatują się’ (277 „*Velocity*. Szybkość”; 365 „*Animality*. Animal. Królestwo zwierzęce”).

Na poziomie haseł nie będą odnotowane następujące, oczywiste dla intuicji, powtórzenia: *смогли — голоса* ‘zamilkli — głosy’; *смогли — слышно* ‘zamilkli — słyszać’; *живое — убитого* ‘żywe — zabitego’; *навшие — живое* ‘polegli — żywe’; *навшие — убитого* ‘polegli — zabitego’; *несчастный — безответная* ‘nieszczęsny — milcząca (nieodpowiadająca)’; *скликаются — сползаются* ‘z(na)wołują się — spełzają się’ oraz *слетаются — сползаются* ‘zlatują się — spełzają się’. Co prawda, większość z nich wykryliśmy już za pomocą innych środków — słownika synonimów oraz słownika parametrycznego. Nic jednak nie stoi na przeszkodzie, aby wszystkie te powtórzenia wykryć również za pomocą [27]. Dotychczas korzystaliśmy bowiem wyłącznie z haseł, choć jak mówiliśmy w § 2.2.4.2, [27] pozwala także na analizę i na poziomach większych klas.

semantycznych. Nie podejmując systematycznej analizy tego rodzaju, wskaźmy tylko, że *смогли — голоса — слышно* stanowi powtórzenie elementów podrozdziału *Spoken Language*, który obejmuje hasła 577—585, natomiast *смогли — слышно — скликаются* stanowi powtórzenie elementów podrozdziału *Sound in General*, obejmującego hasła 399—409. *Павшие — живое — убитого* — to powtórzenie elementów podrozdziałów „*Vitality in General*” (hasła 358—363). I wreszcie *несчастный — безответная* rozpoznajemy jako powtórzenie semantyczne dzięki temu, iż jego człony należą do podrozdziału VI.II.1° *Passive Personal affections*. Słowa *жестокий* i *косится* mogą być zinterpretowane jako powtórzenie tylko na poziomie podrozdziałów poprzez podrozdział *Subservience*, który obejmuje aż 40 haseł 628—668 (wielka moc klasy świadczy o stosunkowo niskim stopniu wspólnoty semantycznej członów dwu ostatnich powtórzeń). Jeżeli natomiast idzie o powtórzenia afiksalnie-semantyczne *скликаются — сползаются, слетаются — сползаются*, to nie zostały one odnotowane wyłącznie z tego względu, iż [29], podając ekwiwalenty dla czasowników niedokonanych, sprowadza *сползать* do *спалзывать* i odpowiednio traci formę zwrotną *сползаться*, proponując jedyny przykład *creep down*, podczas gdy dla formy *сползаться* lepiej zaproponować *creep together*, a to od razu pozwoliłoby na wykrycie szeregu trójczłonowego *скликаются — слетаются — сползаются*, zespolonego przez powtórzenie trzech haseł semantycznych 38, 74 i 89 (zresztą, w języku angielskim powtórzenie to staje się leksykalne, występuje w postaci odrębnego wyrazu *together*).

Tak się przedstawia kwestia powtórzeń, które utraciliśmy podczas analizy na poziomie haseł. Zobaczmy obecnie, czy zaproponowana analiza nie wprowadziła „szumu”, tj. takich powtórzeń, w których członach intuicja nie wykazuje żadnej wspólnoty znaczeniowej. Jak się wydaje, do takich powtórzeń należą pary *страсти — скликаются, носится — косится, живое — слетаются, скликаются — слетаются*, będące powtórzeniem hasła 547 oraz *смогли — смогли* będące powtórzeniem hasła 334³³. Każdy z tych wypadków można wytłumaczyć konkretną przyczyną i z każdym z nich można walczyć (np. wprowadzając jako miarę tożsamości nie zwyczajne trafianie do jednego hasła, lecz ilość słów,

³³ Jako szumowe można rozpatrywać także powtórzenie hasła 744 *одинокие — разнузданы*. Jednakże jego wyodrębnienie może być usprawiedliwione przynajmniej historycznie: związek wyrazu *одинокий* ‘samotny’ z pojęciem „swobody” jest dość czywisty również współcześnie, a *разнузданы* ‘rozpasany’ zgodnie z formą wewnętrzną to *свободный от уздынузданы* ‘wolny od cugli’ lub ‘od pasa’. Mizerny udział tego powtórzenia w tekście wynika raczej z peryferyjnej pozycji wyrazu *одинокие* ‘samotni’ w swej maksyntemie, a także stąd, że *разнузданы*, trafiając na obszar bardzo mocnego pod wszelkimi względami powtórzenia hasła 176 „Przemoc”, jak gdyby ostatecznie wytraca w danym tekście swoją formę wewnętrzną.



Rys. 3

które znajdują się między tymi słowami w wykazie, będącym tekstem hasła, podzieloną przez ogólną ilość słów w hasle). Istotne jest jednak, że ilość takich pozornych powtórzeń nie jest wielka i stanowi albo 5/79, albo 4/64 (tj. w obu wypadkach $\approx 6,3^0\%$, co jest znacznie mniej od wielkości szumu przy analizie mniejszego tekstu „A”) w zależności od tego, czy uwzględniamy 15 par powstałych faktycznie w rezultacie „rozkładu semantycznego” powtórzenia leksykalnego *смолки*.

Wyróżnione na poziomie haseł powtórzenia nieszumowe łącznie z wykrytymi poprzednio powiązaniami syntaktycznymi przedstawia schemat zależności pomiędzy maksyntemami (zob. rys. 3)³⁴. W schemacie tym rzuca się w oczy, że niezależnie od pominięcia w nim znacznej liczby powtórzeń, które łączą nie maksyntemy, lecz wersy w obrębie maksyntem (w celu ich odzwierciedlenia należałoby zbudować graf zależności pomiędzy wersami), mimo to wszystko powiązań jest i tak znacznie więcej, niż tego wymaga zwyczajne stwierdzenie spójności lub nawet powszechnej spójności [37, s. 19—20] tekstu. Zewnętrznie niepretensjonalny, umiarkowany tekst Niekrasowa okazuje się dość mocno nasycony nawiązaniami znaczeniowymi — spośród wszystkich słów tekstu jedynie *доблестно*, *вопявшие* i *гады* nie mają „rymu” semantycznego (zresztą, dwa ostatnie z pewnością wykazałyby odpowiednie powiązania, gdyby nie wypadły z analizy wskutek niewystarczającego ich odzwierciedlenia w [29, 30]). Wielka ilość krawędzi sama przez się sprawia, iż nadzieje na projektywność struktury związków między subtekstami dla całości tekstu są bezpodstawne (por. tezę wysuniętą w [34] dla tekstów naukowo-technicznych). O projektywności można mówić w przypadku takich czy innych podstruktur, uzyskiwanych na drodze wyróżniania pewnych podzbiorów powiązań, np. powtórzeń haseł określonego rozdziału lub dwóch rozdziałów. Zakładając, że powtórzenie łączące subteksty akcentuje w nich jakiś wspólny temat (zob. [47], a także niżej § 5.2.5), wówczas przy wyodrębnianiu takich „podstruktur powiązań” można by pytać o projektywność-nieprojektywność względnej lokalizacji oraz łączliwości dwóch lub większej ilości tematów; o projektywność-

³⁴ Powtórzenie leksykalne występuje w schematach w postaci skrótu: zamiast 11 krawędzi powtórzeń semantycznych wszystkich elementów semantycznego wrazu *смолки*. Z uwagi na jego szczególną rolę w organizacji tekstu „B” spośród tych 11 powtórzeń zostało wyróżnione i oddane przez odrębną krawędź powtórzenie hasła 525. Na obu schematach przedstawione zostało także parametryczne powtórzenie „dźwięk i jego brak”, choć za pomocą [27] daje się ono wykryć nie na poziomie haseł, lecz na poziomie podrozdziałów. Przedstawienie krawędzi grafu z obu stron od prostej poziomej, która łączy wierzchołki, jest czysto umowne i wynika z dążenia do uproszczenia schematu, nie zaś z jakichś strukturalnych właściwości grafu.

-nieprojektywność poszczególnych odcinków ciągu tematycznego lub poszczególnych „jednostek ponadzdaniowych” (jak się wydaje, również w [34] diagramy powiązań stają się projektywne w dużym stopniu dlatego, że odniesione są nie do całych tekstów, lecz do niewielkich i stosunkowo zwartych fragmentów, a ponadto często odzwierciedlają nie wszystkie występujące w tych fragmentach „i-zależności”).

Inną ciekawą osobliwością tekstu „B”, dobrze uwidoczną w schemacie dynamicznym, stanowi to, że gdy przeważająca część powtórzeń semantycznych została zlokalizowana w grupie podmiotu, to najbardziej rozległe i szczególnie istotne dla struktury tekstu powtórzenie klasy „dźwięk i jego brak” (podobnie do powtórzeń gramatycznych) zlokalizowane jest przede wszystkim w grupie orzeczenia.

5. 2. 2. Czynniki, które mają wpływ na moc i znaczenie powtórzeń w strukturze tekstu. Jest rzeczą jasną, że znaczenie wyróżnionych powtórzeń semantycznych i ich rola w strukturze tekstu jest bardzo różna i zmienia się w zależności od szeregu czynników. W najogólniejszej postaci wszystkie te czynniki wymieniliśmy w § 2.4, tutaj zaś przytoczymy konkretne przykłady ich oddziaływania i uściślimy procedurę uwzględniania takich spośród nich, których pomiar poważnie zależy od zastosowanego słownika.

Ponieważ wszystkie wyróżnione powtórzenia zostały wyróżnione i opisane w jednakowy sposób jako powtórzenia haseł [27], dlatego są to powtórzenia jednego typu i dlatego nie ma żadnego sensu mówienie o oddziaływaniu na ich moc takiego czynnika jak typ powtórzenia (por. wyżej „przedideograficzną” analizę tekstu „A”). Jednostajność wyróżniania oraz wysoka liczebność możliwych agensów powtórzenia (tj. haseł [27]) pozwala natomiast dość łatwo zmierzyć i uwzględnić taki czynnik, jak stopień pokrewieństwa członów powtórzenia, ilość wspólnych zawartych w nich cech semantycznych. Za miarę parzystego podobieństwa znaczeniowego najlepiej obrać liczbę wspólnych haseł, do których wchodzi oba człony. Zastosowanie tej miary daje wyniki całkiem dobrze odpowiadające intuicji, por. np. zauważalność i doniosłość w tekście „B” powtórzeń *жестокое — бешенства* (jedno wspólne hasło), *жестокое — злобы* (dwa hasła) oraz *жестокое — кровожадные* (trzy wspólne hasła).

Zresztą na doniosłość tych trzech powtórzeń w danym wypadku ma wpływ również i taki wielokrotnie przez nas spotykany czynnik jak miejsce (ranga) członów powtórzenia w strukturze odpowiednich maksyntem oraz fakt, czy miejsca te są jednakowe. I ten właśnie czynnik decyduje o związku maksyntem I i V: powtórzenie *честные — доброе*, którego oba człony są podmiotem, jest bardziej zauważalne i bardziej ważne od powtórzenia *навише — живое*, jeżeli *навише* traktujemy jako

przydawkę przy *честные*, ale powtórzenia te natychmiast się wyrównują co do swej mocy, jeżeli *нашнее* potraktujemy jako podmiot równorzędny z *честные* (ten ostatni zabieg jest wprawdzie mało prawdopodobny, ale jednak możliwy dzięki syntaktycznej homonimii pierwszego wersu).

Doniosłość powtórzenia, które jednoczy zaledwie dwa człony, związana jest z zasięgiem tego powtórzenia odwrotną zależnością. Ale jeżeli zwiększenie zasięgu spowodowane jest zwiększeniem liczby członów powtórzenia, wówczas zależność staje się prosta (por. szczególną zauważalność powtórzeń według jednej i tej samej cechy, które tworzą długie szeregi — w tekście „B” są to powtórzenia haseł 176 „Przemoc”, 898 „Zła wola, złośliwość”, a także powtórzenia klasy parametrycznej „dźwięk i jego brak”).

Wreszcie, przy wartościowaniu powtórzenia uwzględnić należy również, czy powtarzana w jego członach cecha (agens powtórzenia) jest pojedyncza czy też powtarzają się w tekście i jakieś inne cechy pokrewne (ocenę opartą na tym czynniku nazwaliśmy w § 2.4 stopniem charakterystyczności powtórzenia). Pomiar stopnia charakterystyczności następuje z trudności największe, jednakże w naszym wypadku kwestię tę ułatwia hierarchiczna struktura [27] — połączenie haseł w podrozdziały, a następnie w rozdziały, podklasy i klasy całkiem naturalnie pozwala na wprowadzenie do licznych haseł miary ich pokrewieństwa: hasła należące do jednego podrozdziału są znacznie bliższe od haseł należących do różnych podrozdziałów tego samego rozdziału, te ostatnie zaś są z kolei bliższe od haseł należących do różnych rozdziałów jednej podklasy itd. Dlatego jeżeli z jakiegoś zbioru (np. klasy) występuje w tekście w charakterze agensu powtórzenia cały szereg haseł, wówczas każde z tych powtórzeń będzie dla tekstu bardziej charakterystyczne niż powtórzenie hasła stanowiącego jedyny wykorzystany w tekście agens powtórzenia z jakiejś klasy. Tak np. w tekście „B” 22 pary słów zostały połączone powtórzeniami 10 haseł z klasy VI „*Affections*. Uczucia” (zauważmy jednak, że nie ma powtórzeń ani jednego z rozdziałów „Moralne” i „Religijne uczucia”), 18 par — powtórzeniami 7 haseł z klasy I „*Abstract Relations*. Relacje abstrakcyjne”, podczas gdy z I podklasy „*Formation of Ideas*. Formowanie się idei” klasy IV „*Intellect*. Intelpekt” powtarzają się jedynie 2 rubryki, które łączą tylko dwie pary, a to może być formalnym dowodem uprzedniego wniosku, że jedyna nietautologiczna z tych par *живое — слетаются* stanowi powtórzenie pozorne, jest rezultatem „szumu”.

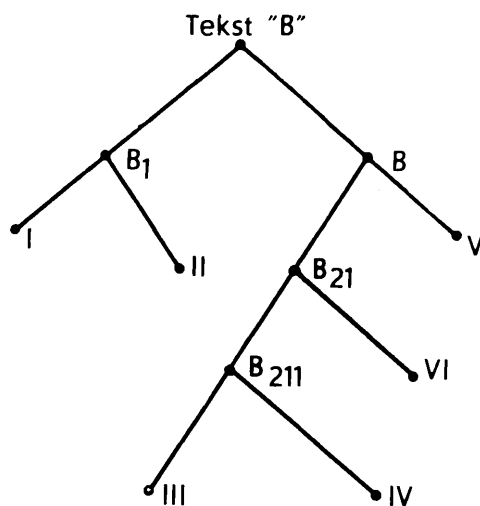
5. 2. 3. Schemat SB i ujawniające się w nim osobliwości struktury tekstu „B”. Możliwość uporządkowa-

nia powtórzeń według stopnia pokrewieństwa, charakterystyczności, ilości członów, zasięgu oraz rangi jest bardzo istotna podczas budowania schematu składowych bezpośrednich, albowiem wielka liczebność powtórzeń semantycznych utrudnia i czyni niejednoznaczny wyróżnianie w tekście „B” subtekstów większych od maksyntemy. Jeżeli rozpoczynając od lewej granicy tekstu, przystąpimy do rozczłonkowania tekstu na SB w oparciu o powiązania semantyczne maksyntemy I, okaże się wówczas niezwłocznie, że nawet dwie składowe pierwszego poziomu, tj. SB samego tekstu, trudno jest wyróżnić jednoznacznie. Największy zasięg mają powtórzenia hasła 129 i 525, które łączą I i VI. Opierając się na tych powtórzeniach będziemy zmuszeni do rozczłonkowania tekstu na $B'_1 = \{I, VI\}$ oraz $B'_2 = \{II-V\}$. Takiemu rozczłonkowaniu przeczy jednakże niski stopień pokrewieństwa oraz znikoma ilość członów w tych powtórzeniach; niska ranga oraz niewielka charakterystyczność pierwszego z nich; sformułowana w teorii analizy syntaktycznej według SB przez Nidę i Wellsa zasada preferencji (przy wielu innych podobnych warunkach) nieprzerwanych składowych [54, s. 88] oraz wyróżniające się dużym stopniem charakterystyczności związki VI z III i IV (ostatni czynnik jest najważniejszy; zauważmy, że jego uwzględnienie wymaga rozpatrywania powtórzeń, które się zaczynają nie od lewej granicy tekstu). Pozostają dwa możliwe rozwiązania: kierując się powtórzeniem hasła 931 *честные — доброе*, możemy rozczłonkować tekst na $B''_1 = \{I-II, V\}$ oraz $B''_2 = \{III-IV, VI\}$, a opierając się na urwaniu się powtórzenia leksykalnego *смокли — смокли* (będącego koniunkcją powtórzeń od razu II haseł) — na $B_1 = \{I-II\}$ oraz $B_2 = \{III-VI\}$. Uwzględniając wielką liczebność urywających się na II powtórzeń i ich charakterystyczność, jak również zbiegające się z nimi urwanie gramatycznego powtórzenia czasu czasownikowego³⁵ oraz wspomnianą już zasadę unikania przerywanych składowych, przyjmujemy drugie z tych dwu rozczłonkowań, należy jednak pamiętać, że pierwsze rozczłonkowanie również posiada pewne uzasadnienie — zarówno formalne (albowiem i w tym wypadku jedna z granic między SB samego tekstu przebiega między II i III, tj. przecina najmniejszą ilość krawędzi grafu zależności), jak i merytoryczne (w pierwszym wariancie rozczłonkowania w odrębną SB wchodzi odcinki dotyczące „losu uczciwych”).

Decydując się na rozczłonkowanie tekstu „B” na B_1 oraz B_2 , przejdźmy z kolei do rozczłonkowania składowych B_1 i B_2 . Złożona zaledwie z dwóch maksyntem B_1 dzieli się na te maksyntemy w jedyny sposób. Bardziej skomplikowana jest kwestia podziału drugiej składowej — B_2 .

³⁵ Gasparow np. [1] upatruje w urywaniu się powtórzeń gramatycznych jedyne sensowne pod względem teoretycznym kryterium segmentacji tekstu na jednostki ponadzdaniowe.

Ponieważ powtórzenie *носятся — косится* uznaliśmy za pozorne, to maksyntemę V z maksyntemą I łączy jedno tylko (na poziomie hasła) powtórzenie semantyczne (odzwierciedla to powrót do tematu „losu uczniowych”). Na tej maksyntemie urywa się powtórzenie hasła 898 „Zła wola, złośliwość”, ponowione w kolejnej maksyntemie. W związku z takim wypośrodkowanym i odizolowanym usytuowaniem V zmuszeni jesteśmy do rozcłonkowania B_2 na składowe przerywane $B_{21} = \{III, IV, VI\}$ oraz $B_{22} = \{V\}$. Ale jeżeli uwzględnimy powstałe na znacznie wyższym poziomie powtórzenie *живое — убитого*, natychmiast powstanie pytanie, czy VI odnieść do tej samej składowej co i III—IV (powtórzenie *жестокое — злобы — кровожадные — ядовитые*), czy też połączyć z V (powtórzenie *живое — убитого*), traktując B'_{21} i B'_{22} jako nieprzerwane. Mimo wszystko, jak się wydaje, powinno się wybrać pierwsze rozcłonkowanie, mając na uwadze fakt, że powtórzenie hasła 898 obejmuje więcej maksyntem, że jego człony zajmują, przeciętnie, wyższe miejsce w strukturze swych maksyntem, że stopień jego charakterystyczności jest wyższy niż powtórzenia *живое — убитого*. Na trzecim etapie B_{21} w sposób całkiem naturalny rozpada się na B_{211} , obejmującą maksyntemy III—IV, oraz B_{212} , obejmującą maksyntemę VI. I wreszcie na ostatnim, czwartym etapie rozcłonkowania B_{211} w jedyny sposób rozpada się na składające na nią maksyntemy. Uzyskany schemat końcowy³⁶ SB przedstawiamy na rys. 4.



Rys. 4

³⁶ Gdybyśmy w charakterze jednostek terminalnych struktury tekstu rozpatrywali nie maksyntemy, lecz wersy, wówczas niższe piętra schematu SB wyglądałyby nieco inaczej. Ponieważ jednak w tekście „B” (w odróżnieniu np. od tekstu „A”)

Charakterystyczne i bezpośrednio zauważalne cechy tego schematu świadczą o skomplikowanej i dysharmonicznej kompozycji struktury danego wiersza i polegają na zdecydowanej asymetrii (widocznej szczególnie wyraźnie na tle ściśle symetrycznego, zrównoważonego schematu SB tekstu „A”), regresywności (rozgałęzienie na lewo) prawej części tego schematu a odpowiadającej wersom 4—13³⁷ oraz na jego „nieprojektywności”³⁸ — w tym ostatnim wypadku chodzi o fakt, że maksyntema VI, która w tekście występuje po (tj. z prawa) maksyntemie V, tutaj wskutek przerwania B_{21} znalazła się po lewej stronie V.

5. 2. 4. O konflikcie między rozczłonkowaniem kompozycyjno-znaczeniowym a stroficznym. Obok wysokiej liczebności powiązań semantycznych oraz wynikającego stąd niejedynego rozczłonkowania kompozycyjnego, a także asymetrii, regresywności i „nieprojektywności”, charakterystyczną cechą tekstu „B’” stanowi konflikt między rozczłonkowaniem kompozycyjno-znaczeniowym a rozczłonkowaniem stroficznym. Nawet podstawowy strukturalno-kompozycyjny podział na B_1 i B_2 przebiega między wersem 3 i 4 — zakłóca więc spistość czterowiersza. Ale tutaj zwycięża jeszcze układ stroficzny — nawiązanie za pomocą spójnika *но* wystarczająco mocno przytrzymuje wers 4 i dlatego Niekrasow mógł sobie pozwolić na odstęp po utworzonym czterowierszu. Pod koniec drugiego czterowiersza zwycięża natomiast podział kompozycyjny: sens wersu 8 pozostaje niepełny i dlatego właśnie wers ten jest tak bardzo powiązany z następnymi wersami, autor musi więc zrezygnować z odstępu międzystroficznego i czterowiersze 2 i 3 połączyć w jedną całość. I wreszcie po wersie 12 inercja łańcucha rymowego sugeruje nam albo zakończenie wiersza, albo też początek kolejnego czterowiersza z jakimś nowym rymem. Tymczasem tekst podaje (znowu bez odstępu) jeszcze jedno zdanie podrzędne o tym samym rymie, które ostatecznie łamie organizację stro-

każda jednostka syntaktyczna zawiera pełną ilość wersów, to schemat doprowadzony do wersów może być uzyskany ze schematu na rys. 4 [w oryginale rys. 5] przez dalsze rozgałęzienia punktów końcowych i nie wymaga zatem ich łączenia lub przegrupowania.

³⁷ Rozgałęzienie w lewo, w odróżnieniu od rozgałęzienia w prawo, zwiększa tzw. głębokość drzewa SB, a przez to także pojemność pamięci operatywnej, którą powinno posiadać urządzenie zdolne do wygenerowania struktury (tekstu lub zdania), opisywanej przez dane drzewo SB (zob. [55, s. 206—208] z uwzględnieniem poprawki zaproponowanej w [56, s. 195]).

³⁸ Jak wiadomo, pojęcie projektywności zostało wypracowane dla drzewa zależności. Toteż nasze zastosowanie go wobec drzewa SB jest tylko metaforą, opartą nie tyle na ścisłym określeniu projektywności (zob. np. [56, s. 146—151, 189—191]), co raczej na pierwotnym zaproponowanym przez Leserfa [57] czysto geometrycznym współodniesieniu linearnego porządku elementów (w naszym wypadku — SB terminalnych) w tekście i w ortogonalnej projekcji drzewa.

ficzną i sprawia, że ogólna liczba wersów pozostaje nieparzysta³⁹, ale jednocześnie wprowadza powtórzenie *смогли — сползаются*, które — tworząc swoistą „ramę” tekstu — wyakcentowuje temat ciszy, przyniatającego milczenia: jakże powinno być cicho, skoro słyhać, jak się ktoś czołga, skrada.

5. 2. 5. *Quasi*-streszczenie tekstu „B” na podstawie powtórzeń semantycznych. Jak widać, wyrażenie sensu zwierzęcości, okrucieństwa, dysharmonii współczesnych Niekrasowowi wydarzeń, czyli zadanie tematyczne określa właściwości struktury, kompozycyjnego układu tekstu. Na zakończenie spróbujemy wykorzystać wykryte przez nas właściwości organizacji w celu zrekonstruowania, przynajmniej częściowo, owego zadania tematycznego, czyli tego, co się określa terminem „idea” lub „temat” wiersza. Oznaczając każdy z subtekstów symbolem zespalającego go powtórzenia, uzyskamy coś w rodzaju adnotacji następujących po sobie subtekstów i przez to — adnotację tekstu.

Dla całości tekstu, który rozpatrujemy jako niepodzielną jednostkę, odnaleźć możemy (z pewnym zastrzeżeniem, gdyż pomija się tu maksynty III i V) tylko klasę „dźwięk i jego brak” (lub „wrażenia słuchowe”), albo też zespalającą maksynty I—II oraz VI hasło „ukrywać się” („chować się”, „kryć się” — czasowniki lepiej bowiem oddają angielskie „*concealment*”, niż rzeczowniki o tym samym rdzeniu), hasło, które odgrywa w tym wypadku rolę antonimu „dźwięku”, tj. „ciszy”.

Bardziej dokładną adnotację wiersza można uzyskać w wypadku sporządzenia osobnej adnotacji dla każdej z tych dwu podstawowych części tekstu. Wówczas dla pierwszych trzech wersów uzyskalibyśmy powtarzające się w nich elementy ‘do chwili obecnej, już’; ‘zamilknąć’; ‘uczciwy’; natomiast dla wersów 4—13 — elementy ‘obecnie’ oraz 898 ‘Zła wola, złośliwość’. I jeżeli spójkę syntaktyczną *но* potraktujemy w naszym zapisie jako znak antonimiczny, wówczas dla całości tekstu uzyskamy następującą postać:

‘już zamilkli uczciwi’ ‘Ale’ ‘teraz (obecnie) złośliwość (szyderstwo)’.
[...] ⁴⁰

7. Warunki, w jakich przeprowadzaliśmy naszą analizę, były bardzo niedogodne, albowiem słownik ideograficzny oraz rozpatrywane teksty należą do różnych języków. Nawet w takich jednak warunkach wyniki wykrywania powtórzeń za pomocą przyjętej metodyki okazały się dość adekwatne, a wyniki eksperymentów kontrolnych, w których

³⁹ Por. hipotezę Jakobsona [58], zgodnie z którą nieparzysta liczba wersów stanowi środek odzwierciedlenia dysharmonii przedmiotu wypowiedzi.

⁴⁰ [§ 6. Контрольный эксперимент: применение альтернативной переводной методики в niniejszym przekładzie został pominięty.]

zastosowaliśmy i inną metodykę analizy, i inny słownik (zob. przypisy 11 i 12), nie są z nimi sprzeczne. Dlatego, choć nie uważamy zaproponowanej tu metodyki ani za zakończoną, ani tym bardziej za wyłączną (co prawda, pewne jej perspektywy — uwzględnienie wewnętrznej budowy słowa, wykorzystanie różnych poziomów słownika hierarchicznego — są oczywiste już dzisiaj), wolno nam sądzić, iż zdołaliśmy udowodnić samą możliwość obiektywnej analizy struktury tekstu za pomocą słowników semantycznych.

Przełożył Jerzy Faryno

Bibliografia ⁴¹

- *1. Б. М. Гаспаров, *О некоторых лингвистических аспектах изучения структуры текста. Третья летняя школа по вторичным моделирующим системам. Тезисы*. Тарту 1968.
- *2. В. М. Жирмунский, *Композиция лирических стихотворений*. Петроград 1921.
- *3. И. П. Севбо, *Структура связного текста и автоматизация реферирования*. Москва 1969.
- *4. Z. S. Harris, *Discourse Analysis Reprints*. The Hague 1963.
- *5. Z. S. Harris, *Discourse Analysis*. „Language” (Baltimore) 1, 1952. Przedruk w: Z. S. Harris *Papers in Structural and Transformational Linguistics*. Dordrecht 1970.
- *6. С. И. Гиндин, *Внутренняя организация текста. Элементы теории и семантический анализ*. Москва 1972 (maszynopis pracy doktorskiej).
- *7. С. И. Гиндин, *Онтологическое единство текста в виду внутритекстовой организации*. „Машинный перевод и прикладная лингвистика” 14, 1971.
- *8. С. И. Гиндин, [гес.:] И. П. Севбо, *Структура связного текста и автоматизация реферирования*. „Вопросы Языкознания” 5, 1971.
- *9. С. И. Гиндин, *Семантический повтор как особый тип связей между частями текста*. Москва 1972.
- *9а. С. И. Гиндин, *Анализ структуры текста Н. А. Некрасова „Смолкли честные, доблестно павшие...” с помощью семантического словаря*. Ярославль 1973.
10. А. Я. Шайкевич, *Распределение слов в тексте и выделение семантических полей*. „Иностранные языки в школе” 2, 1963.
11. Ю. А. Шрейдер, *Машинный перевод на основе смыслового кодирования*. „Научно-техническая информация” 1, 1963.
12. С. И. Гиндин, *Опыт статистической реконструкции семантики поэтического диалекта по корпусу связных текстов*. W: *Автоматическая переработка текста методами прикладной лингвистики*. Материалы Всесоюзной конференции. Кишинев 1971.
13. Ю. А. Шрейдер, *О семантических аспектах теории информации*. W: *Теоретические проблемы информатики*. Москва 1968.
14. В. В. Морковкин, *Идеографические словари*. Москва 1970.
15. А. А. Леонтьев, *От редактора*. W: *Семантическая структура слова. Психолингвистические исследования*. Москва 1971.
16. А. А. Потехина, *Из записок по русской грамматике*. W: В. А. Звегинцев, *История языкознания XIX—XX веков в очерках и извлечениях. Часть I*. Москва 1964.

⁴¹ Numery prac dotyczących struktury całości tekstu zaznaczyliśmy gwiazdką *.

17. Г. С. Шур, *О типах лексических ассоциаций в языке*. W: *Семантическая структура слова. Психолингвистические исследования*. Москва 1971.
18. И. Б. Долинина, *Проблема представления синтаксической структуры в грамматике членов предложения*. „Проблемы моделирования языка” 3, 1969.
19. D. S. Worth, A. S. Kozak, D. V. Tomson, *Russian Derivational Dictionary*. New York 1970.
20. З. А. Потиха, *Школьный словообразовательный словарь*. Москва 1964.
21. З. Е. Александрова, *Словарь синонимов русского языка*. Москва 1968.
22. *Словарь синонимов русского языка в 2 томах*. Ленинград 1970—1971.
23. Л. А. Введенская, *Словарь антонимов русского языка*. Ростов-на-Дону 1971.
24. А. К. Жолковский, И. А. Мельчук, *О семантическом синтезе*. “Проблемы кибернетики” 19, 1967.
25. Ю. Д. Апресян, А. К. Жолковский, И. А. Мельчук, *О системе семантического синтеза. III. Образцы словарных статей*. „Научно-техническая информация”, серия 2, 11, 1968. [Przekład polski: J. D. Apresjan, I. A. Mielczuk, A. K. Żolkowski, *Próba objaśniająco-kombinatorycznego słownika języka rosyjskiego*. Przeł. J. Faugno. W zbiorze: *Semantyka i słownik*. Pod redakcją A. Wierzbickiej. Wrocław—Warszawa—Kraków—Gdańsk 1972.]
26. *10 словарных статей толково-комбинаторного словаря русского языка*. (*Борьба, Борьба, Кадр, Кадры, Надежда, Надеяться, Отчаяние, Сердце, Страх, Стыдиться*). Москва 1971.
27. *Roget's Thesaurus of English Words and Phrases*. New Edition, Completely Revised and Modernized by Robert A. Dutch. 3d ed. London 1965.
28. Ch. Maquet, *Dictionnaire Analogique*. 10^e tirage. Paris 1961.
29. *Русско-английский словарь*. Составили С. К. Боянус и В. К. Мюллер. 50 000 слов ... Изд. 2, испр. и доп. Москва 1932.
30. *Англо-русский словарь*. Составил В. К. Мюллер. Около 60 000 слов и выражений. Изд. 2, испр. и доп. Москва 1946.
- *31. Э. Ф. Скороходько, *Автоматическая компрессия текста на основе анализа его семантической структуры*. W: *Проблемы прикладной лингвистики*. Тезисы межвузовской конференции, Часть 2, Москва 1969.
- *32. Э. Ф. Скороходько, *Анализ семантической структуры текста в связи с автоматизацией индексирования и реферирования*. W: *Автоматическая переработка текста методами прикладной лингвистики*. Материалы всесоюзной конференции. Кишинев 1971.
- *33. Э. Ф. Скороходько, *Лінгвістичні основи автоматизації інформаційного пошуку*. Київ 1970.
- *34. В. Е. Берзон, *О свойстве проективности языка на уровне структуры связного текста*. „Научно-техническая информация”, серия 2, 12, 1971.
- *35. А. Е. Кибрик, *Об иерархической структуре смысла связного текста*. „Проблемы моделирования языка” 3, 1969.
36. У. Росс Эшби [W. R. Ashby], *Конструкция мозга*. Москва 1964.
- *37. С. И. Гиндин, *Связный текст: формальное определение и элементы типологии*. Москва 1971.
- *38. Г. В. Бондаренко, *Об одном способе членения связного текста и представления его структуры*. „Машинный перевод и прикладная лингвистика” 15, 1972.
- *39. М. И. Белза, *К вопросу о некоторых особенностях семантической структуры связных текстов*. W: *Семантические проблемы автоматизации информационного поиска*. Киев 1971
- *40. Н. Д. Арутюнова, *К проблеме связности прозаического текста*. W: *Памяти академика В. В. Виноградова*. Москва 1971.
- *41. Н. И. Серкова, *О некоторых вопросах функциональной перспективы предложения в терминах „сверхфразовых единств”*. „Вопросы языкознания” 3, 1967.

42. Н. А. Некрасов, *Лирика*. Москва 1968.
43. *Тезисы Пражского лингвистического кружка*. W: *Пражский лингвистический кружок. Сборник статей*. Москва 1967.
44. Ю. Н. Тынянов, *Проблема стихотворного языка. Статьи*. Москва 1965.
45. Ю. И. Левин, *К проблеме семантического анализа поэтического текста*. W: *Н. А. Некрасов и русская литература. Тезисы, доклады и сообщения межвузовской научной конференции*. Кострома 1971.
- *46. Б. В. Томашевский, *Теория литературы. Поэтика*. Изд. 5-е, исправленное, Ленинград 1930.
- *47. Г. А. Шенгели, *О лирической композиции*. W: *Проблемы поэтики*. Москва 1925.
- *48. Ю. Л. Фрейдин, *Интервал у Пушкина*. W: *Тезисы докладов 5-й научной студенческой конференции Новосибирского Университета. Экономика, история, филология*. Новосибирск 1967.
- *49. А. В. Прохоров, Ю. Л. Фрейдин, *Малоизученные особенности членения пушкинских стихов*. W: *Slavic Poetics*, The Hague 1973.
50. И. И. Ревзин, В. Ю. Розенцвейг, *Основы общего и машинного перевода*. Москва 1964.
51. Э. Хемп, *Словарь американской лингвистической терминологии*. Перевод и дополнения В. В. Иванов. Москва 1964.
52. Г. Вейль, *Симметрия*. Москва 1968.
53. Т. Д. Корельская, Е. В. Падучева, *О формальном аппарате синтаксических преобразований*. W: *Исследования по математической лингвистике, математической логике и информационным языкам*. Москва 1972.
54. Р. М. Венцкович, А. Я. Шайкевич, *История языкознания*, Вып. 6. Москва 1970.
55. Ю. Д. Апресян, *Идеи и методы современной структурной лингвистики*. (Краткий очерк). Москва 1966.
- *56. Ю. А. Шрейдер, *Равенство, сходство, порядок*. Москва 1971.
57. И. Лесерф, *Применение программы и модели конфликтной ситуации к автоматическому синтаксическому анализу*. „Научно-техническая информация” 10, 1963.
58. R. Jakobson, *Une microscopie de dernier "Spleen" dans les „Fleurs du Mal"*. „Tel Quel” 29, 1927.