

JADWIGA SAMBOR

Z PROBLEMÓW WSPÓŁCZESNEJ SEMANTYKI KWANTYTATYWNEJ

1. W ostatnim dziesięcioleciu (1980-1990) rozwinęły się bardzo intensywne badania w zakresie lingwistyki kwantytatywnej, co spowodowane zostało m. in. wzrastającym wykorzystywaniem komputerów w badaniach lingwistycznych, umożliwiającym opracowanie statystyczne dużych korpusów tekstów. Równocześnie badania te, w początkowych etapach przede wszystkim materiałowe, mają obecnie w coraz większym stopniu charakter poszukiwań teoretycznych. Nie jest przypadkiem, że badania statystyczno-językowe zaczynały się od tworzenia statystyki fonemów, liter, słowników frekwencyjnych itp.; równocześnie obserwowano też prawidłowości zachodzące między cechami liczbowymi jednostek języka (takimi np., jak: frekwencja a długość, frekwencja a stopień polisemii itp.), które odnajdywano w materiale empirycznym.

W ostatnich latach odkryto wiele bardzo interesujących prawidłowości statystycznych, dotyczących zarówno jednostek języka, jak i tekstu; celem najnowszych badań jest nie tylko opis tych prawidłowości za pomocą określonych modeli probabilistycznych, ale także – docelowo – odkrycie całego systemu praw statystycznych, decydujących o strukturze kwantytatywnej języka (por. Hammerl 1987 a, Köhler 1986).

2. Wszelkie dotychczasowe badania statystyczno-językowe dotyczyły niemal wyłącznie planu wyrażenia; najwięcej uwagi poświęcono statystyce fonemów oraz leksemów, mniej natomiast rozwinięta była kwantytatywna morfologia oraz składnia. Leksykologia kwantytatywna, ze względu na możliwość badania dużych zbiorów jednostek, zawsze była działem bardzo uprzywilejowanym w językoznawstwie statystycznym. Wystarczy tu wspomnieć sformułowane u zarania jego rozwoju znane prawa G. K. Zipfa, dotyczące zależności między częstością a rangą, długością, liczbą znaczeń, wiekiem i pochodzeniem wyrazów (Zipf 1949; omówienie tych praw w: Sambor 1972; 1988).

W ostatnich kilku latach obserwuje się wzrost zainteresowania badaczy strukturą kwantytatywną planu semantycznego leksyki. Jest to obiekt badań bardzo trudny dla wszelkich analiz ilościowych ze względu na niemożność wyodrębnienia ścisłego zakresu pojęcia "znaczenie leksemu". Jak wiadomo, w semantyce nie wypracowano dotąd dostatecznie ścisłych kryteriów wydzielenia znaczeń, a jedynym źródłem ujmującym w miarę systematycznie strukturę znaczeniową leksyki są słowniki jednojęzyczne, w których opisy planu semantycznego są formułowane często według indywidualnych intuicji leksykografów. Zdaniem niektórych językoznawców sytuacja ta z góry wyklucza sensowność wszelkich kwantytatywnych badań semantyczno-leksykalnych. Sądzę, że jest to stanowisko zbyt skrajne, okazuje się bowiem, że w obfitym materiale empirycznym pewne prawidłowości semantyczne ujawniają się z zadziwiającą regularnością m i m o indywidualnych różnic między słownikami, dotyczących opisów poszczególnych leksemów. Może to być efektem działania prawa wielkich liczb, dzięki któremu w dużej masie obserwacji ujawnia się dana prawidłowość jako czysta tendencja statystyczna.

Celem obecnej pracy jest przedstawienie na materiale empirycznym kilku odkrytych i badanych obecnie regularności statystyczno-semantycznych dotyczących leksyki. Impulsem dla tych poszukiwań stały się przede wszystkim prace G. Altmanna i jego współpracowników na uniwersytecie w Bochum, dotyczące problemu tzw. dywersyfikacji semantycznej (Altmann 1985, 1991; Köhler, Altmann 1991).

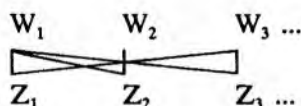
3. W najnowszych niemieckich pracach teoretycznych, analizujących strukturę kwantytatywną języka, postuluje się nie tylko badanie stanu tej struktury, ale także jej wyjaśnienie. Badacze odwołują się tu do znanej koncepcji G. K. Zipfa, który interpretował strukturę kwantytatywną języka na wszystkich jego poziomach (fonologicznym, morfologicznym, leksykalnym) jako efekt stosowania przez nadawców i odbiorców w procesie mówienia zasady najmniejszego wysiłku (por. Zipf 1949). Prawo najmniejszego wysiłku, ujmowane z punktu widzenia obu uczestników aktu mowy, sprowadza się do dwóch przeciwstawnych tendencji: nadawca, chcąc zminimalizować wysiłek nadawania, dąży do używania jak najmniejszej liczby znaków o możliwie prostej budowie, natomiast minimalizacja wysiłku odbiorcy przy dekodowaniu tekstu może być zapewniona przez używanie znacznej liczby znaków maksymalnie różniących się od siebie. Te dwie wzajemnie przeciwstawne tendencje prowadzą do tego, że w efekcie powstaje osobliwa, ale dobrze znana struktura kwantytatywna języka; np. w strukturze statystycznej tekstu istnieje zawsze nieliczna grupa jednostek o budowie prostej, mających bardzo wysokie częstości (efekt tendencji ze strony nadawcy do upraszczania czynności kodowania), a równocześnie występuje też – z małymi częstościami – znaczna liczba jednostek o budowie złożonej (efekt tendencji ze strony odbiorcy do zapewnienia maksymalnej wyrazistości odbioru).

Te dwie tendencje, sformułowane przez Zipfa (Zipf 1935, 1949), określa się w literaturze mianem sił Zipfa; sam autor tej koncepcji wyjaśniał poprzez zasadę najmniejszego wysiłku zaobserwowaną w tekstach strukturę statystyczną fonemów (częstsze są zawsze dźwięki artykulacyjnie prostsze).

Działające przeciwstawnie siły Zipfa prowadzą do procesów unifikacji i dywersyfikacji elementów języka. W najnowszych pracach poświęconych analizie semantycznej leksykonu traktuje się strukturę kwantytatywną znaczeń leksemów także jako efekt działania praw Zipfa. W jednej ze swych najbardziej znaczących prac tak pisze o tym G. Altmann (Altmann 1985): uwzględniając sytuacje skrajne, optymalnym rozwiązaniem dla odbiorcy byłoby używanie wyrazów (leksemów, morfemów, rdzeni) absolutnie jednoznacznych:

wyrazy:	W_1	W_2	$W_3 \dots$
znaczenia:	Z_1	Z_2	$Z_3 \dots$

(jest to niejako "stan narodzin" wyrazów), natomiast dla nadawcy minimum wysiłku nadawania zapewniłby 1 wyraz maksymalnie polisemiczny. Ten ostatni przykład jest wyłącznie modelem, gdyż taka sytuacja nigdy w języku nie zachodziła, natomiast wyrazy pierwotnie jednoznaczne poprzez częstsze używanie nabierają z czasem innych znaczeń, co w efekcie prowadzi do powstania polisemii typu:



Tendencją nadawcy jest powiększanie polisemii, tendencją odbiorcy – jej ograniczanie, wskutek czego w leksykonie istnieje bardzo wysoki udział leksemów jednoznacznych (o prostej budowie semantycznej), a równocześnie niski udział leksemów wysoce wieloznacznych (mających złożoną budowę semantyczną). Przykład powyższy dotyczy dywersyfikacji znaczeń leksemów w słowniku. Powróćmy do niego, rozważając poniżej rozkład leksemów w słowniku według liczby znaczeń, znany pod nazwą prawa Kryłowa.

4. W dalszej części artykułu ujmijemy pokrótce zarys problemów semantyki kwantytatywnej, jakie stały się przedmiotem badań w ostatnich latach, a następnie omówimy je kolejno, prezentując przede wszystkim wyniki empiryczne, uzyskane z odpowiednio dobranych, reprezentatywnych materiałów. Celem bowiem obecnej pracy jest unaocznienie omawianych zależności statystycznych poprzez dane empiryczne.

Pierwszy ważny problem dotyczy struktury statystycznej leksemów w słowniku, analizowanych ze względu na liczbę znaczeń, a więc wspomnianego wyżej prawa Kryłowa. Prawidłowość dotycząca udziału w słowniku leksemów mających x znaczeń była sformułowana już przez Zipfa, obecnie jest ona przedmiotem dużego zainteresowania, jakkolwiek badano ją dotąd na niewielkim materiale kilku języków.

Prawo Kryłowa to struktura semantyczno-statystyczna całego słownika, a więc makrosemantyka. Równocześnie dostrzeżono interesujący aspekt ilościowy na poziomie mikrosemantyki, tzn. na poziomie znaczeń konkretnego leksemu polisemicznego, którego poszczególne znaczenia są używane w tekście z różnymi częstościami, a zatem znaczenia te mają w tekście określony rozkład rangowy. Okazuje się, że i ta zależność – mimo niewątpliwych trudności w rozróżnianiu znaczeń wyrazu polisemicznego – ma w tekście przebieg bardzo regularny; w literaturze nadano jej nazwę prawa Beöthy.

Wreszcie ostatni zespół problemów dotyczy hierarchicznej struktury znaczeń leksemów w słowniku i wiąże się z badaniami tzw. ciągów definicyjnych typu *amoniak* → *gaz* → *substancja*, na temat których istnieje już spora literatura. Jak się okazało, budowa ciągów definicyjnych łączy się najściślej z tzw. gniazdami leksykalnymi, przy czym typ ich budowy wewnętrznej zdaje się wskazywać na istotne różnice między hasłami polisemicznymi a homonimicznymi.

Omówieniu powyższych problemów poświęcimy dalszą część pracy, ilustrując je za pomocą odpowiednich danych empirycznych, uzyskanych z dotychczasowych badań, i wskazując na perspektywy badawcze, jakie się w tej dziedzinie otwierają.

5. Rozkład wyrazów liczby znaczeń w słowniku, czyli prawo Kryłowa

Osobliwą strukturę statystyczną leksemów w słowniku, mających x znaczeń, odkrył Zipf (1949), stwierdzając, że istnieje zależność między liczbą znaczeń leksemu (x) a liczbą leksemów w słowniku (n), mających x znaczeń. Dane empiryczne wskazywały na zależność odwrotnie proporcjonalną: najwięcej jest w słowniku leksemów jednoznacznych, zaś w miarę wzrostu stopnia polisemiczności udział leksemów maleje. Zależność tę badał później węgierski językoznawca F. Papp na materiale 60 000 leksemów słownika języka węgierskiego, potwierdzając monotonicznie malejący kształt krzywej, ilustrującej tę prawidłowość (Papp 1967; omówienie wyników tej pracy w: Sambor 1972, 75-76).

Na początku lat 80-tych rozkładem wyrazów polisemicznych w słowniku zajął się J. K. Kryłow, badając ten rozkład w dwóch słownikach języka rosyjskiego. Jego dwie teoretyczne prace poświęcone tej prawidłowości wzbudziły duże zainteresowanie (Kryłow 1982, 1984). Materiałem badanym było słownictwo, wylosowane z całego słownika S. J. Ożegowa (Ożegow 1963, dalej jako SO) (próbki losowano dla każdej litery oddzielnie), a oprócz tego zbadano porównawczo

słownictwo na litery I, K, S z tegoż słownika. Drugą próbę stanowiły leksemy wylosowane z wielkiego 17-tomowego słownika języka rosyjskiego (dalej jako SSRLJa).

Uzyskane dane skłoniły autora do przyjęcia bardzo ciekawej hipotezy dotyczącej kolejnych liczebności leksemów mających x znaczeń, a mianowicie liczebności te miałyby się układać według postępu geometrycznego o ilorazie $\frac{1}{2}$, czyli pozostałyby w relacjach liczbowych $\frac{1}{2} : \frac{1}{4} : \frac{1}{8} : \frac{1}{16} \dots$. Zgodnie z tą hipotezą Kryłow przyjął modelową zależność między liczbą znaczeń x leksemu a prawdopodobieństwem p_x leksemów mających x znaczeń, wyrażającą się funkcją o postaci:

$$p_x = \frac{1}{2^x} \quad (x = 1, 2, \dots, k) \quad (1)$$

z której wynika, że 50% leksemów słownika stanowią leksemy jednoznaczne, 25% – leksemy mające dwa znaczenia, 12,5% – leksemy o trzech znaczeniach itd.

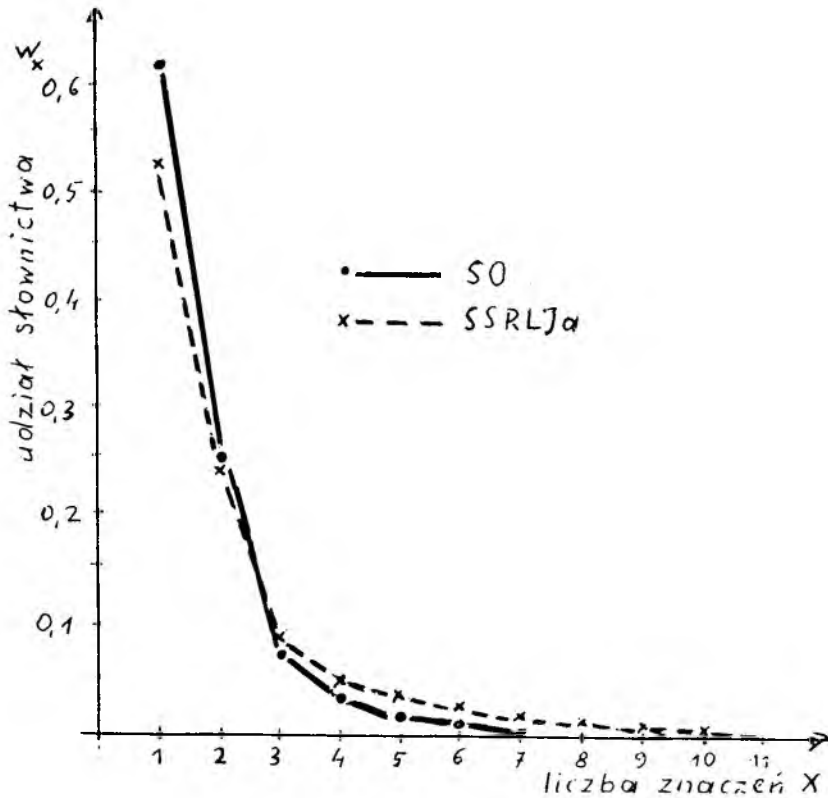
Poniżej w tabeli 1. podajemy dane empiryczne w_x , dotyczące udziałów słownictwa mającego x znaczeń w trzech próbach, pochodzących z obu słowników rosyjskich (S0 oraz SSRLJa), a także dane teoretyczne p_x , obliczone z funkcji (1) (Kryłow 1982, 251):

Tabela 1. Empiryczne i teoretyczne rozkłady leksemów według liczby znaczeń w dwóch słownikach języka rosyjskiego

liczba znaczeń x	Słownik			
	SO(A–Ja)	SO(IKS)	SSRLJa	P_x
	w_x	w_x	w_x	
1	0,615	0,614	0,525	0,500
2	0,254	0,249	0,248	0,250
3	0,071	0,070	0,084	0,125
4	0,030	0,031	0,045	0,063
5	0,013	0,012	0,029	0,031
6	0,007	0,005	0,018	0,016
7	0,003	0,004	0,012	0,008
8	0,002	0,003	0,009	0,004
9	0,002	0,000	0,008	0,002
10	0,002	0,004	0,004	0,001
11	0,001	0,003	0,002	0,0005

W tab. 1 uderza wyjątkowa zgodność danych dla całego słownika Ożęgowa i dla jego fragmentu, natomiast występują pewne niewielkie różnice w strukturze słownictwa polisemicznego, dotyczące obu badanych słowników, co uwidacznia

też rys. 1. Łatwo jednak zauważyć, że w obu wypadkach dane empiryczne układają się w ciąg malejący i są zbliżone do wartości teoretycznych, obliczonych z modelu (1).



Rys. 1. Rozkłady empiryczne leksemów według liczby znaczeń w dwóch słownikach języka rosyjskiego (Kryłow 1982, 251)

Prawo Kryłowa badano też na materiale języka polskiego i – porównawczo – rosyjskiego, uzyskanym z prób losowych, obejmujących 1000 leksemów, wylosowanych z *Małego słownika języka polskiego* (dalej jako MSJP) oraz ze słownika Ożegowa. Z MSJP losowano 3 próby po ca 1000 leksemów, obejmujące: a) wszystkie leksemy, b) rzeczowniki, c) czasowniki; ze słownika SO losowano 2 próby po 1000 leksemów, obejmujące wszystkie leksemy oraz rzeczowniki. Uzyskane rozkłady empiryczne zawiera tabela 2. (Sambor 1989 d), dane porównawcze przedstawiemy w postaci rysunków.

Ostatnio uzyskano też analogiczne dane dotyczące struktury semantycznej słownictwa w języku angielskim na podstawie próby 1000 leksemów ze słownika A. S. Hornby'a (*Oxford Advanced Learner's Dictionary of Current English*, Oxford

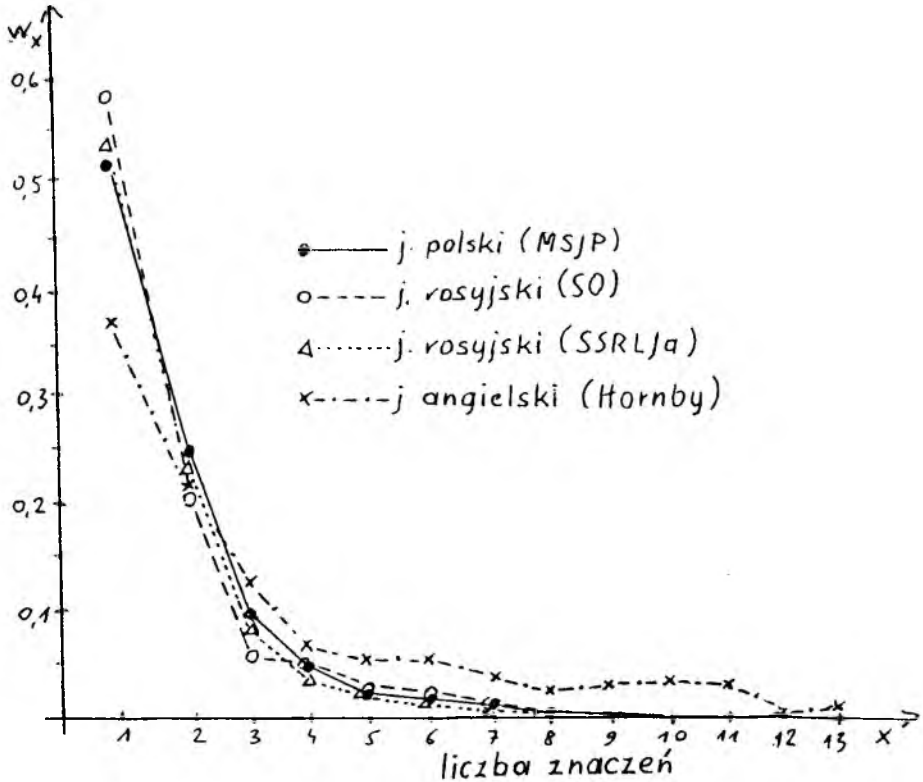
1987), opracowane przez T. Zagrodzką i A. Szelenbaum na seminarium z lingwistyki kwantytatywnej UW. Strukturę statystyczną leksemów mono- i polisemicznych w angielszczyźnie zawiera ostatnia kolumna w tab. 2. zaś interpretację graficzną krzywych dla wszystkich trzech języków podano na rys. 2. (dane dotyczą całego leksykonu; por. też Hammerl, Sambor 1991, 111-122).

Tab. 2. Rozkłady empiryczne leksemów mających x znaczeń w jęz. polskim, rosyjskim i angielskim (dane dla trzech słowników)

liczba znaczeń	udziały w_x leksemów mających x znaczeń					
	MSJP (j. pol.)			SO (j. ros.)		(j. ang.)
	cały słownik	rzecz.	czas.	cały słownik	rzecz.	cały słownik
x						
1	0,517	0,561	0,564	0,588	0,553	0,378
2	0,249	0,272	0,246	0,203	0,296	0,206
3	0,096	0,085	0,087	0,058	0,079	0,112
4	0,054	0,039	0,047	0,050	0,027	0,064
5	0,026	0,014	0,017	0,028	0,020	0,050
6	0,020	0,011	0,013	0,026	0,008	0,050
7	0,018	0,012	0,005	0,015	0,007	0,022
8	0,006	0,004	0,007	0,006	0,003	0,012
9	0,005	0,001	0,005	0,010	0,004	0,016
10	0,003	—	0,002	0,005	0,001	0,020
11	0,003	0,001	0,003	0,004	0,002	0,018
12	—	—	0,001	0,002	—	0,006
13	0,002	—	0,002	—	—	0,008
14	0,001	—	0,001	—	—	0,010
15	—	—	—	—	—	0,006
16	—	—	—	—	—	0,004
17	—	—	—	—	—	0,002
18	—	—	—	—	—	0,006
19	—	—	—	0,005	—	0,002
≥ 20	—	—	—	—	—	0,008
suma	1,000	1,000	1,000	1,000	1,000	1,000

Łatwo zauważyć, że we wszystkich próbach ogromną przewagę ma słownictwo o małym stopniu polisemii; słownictwo wysoce polisemiczne stanowi zupełny margines słownika. Otrzymane wyniki są analogiczne do danych z tab. 1., a zatem nasze wyniki są zbliżone z otrzymanymi przez Kryłowa.

Równocześnie ujawnia się wyraźnie odmienna struktura danych liczbowych dla angielszczyzny: obserwujemy o wiele mniejszy udział słownictwa monosemicznego, a równocześnie występowanie leksemów o bardzo wysokiej polisemii, co



Rys. 2. Rozkłady empiryczne leksemów mających x znaczeń w czterech słownikach (j. polski: MSJP; j. rosyjski: SO, SSRLJa; j. angielski: słownik Hornby¹⁾)

ma swoje konsekwencje w odmiennym przebiegu krzywej w porównaniu z krzywymi dla j. polskiego i rosyjskiego. Ujawnia się tu znana cecha leksyki j. angielskiego, w której występuje relatywnie znaczna liczba tzw. "leksemów-omnibusów", tj. leksemów bardzo pojemnych znaczeniowo typu *to go*, *to take*, których obecność bardzo "wydłuża" przebieg krzywej na rys. 2. Natomiast różnice między leksemami w trzech badanych słownikach obu języków słowiańskich nie są już tak jaskrawe.

Z kolei monotoniczny spadek wszystkich krzywych na rys. 2. świadczy niewątpliwie o identycznej tendencji statystycznej, będącej efektem działania prawa wielkich liczb, które określają strukturę ilościową słownictwa, mającego x znaczeń w konkretnym słowniku.

Jeszcze bardziej wyrazistym przykładem tej tendencji jest porównanie prób z obu języków słowiańskich, dotyczących tej samej części mowy. W obu słownikach – polskim i rosyjskim – zbadano strukturę statystyczną leksemów mających

x znaczeń w odniesieniu do rzeczowników i czasowników (na podstawie próbek po 1000 leksemów). Graficzną interpretację wyników podano w cytowanych pracach (Sambor 1989d;: Hammerl, Sambor 1991, 116), przy czym okazało się, że otrzymane dane porównawcze były zaskakująco zbieżne: krzywe dla rzeczowników i czasowników w polskim słowniku MSJP omal się pokrywały; taki sam przebieg krzywych otrzymano dla rzeczowników, badanych w słowniku polskim i rosyjskim (dane ze słownika Ożegowa).

Wyniki te trzeba uznać za niezwykle intrygujące, nic też dziwnego, że Kryłowa interesowało przede wszystkim poszukiwanie modelu tej zależności. Niestety, badanej prawidłowości nie sprawdzał ani model dany wzorem (1), ani jego poprawiona wersja (omówienie tych modeli w pracach: Hammerl, Sambor 1990, 345-350;:1991, 117-119).

Poszukiwania odpowiedniego modelu dla prawa Kryłowa zostały ostatnio uwiecznione powodzeniem; w jednej z nowszych prac R. Hammerl zaproponował model bardzo dobrze opisujący tę zależność (Hammerl 1991, 139-145).

Obecnie ważnym zadaniem jest zgromadzenie obfitszego materiału empirycznego, dotyczącego prawa Kryłowa w leksyce różnych języków. Na podstawie danych, jakimi dotąd dysponujemy, można sądzić, iż prawo to zdaje się wyznaczać strukturę leksemów mających x znaczeń w leksykonie każdego języka naturalnego i jako takie może mieć charakter uniwersalny. Tezę tę muszą potwierdzić dalsze badania.

6. Rozkład rangowy znaczeń wyrazu polisemicznego, czyli prawo Beðthy

Omawiane powyżej prawo Kryłowa dotyczy wszystkich leksemów, a więc obejmuje cały leksykon, natomiast przedstawiana obecnie zależność – stanowiąca problem semantyki kwantytatywnej – odnosi się do konkretnego leksemu (morfemu) polisemicznego, a mianowicie bada się strukturę częstości poszczególnych znaczeń leksemu (morfemu) polisemicznego, z jakimi znaczenia te pojawiają się w tekście lub w słowniku (ten ostatni wypadek dotyczy tylko morfemów). Okazuje się, że znaczenia te można uszeregować według malejącego porządku ich częstości, a zatem mają one rozkład rangowy.

Podobnie jak w wypadku prawa Kryłowa, także rozkład rangowy znaczeń konkretnego wyrazu polisemicznego tłumaczy G. Altmann działaniem sił Zipfa (Beðthy, Altmann 1984a). Jeśli dany wyraz ma wiele znaczeń, lub pełni kilka funkcji gramatycznych, to znaczenia i funkcje te nigdy nie występują w tekście równie często, ale częstości ich mają rozkład rangowy, który jest wynikiem z jednej strony tendencji nadawcy do unifikacji znaczeń (a więc wyodrębniania małej grupy znaczeń bardzo częstych), zaś z drugiej – tendencji odbiorcy do maksymalnej ich dywersyfikacji (co prowadzi do utrzymywania znacznego zbioru znaczeń wyraźnie zróżnicowanych, które w efekcie występują z małymi częstościami).

W przeciwieństwie do prawa Kryłowa, które – obejmując cały leksyk – dotyczyłoby jakby planu makrosemantyki, rozkład rangowy znaczeń konkretnego leksemu (morfemu) polisemicznego jest przykładem zależności w planie mikrosemantyki. Odkrywcami tej zależności są G. Altmann i E. Beöthy, którzy badali występowanie w tekście noweli I. Örkény'ego *A rózsza kiállitás* znaczeń kilku przedrostków węgierskich (Beöthy, Altmann 1984 a, b, c).

W tabeli 3. przedstawiamy uzyskane przez nich wyniki, dotyczące przedrostków węgierskich *meg-* oraz *ki-*, wraz z podanymi przykładami wyróżnionych znaczeń (Beöthy, Altmann 1984 b, 36).

Autorzy podali też kilka przykładów czasowników, których znaczenia są modyfikowane przez przedrostek *ki-* (Beöthy, Altmann 1984b, 36) (cytujemy za autorami tłumaczenie angielskie węgierskich leksemów – oraz dodatkowo – tłumaczenie polskie, dokonane przez ungarystkę, znającą język polski):

- képez* – "shape, form" (formować, kształtować); *kiképez* – "develop, drill, train" (wytwarzać, tworzyć jakąś nową formę) (10. – znaczenie metaforyczne + nowe znaczenie);
- vív* – "fight" (bić się, walczyć); *kivív* – "achieve, obtain" (wywalczyć) (7. – czynność wykonana + jej rezultat);
- vonul* – "proceed, move" (ruszać); *kivonul* "march, withdraw from" wyruszać w jakimś kierunku) (4. – czynność wykonana + kierunek);
- jön* – "come" (przybywać); *kijön* – "agrees, get along" (zgodzić się na co) (10. – znaczenie metaforyczne + nowe znaczenie);
- octat* – "educate, teach" (nauczać); *kiocstat* – "reprimand" (dać komuś nauczkę) 6. – czynność wykonana + modyfikacja znaczenia);
- facul* – "lose colour" (blaknąć); *kifacul* – "lose colour" (wyblaknąć) (2. – czynność wykonana);
- küld* – "send" (przesyłać); *kiküld* – "send out" (wysłać na zewnątrz) (3. – kierunek);
- hoz* – "bring" (przynieść); *kihoz* – "exasperate" (wygasnąć, zakończyć się) (9. – znaczenie metaforyczne).

W wypadku przedrostka *meg-* wyróżniono jeszcze 4 inne znaczenia, wszystkie rzadko występujące, których w tab. 3. nie wyszczególniono. Łatwo zauważyć, że rozkłady rangowe obu przedrostków znacznie się różnią: cztery pierwsze znaczenia przedrostka *ki-* mają podobne częstości, tj. żadne z tych znaczeń nie dominują wyraźnie nad pozostałymi, natomiast w wypadku przedrostka *meg-* znaczenie 2. wystąpiło w znacznej większości użyć (107 razy na 141 użyć).

Trzeba podkreślić, że autorzy traktują jako znaczenie danego morfemu zarówno jego właściwe znaczenie signifikatywne, jak i funkcje gramatyczne; rozróż-

nienie to jest ważne przede wszystkim w odniesieniu do morfemów gramatycznych, które występują z dużymi częstościami i pełnią różnorodne funkcje. Rozkład rangowy znaczeń ma więc obejmować zarówno znaczenia, jak i funkcje gra-

Tabela 3. Znaczenia węgierskich przedrostków *ki-* oraz *meg-* i ich częstości w tekście *A rózsza kiállitás* I. Orkény'ego

ranga	modyfikacje znaczenia czasowników za pomocą przedrostków <i>ki-</i> oraz <i>meg-</i>	częstości znaczeń <i>ki-</i>	częstości znaczeń <i>meg-</i>
1	czynność wykonana + nowe znaczenie	12	5
2	czynność wykonana	10	107
3	kierunek	8	0
4	czynność wykonana + kierunek	8	0
5	czynność wykonana + znaczenie metaforyczne	5	0
6	czynność wykonana + modyfikacja znaczenia	4	8
7	czynność wykonana + jej rezultat	3	5
8	kierunek + znaczenie metaforyczne	3	0
9	znaczenie metaforyczne	2	0
10	znaczenie metaforyczne + nowe znaczenie	2	0
11	nowe znaczenie	2	8
12–15	cztery inne znaczenia	0	8
	łączna liczba użyć	59	141

matyczne leksemu (morfemu), określane przez autorów jako znaczenia signifikatywne i gramatyczne.

Kolejnym przykładem, ilustrującym rozkład znaczeń określonego morfemu, jest praca, w której analizuje się rozkład rangowy znaczeń niemieckiego sufiksu *-os// -ös* w zbiorze derywatów przymiotnikowych języka niemieckiego (Altmann, Best, Kind 1987); materiał zebrano na podstawie 3-tomowego opracowania *Deutsche Wortbildung* (DW). W tab. 4 podajemy rozkład rangowy znaczeń tego sufiksu (Altmann, Best, Kind 1987, 131).

Rozkłady rangowe znaczeń określonego leksemu bądź morfemu polisemicznego są o tyle intrygujące, że we wszystkich – nielicznych dotąd – przy-

kładach zdają się one przebiegać według pewnego ściśle określonego modelu probabilistycznego, a więc nie są one przypadkowe, co sprawdzono w odniesieniu zarówno do rozkładów znaczeń przedrostków węgierskich, jak i sufiksu niemieckiego *-os// -ös*. Podobnie, jak w wypadku prawa Kryłowa, dla opisu którego R. Hammerl znalazł ostatnio odpowiedni model probabilistyczny, także i omawiane prawo Beöthy można opisać z dużą dokładnością za pomocą modelu, jakim jest ujemny rozkład dwumianowy, czego dowiedli autorzy obu wyżej omawianych prac (omówienie tego modelu w: Sambor, Hammerl 1991, 67-73, 139-145).

Tabela 4. Częstości znaczeń sufiksu *-os// -ös* w niemieckich derywatach przymiotnikowych (Altmann, Best, Kind 1987, 131)

ranga	grupy nominalne Adj + Subst.	parafraza słowotwórcza	podstawa słowotwórcza	częstość
1	<i>ambitiöser Sportler</i>	<i>Sportler, der Ambitionen hat</i>	<i>Ambition</i>	59
2	<i>gelatinöses Nebenprodukt</i>	<i>Nebenprodukt, das wie Gelatine ist</i>	<i>Gelatine</i>	31
3	<i>inzestuöse Verbindung</i>	<i>Verbindung, die ein Inzest ist</i>	<i>Inzest</i>	20
4	<i>ruinöser Wettbewerb</i>	<i>Wettbewerb, der den Ruin bewirkt</i>	<i>Ruin</i>	6
5	<i>tuberkulöse Erkrankung</i>	<i>Erkrankung, die die Tuberkel (Bakterie) hervorrufen</i>	<i>Tuberkel</i>	5
6	<i>seröse Flüssigkeit</i>	<i>Flüssigkeit, die aus Serum besteht</i>	<i>Serum</i>	4
7	<i>religiöser Zweifel</i>	<i>Zweifel, der die Religion betrifft</i>	<i>Religion</i>	2
8	<i>venöses Blut</i>	<i>Blut, das sich in der Vene befindet</i>	<i>Vene</i>	2
9 i więcej	<i>medikamentöse Behandlung</i>	<i>Behandlung, die durch Medikamente geschicht</i>	<i>Medikament</i>	2
łącznie liczba derywatów:				131

Prawidłowość powyższa nie jest jednak łatwa do badania, a uwagi krytyczne i wątpliwości mogą się nasuwać językoznawcy już przy lekturze danych obu tabel, gdyż wiadomo, jak wielkie trudności nasuwa klasyfikacja znaczeń jednostki polisemicznej lub wielofunkcyjnej, jak niełatwą sprawą jest wyznaczenie rozłącz-

nych zakresów poszczególnych znaczeń; problem stanowią też znaczenia metaforyczne, które na ogół klasyfikuje się łącznie ze znaczeniami dosłownymi itd.

Parafrazy semantyczne, podane w tab. 4., nie stanowią definicji semantycznych, bo zdają się dotyczyć tylko konkretnej grupy nominalnej, choć można się domyślać, że kryją się za nimi jako definicje jakieś eksplikacje semantyczne typu "X, który wygląda jak Y" (*gelatinöses Nebenprodukt*), "X, który powoduje Y" (*ruinöser Wettbewerb*), "X, który wywołuje Y" (*tuberkulöse Erkrankung*), "X, który składa się z Y-a" (*seröse Flüssigkeit*), "X, który dotyczy Y-a" (*religiöser Zweifel*) itd.

Ciekawe rozwiązanie metodologiczne dotyczące sposobu wyodrębniania znaczeń morfemu wielofunkcyjnego zawiera praca U. Rothe, poświęcona rozkładowi rangowemu znaczeń francuskiego spójnika *et* w powieści *Le petit prince* A. Saint-Exupéry'ego; przytaczamy fragment początkowy tego rozkładu (Rothe 1986, 66):

Tabela 5. Rozkład rangowy znaczeń spójnika *et* w powieści *Le petit prince* A. Saint-Exupéry'ego

ranga x	parafraza (= znaczenie)	częstość f_x
1	<i>alors</i>	28
2	<i>puis</i>	20
3	<i>mais</i>	15
4	<i>mais</i> , prg.	11
5	konstrukcje z gerundium	9
6	<i>de même que</i>	6
7	<i>enfin</i>	6
8	<i>id-dessus</i>	6
9	<i>alors</i> , prg.	5
10	<i>c'est pourquoi</i>	5
.	.	.
.	.	.
.	.	.
49	<i>tandis que</i>	1
łączna suma użyć		202

Jak widać, autorka nie definiuje żadnych znaczeń tego spójnika, podając jedynie jego różne odpowiedniki formalne, a więc posługuje się tu dystrybucyjną metodą substytucji. Okazało się, że także i ten empiryczny rozkład rangowy potwierdzał prawo Beöthy, co wykazano odpowiednim testem statystycznym. Tak więc ujawniająca się – mimo oczywistych trudności w klasyfikacji znaczeń – regularność statystyczna w ich rozkładzie rangowym jest czymś wręcz intrygującym.

Obecnie planujemy przebadanie prawa Beðthy na materiale polskich przyimków. Pierwszą próbą jest praca poświęcona przyimkowi *w*, którego strukturę znaczeń (funkcji) sprawdzono we fragmencie powieści *Droga wiodła przez Narvik* K. Pruszyńskiego (uwzględniono 7 pierwszych rozdziałów, uzyskując ca 500 użyć przyimka) (Sambor 1989 c).

Zwracaliśmy już uwagę na trudności w rozróżnianiu znaczeń jednostek poli-semicznych. Semantyka najczęstszego i wielofunkcyjnego przyimka *w* jest ogromnie skomplikowana, dlatego też we wstępnym etapie badań nie proponowano własnej klasyfikacji znaczeń funkcji tego przyimka, naprzód bowiem chodziło o sprawdzenie opisów już istniejących. Wykorzystano mianowicie i porównano opisy funkcji przyimka *w*, zawarte w MSJP, w *Słowniku poprawnej polszczyzny* oraz w trzynomowym *Słowniku Języka polskiego* (dalej jako SPP i SJP).

Porównanie to wykazało, iż klasyfikacje funkcji przyimka *w*, zawarte w MSJP oraz w SJP, są wysoce niezadowolające (wyników dla SJP nie opublikowano), zakresy niektórych funkcji są nierozłączne itp., dlatego też dane liczbowe dotyczące częstości tych funkcji w tekście mają charakter przybliżony; np. w wypadku niektórych funkcji trzeba było przyjąć rozstrzygnięcia konwencjonalne, poprawiając niejako słowniki. Jediną klasyfikacją, którą można było stosować z pewnymi modyfikacjami, był opis tego przyimka podany w SPP, dlatego też poniżej omówimy bardziej szczegółowo wyniki, uzyskane na podstawie tego właśnie opisu.

Autorzy SPP wyróżnili następujących 11 znaczeń (funkcji) przyimka *w* (podajemy je w skrócie):

1. Przyimek *w* w połączeniu z miejscownikiem rzeczowników tworzy wyrażenia odpowiadające na pytanie "gdzie?" i określające miejsce dziania się, położenie czegoś, przebywanie kogoś gdzieś, zwłaszcza we wnętrzu, w środku, w głębi przedmiotu nazwanego przez ten rzeczownik, np. *w kotlinie*, *w kącie*, *w szufladzie* (miejsce lokatywne).

2. Przyimek *w* w połączeniu z rzeczownikiem w bierniku tworzy wyrażenia odpowiadające na pytanie "dokąd?" i określające miejsce, które jest celem jakiegoś ruchu. Wyrażenia te odpowiadają miejscownikowym połączeniom z przyimkiem *w*, np. *być w górach* – *jechać w góry*, *w zaroślach* – *uciec w zarośla* itd. (miejsce adlatywne).

3. Przyimek *w* tworzy wyrażenia oznaczające czas, porę dziania się czegoś – z miejscownikiem i biernikiem; np. *w ostatniej chwili*, *w zeszłym miesiącu*, *w żniwa*, *w pogodę*, *we wtorek*, *w południe* (czas).

4. Wyrażenia określające postać, charakter, układ, kolor czegoś, np. a) z miejscownikiem: *cukier w kostkach*, *format w ósemce*, *dzieło w małej liczbie egzemplarzy*; *rzeźbić w marmurze*, *materiał w piaskowym kolorze*; b) z biernikiem:

złożyć coś w kostkę, broń ustawiona w kozły, stoły w podkowę, materiał w kwiatki, w kratę.

5. Wyrażenia związane ze stanami, uczuciami: a) z miejscownikiem: *w szale, w gorączce, w bólach, we śnie, w rozpacz, w humorze*; b) z biernikiem: *wpaść w dobry humor, w pasję* itp. (stan).

6. Wyrażenia funkcjonujące jak przymiotniki lub przysłówki: a) z miejscownikiem: *być w modzie, w cenie, żyć w nędzy, chodzić w stawie; iść w podskokach, w milczeniu, w gromadzie* itp.; b) z biernikiem: *iść w pojedynkę, we dwoje, śmiać się w głos, pojąć coś w lot* itp.

7. Przyimek łączy się z nazwami ubioru, jego części, narzędzi lub narzędzia chwytanego a) z miejscownikiem: *w czapkach, w gumowych butach, trzymać coś w zębach, mieć coś w imadle*, b) z biernikiem: *ubrany był w płaszcz, w gumowe buty*.

8. Wyrażenie uściślające treść, zakres wyrazu, do którego się odnosi (tylko z miejscownikiem): *miły w dotknięciu, pochylony w ramionach, nie uchybiał nikomu w grzeczności, ostygł w uczuciach*.

9. Wyrażenia oznaczające (tylko z biernikiem): a) cel czynności: *rzucić się w pogoń, iść w odwiedzin, w zawody*; b) wynik, skutek czynności, inny niż poprzednio stan czegoś: *szmer przechodził w hałas, z nicponia przeobraził się w porządnego człowieka. Połamał stół w drzazgi. Rozpaść się w gruzy, w kawałki*.

10. Przyimek w występuje w równoważnikach zdań – z biernikiem: *Zobaczyła go i w krzyk! I tak w kółko. W nogi! W konie!*

11. Przyimek w występuje w utartych wyrażeniach: a) z miejscownikiem: *w ciągu, w końcu, w ogóle, w oka mgnieniu, w tych dniach* itp. b) z biernikiem: *krok w krok, kropka w kropkę, chłop w chłop, w bród, w poprzek, w zamian, ni w pięć ni w dziewięć* itp.

W porównaniu z opisami w dwóch pozostałych słownikach podział podany w SPP budzi o wiele mniej wątpliwości, jednakże nie wyróżniono tu np. wyrażen typu *w + nazwy instytucji, grupy społecznej (w szkole, w wojsku)*, które przy badaniu materiału empirycznego oczywiście trzeba było uwzględnić (w tabeli 6. dane te podano w nawiasach).

Badając użycia przyimka *w* we fragmencie powieści *Droga wiodła przez Narvik* według powyższej klasyfikacji otrzymano następujące dane empiryczne (tab. 6., por. Sambor 1989 c). Jak wynika z tabeli, kolejność znaczeń przyimka *w*, przyjęta w badanym słowniku, jest zupełnie odmienna od kolejności tych znaczeń w tekście, ustalonej na podstawie ich częstości, przy czym częstości te mają rozkład rangowy, czyli niektóre z nich wyraźnie dominują w badanym tekście.

W cytowanej pracy chodziło o sprawdzenie stopnia zgodności kolejności (=rang r_1) odpowiednich znaczeń, podawanych przez leksykografa, z ich rzeczywistą kolejnością według częstości w tekście (rangi r_2). W tym celu wykorzy-

stano wzajemny układ dwóch rozkładów rangowych według rang r_1 i r_2 , podanych w tabeli 6., i obliczono odpowiednim współczynnikiem korelacji rangowej R stopień zgodności tych rang.

Dla danych tab. 6. współczynnik R wynosił + 0,6058, natomiast dla MSJP – tylko + 0,2086 (bliższy opis obliczania współczynnika R i jego interpretacja – por. Sambor 1989 c, także Hammerl, Sambor 1990, 185). Wynik ten oznacza, iż w SPP zaobserwowano daleko wyższy stopień zbieżności obu rozkładów rangowych, a zatem uszeregowanie znaczeń przyimka *w* w SPP jest o wiele bliższe rozkładowi rangowemu tych znaczeń według ich częstości w badanym tekście.

Tabela 6. Rangi znaczeń przyimka *w* w SPP i częstości tych znaczeń we fragmencie powieści *Droga wiodła przez Narvik* K. Pruszyńskiego

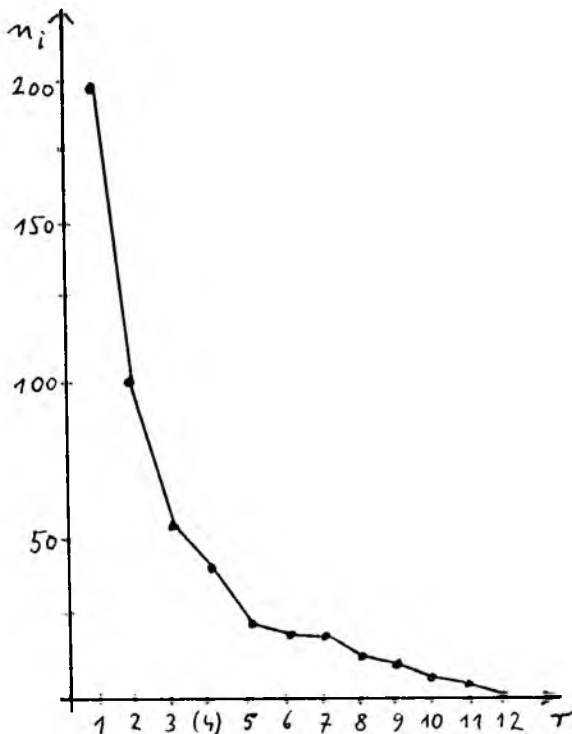
rangi znaczeń w SPP	częstości w tekście oraz ich rangi	rangi znaczeń w SPP	częstości w tekście oraz ich rangi
r_1	r_2	r_1	r_2
1. miejsce lok.	1. 198	9. cel, wynik, skutek	7. 20
2. miejsce adlat.	2. 100	10. równow. zdania	12. 1
3. czas	3. 54	11. wyrażenia zleksykaliz.	8. 15
4.	11. 7		
5. stan	5. 21	(12) grupa, instytucja	(4). (40)
6. jaki? jak? (sposób)	6. 20		
7.	10. 8		
8.	9. 13		
użycia zgodne z klasyfikacją w SPP:			457
łączna suma wszystkich użyć:			(497)

Poniżej podajemy interpretację graficzną empirycznego rozkładu rangowego znaczeń przyimka *w* w badanym tekście według opisu w SPP (por. rys. 3).

Jest rzeczą godną uwagi, iż mimo ewidentnych błędów w klasyfikacjach znaczeń przyimka *w*, przyjętych w MSJP i w SPP, w rozkładach rangowych ułożonych dla *w s z y s t k i c h t r z e c h s ł o w n i k ó w* na pierwszych miejscach pojawiły się te same znaczenia w identycznej kolejności, wyliczymy je według rang: 1. miejsce, 2. czas, 3. grupa, instytucja, 4. sposób, 5. stan. Wysokie częstości znaczenia 'grupa, instytucja' są spowodowane niewątpliwie charakterem badanego tekstu: powieść z czasów wojny zawierała wiele określeń typu *w wojsku*, *w sztabie*, *w batalionie*, *w kompanii* itp., w innym typie tekstu wyrażenia te mogłyby się pojawić z bardzo małymi częstościami.

W dalszym etapie prac sprawdzono rozkłady empiryczne znaczeń (funkcji) przyimka *w* w całej powieści Pruszyńskiego, wykorzystując jako ulepszone kryte-

rium ich wyróżniania bardzo pomocny test pytań A. Weinsberga (Sambor 1991). Okazało się, że rozkład empiryczny tych znaczeń, otrzymany dla fragmentu powieści, niewiele różnił się od rozkładu dla całego tekstu, podobna też była kolejność znaczeń najczęstszych.



Rys. 3. Empiryczny rozkład rangowy znaczeń w tekście przyimka *w* według ich klasyfikacji w SPP (w nawiasie ranga znaczeń w SPP pominiętych)

Kolejną hipotezą było przypuszczenie, że częstości znaczeń morfemu wielofunkcyjnego mają wprawdzie w tekstach zawsze rozkład rangowy, ale kolejności samych znaczeń zależą od rodzaju tekstu.

Hipotezę tę sprawdziła M. Golanowska (1992), badając rozkłady rangowe funkcji przyimka *w* w powieści współczesnej (*Siekierzada* E. Stachury), w tekście popularnonaukowym (podręcznik szkolny do astronomii) oraz w felietonach (teksty Z. Kałużyńskiego). Okazało się, że:

1) we wszystkich trzech typach tekstów 2 najczęstsze funkcje przyimka *w* to tworzenie wyrażzeń oznaczających miejsce i czas (*w domu*, *w maju*);

2) kolejność pozostałych funkcji istotnie zależy od rodzaju tekstu, być może, od stylu osobniczego itp.

7. Ciągi definicyjne

7. 1. Ciągi definicyjne i tzw. prawo Martina

Pojęcie ciągu definicyjnego zawdzięczamy pracy francuskiego lingwisty R. Martina (1974). Dla danego leksemu typu *powietrze* szukamy w określonym słowniku jednojęzycznym (np. w MSJP) jego definicji, wyodrębniając w niej hiperonim, czyli leksem wskazujący na pojęcie nadrzędne (*gaz*). Z kolei w definicji dla leksemu *gaz* odnajdujemy leksem o znaczeniu nadrzędnym – *substancja*. Układając ciąg definicyjny dla leksemu *powietrze* według MSJP, natrafiamy dalej na definicje cyrkularne (*substancja*→*materia*, *materia*→*substancja*, a zatem ciąg *powietrze*→*gaz*→*substancja* uznajemy za zakończony. Łatwo zauważyć, że kolejne leksemy w ciągu odpowiadają pojęciom bardziej ogólnym, ciąg definicyjny odzwierciedla więc hierarchiczną strukturę słownictwa, gdyż ujmuje on relacje hiponimiczno-hiperonimiczne w leksykonie.

Ciągi definicyjne, ułożone dla znacznej liczby leksemów danego języka, nasuwają cały szereg ważnych problemów. R. Martin, który badał próbę 1723 ciągów, ułożonych dla rzeczowników, wylosowanych ze słownika języka francuskiego Gougenheima, analizował regularność w budowie statystycznej tych ciągów, który to problem podjęli następnie G. Altmann i E. Kind (1983) – do tematu tego wrócimy jeszcze w dalszych rozważaniach.

Kolejnym badaniem była analiza 1000 ciągów dla polskich rzeczowników, wylosowanych z MSJP (Sambor 1986), materiałem porównawczym zaś – 1000 ciągów dla ekwiwalentów tych samych rzeczowników w języku niemieckim, jakie ułożył R. Hammerl (1987 a, b, c). Ponieważ problemy związane z ciągami definicyjnymi są dość szczegółowo omówione w opublikowanych już polskich pracach (Sambor 1986, Hammerl 1987 c, Hammerl, Sambor 1991, 42-87), tu ograniczymy się tylko do skrótowego ich naszkicowania, odsyłając Czytelników zainteresowanych tą tematyką do wskazanych publikacji.

Przy układaniu ciągów definicyjnych dla rzeczowników polskich okazało się, że definicje leksemów w MSJP nie zawsze miały strukturę dwudzielną typu *genus proximum + differentia specifica*, wskutek czego w niektórych wypadkach autorka opracowania "dotwarzała" te ciągi według własnej intuicji (dotyczyło to około 20% rzeczowników). W dalszych badaniach wypracowano już bardziej ścisłe kryteria tworzenia ciągów definicyjnych, zalecając np. pomijanie rzeczowników o ciągach wątpliwych, skreślenie obu ostatnich członów w definicjach cyrkularnych i in. (Hammerl, Sambor 1991, 51-56).

Otrzymane ciągi miały różne długości, przykłady zawiera tab. 7.

Tabela 7. Struktura ciągów definicyjnych

poziom leksemowy x	ciągi definicyjne					
1	powietrze ↓	amoniak ↓	fauna ↓	flora ↓	amator ↓	ciastkarnia ↓
2	gaz ↓	gaz ↓	ogół (zwierząt)	ogół (roślin)	zwolennik ↓	sklep ↓
3	substancja	substancja			człowiek ↓	lokal ↓
4					istota żywa	miejsce ↓
5						przestrzeń ↓
6						obszar ↓
7						część

Łatwo zauważyć, że w badanych ciągach pojawiają się kolejne poziomy abstrakcji. Poziomowi $x = 1$ odpowiadają leksemy badane a więc wylosowane ze słownika, poziomowi $x = 2$ ich hiperonimy, czyli leksemy dla nich nadrzędne znaczeniowo, odszukane z definicji itd. Niech z kolei y_x oznacza liczbę leksemów występujących w ciągach na poziomie x . Przytoczone powyżej przykłady wskazują, iż na pozycjach $x \geq 2$ pewne leksemy mogą się powtarzać (por. leksemy *gaz*, *substancja*, *ogół*), niech zatem zmienna y_x oznacza liczbę leksemów, występujących na poziomie x b e z p o w t ó r z e ń.

Okazuje się, że dla trzech badanych języków (francuskiego, polskiego i niemieckiego) struktura liczbowa analizowanych ciągów miała postać następującą (dane liczbowe z prac: Martin 1974, Sambor 1986, Hammerl 1991 a): (tab. 8).

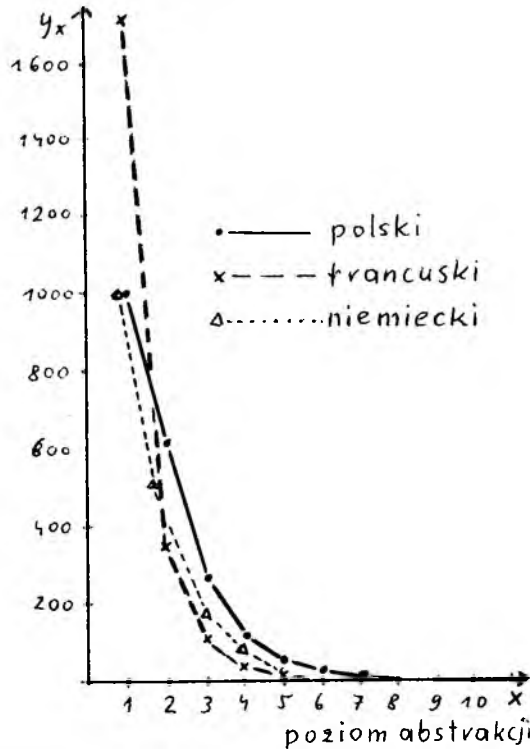
Jak widać, w miarę wzrostu zmiennej x zmienna y_x gwałtownie maleje, zatem im wyższy jest poziom abstrakcji x , tym mniejsza jest liczba występujących na tym poziomie leksemów. Jest to zależność łatwo wytłumaczalna: im wyższy jest poziom x , tym bardziej ogólne znaczenia mają występujące na danym poziomie leksemy, a wiadomo, że leksemy o znaczeniach wysoce abstrakcyjnych stanowią w każdym słowniku zbiór bardzo niewielki.

Dane dla wszystkich trzech języków sprawdzają tę regularność, choć struktura liczbowa tych ciągów jest z gruntu odmienna. Ciągi francuskie są o wiele krótsze, poza tym uderza znaczna liczba leksemów na poziomach wyższych, stąd wniosek że w ciągach francuskich wiele leksemów powtarzało się na tych pozio-

Tabela 8. Struktura liczbowa ciągów definicyjnych dla języka francuskiego, polskiego i niemieckiego (dane bez powtórzeń na poziomie x)

poziom x	liczba leksemów y_x					
	francuski		polski		niemiecki	
	abs.	%	abs.	%	abs.	%
1	1723	100,0	1000	100,0	1000	100,0
2	348	20,2	618	61,8	512	51,2
3	108	6,3	271	27,1	183	18,3
4	39	0,02	110	11,0	72	7,2
5	13	0,007	44	4,4	31	3,1
6	3		16	1,6	14	1,4
7			9	0,9	9	0,9
8			3	0,3	5	0,5
9			1	0,1	3	0,3
10					2	0,2
11					1	0,1

mach. Dla kilkunastu leksemów niemieckich uzyskano ciągi bardzo długie, ciągi polskie zajmują tu pozycję pośrednią.



Rys. 4. Zależność Martina (dane dla polskiego, francuskiego i niemieckiego)

Zależność między poziomem x a liczbą leksemów y_x , występujących bez powtórzeń na poziomie x , nazwano w literaturze prawem Martina (Altmann, Kind 1983); dane empiryczne absolutne z tabeli 8. ilustruje rys. 4. (Hammerl 1991):

Zależność ta zainteresowała przede wszystkim badaczy-teoretyków; przedmiotem pracy Altmana i Kinda było poszukiwanie modelu funkcyjnego, opisującego w sposób optymalny tę zależność (Altmann, Kind 1983). Model ten ulepszył potem R. Hammerl, proponując modyfikację funkcji obu autorów (Hammerl 1987 c). Autor ten sporządził także statystykę ciągów definicyjnych o długości x (dane dla języka polskiego i niemieckiego, Hammerl 1987 c, 423; 1989 a), którą ujmuje tabela 9.:

Tabela 9. Liczba ciągów definicyjnych n_x o długości x leksemów (jęz. polski i niemiecki)

długość ciągu w leksemach x	polski n_x	niemiecki n_x
1	0	104
2	238	257
3	251	267
4	249	183
5	159	96
6	68	60
7	30	19
8	4	9
9	1	3
10	0	1
11	0	1

Różnice w długości ciągów, ujawniające się między oboma językami dla zmiennej $x = 1$, wynikają z odmiennego traktowania w obu opisach ciągów zakończonych wyrażeniami, nie stanowiącymi leksemów (typu: *ciemieżyciel* → *ten, kto ciemięży*). W ciągach niemieckich skreślano takie człony, uzyskując ciągi jednoelementowe, w opisie polskim ciągi te traktowano jako dwuelementowe.

Łatwo zauważyć, że w obu językach przeważają ciągi o długości trzech leksemów, zaś ogromna większość ciągów liczy od dwóch do czterech leksemów. Hammerl zaproponował też bardzo interesujący wskaźnik siły podporządkowania g_x , określający, ile przeciętnie leksemów z poziomu $x - 1$, dla których istnieje leksem nadrzędny na poziomie x , może być podporządkowanych jednemu leksemowi z poziomu x . Szczegóły dotyczące sposobu obliczania tego wskaźnika i jego interpretacji zawierają odpowiednie prace autora (Hammerl 1987 b, c.). Tu nadmienimy jedynie, że właśnie wartości tego wskaźnika podważyły sensowność uwzględniania w ciągach polskich aż 9 poziomów abstrakcji – można sądzić, że

poziomy $x = 7$, $x = 8$ oraz $x = 9$ są zbędne z punktu widzenia relacji hiponimicznych i stopnia ogólności leksemów występujących na tych poziomach.

7. 2. Indefinibilia semantyczne

Jak wynika z cytowanego przeglądu prac, zarówno Martina jak i jego kontynuatorów interesowały w ciągach definicyjnych przede wszystkim problemy statystyczno-teoretyczne. Jednakże ciągi te dostarczają także materiału do bardzo interesujących poszukiwań semantycznych o charakterze jakościowym, dotyczących indefinibiliów semantycznych.

Przy układaniu ciągów zauważono mianowicie, że na odpowiednio wysokich poziomach abstrakcji pojawiają się – chyba nieprzypadkowo – definicje cyrkularne, zawierające leksemy o znaczeniach bardzo ogólnych. Przykładami mogą być ciągi: *powietrze*→*gaz*→*substancja* (→*materia*→*substancja*), *agawa*→*roślina*→*istota żywa* (→*organizm*→*istota żywa*), *akumulator*→*urządzenie* (→*mechanizm*→*urządzenie*). Sądzić można, że leksemy typu *substancja*, *materia*, *organizm*, *istota żywa* (także *przedmiot*, *rzecz*) są bardzo trudne do zdefiniowania, między członami odpowiednich par nie wyczuwa się też relacji hiperonimii – stąd "ratowanie się" leksykoграфów definicjami cyrkularnymi.

Uważamy, że te właśnie leksemy odpowiadają pojęciom, które można byłoby uznać za indefinibilia semantyczne. W badanych polskich ciągach wyodrębniono też leksemy występujące jako *o s t a t n i e* człony ciągów i to bez powtórzeń (tzn. jeżeli leksem wystąpił na poziomie x , to nie liczone go dla poziomów niższych, czyli dla poziomów $x - 1$, $x - 2$ itd.), odpowiednie dane liczbowe i wyszczególnienie leksemów końcowych zawiera tabela 10 (Sambor 1986, 163). Zwróćmy uwagę na fakt, że w materiale naszym występują dwa rodzaje zbiorów leksemów "bez powtórzeń": kolumna 2 tabeli 10. zawiera zbiory leksemów nie powtarzających się na poziomie x (por. powtórzenie w tabeli 7.), natomiast powtarzających się na pozostałych poziomach. Kolumna 3 obejmuje zbiory leksemów występujących tylko raz na poziomie x , nie liczone ich dla poziomów niższych.

Tabela 10. zawiera bardzo interesujący materiał, gdyż ujawniają się tu leksemy końcowe, występujące na najwyższych poziomach abstrakcji. Niektóre z nich występowały w cytowanych definicjach cyrkularnych, a wszystkie one odpowiadają pojęciom o bardzo wysokim stopniu ogólności. Najczęściej powtarzającym się leksemem końcowym w ciągach była *część* – nie jest przypadkiem, że w znanym modelu 13 semów, proponowanym przez A. Wierzbicką (1972), występuje ten właśnie sem, pojawiający się niezwykle często w definicjach słownikowych.

Jeśli wymienione przez nas wyżej leksemy miałyby wskazywać na indefinibilia semantyczne, należałoby się spodziewać, że powinny one występować także w ciągach definicyjnych, układanych dla innych języków. W 1987 r. powstała praca

Tabela 10. Struktura liczbowo leksmów końcowych bez powtórzeń

poziom x	ciągi z powtórzeniami na poszczególnych poziomach	ciągi bez powtórzeń	leksemy końcowe
1	1000		
2	618	347	
3	271	161	
4	110	66	
5	44	28	np.: powierzchnia, proces, konstrukcja, aparat, ciecz, właściwość, sprzęt, narząd, pomieszczenie, zbiór, ogół/suma, grupa, kawalek
6	16	7	ciało, miejsce, przedmiot, układ, urządzenie, cecha, wytwór
7	9	6	istota żywa, mechanizm, organizm, przestrzeń, struktura, substancja
8	3	2	materia, obszar
9	1	1	część

poświęcona ciągom dla języka serbsko-chorwackiego, w której analizowano 300 ciągów definicyjnych dla 300 rzeczowników, stanowiących ekwiwalenty serbsko-chorwackie odpowiednich rzeczowników polskich, wylosowanych przez nas do badania ciągów definicyjnych (Komuda 1987). Autorka tego opracowania otrzymała interesujące wyniki, okazało się mianowicie, że w badanym słowniku jednojęzycznym języka serbsko-chorwackiego wystąpiły te same definicje cyrkularne typu *materija*→*tvar* (*materia*→*substancja*), *predmet*→*stvar* (*przedmiot*→*rzecz*), *organizm*→*biće* (*organizm*→*istota żywa*): również w ciągach serbsko-chorwackich najczęstszym leksemem końcowym był leksem *deo* (*część*) itp. Zbieżność otrzymanych wyników była uderzająca.

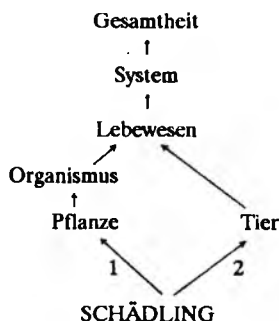
Na zakończenie dodajmy, że w ciągu ostatnich kilku lat dokonano także ważnych badań porównawczych nad ciągami definicyjnymi języka polskiego i niemieckiego, rozszerzono zakres badanych pojęć końcowych o dane uzyskane dla niemieczyny, przy czym pojęcia te stały się punktem wyjścia do odkrycia tzw. gniazd leksykalnych i pojęciowych (o czym niżej w par. 7. 3). Czytelników zainteresowanych tą tematyką odsyłamy do monografii (Hammerl, Sambor 1991, 42-87 Sambor, Hammerl 1991), zawierających m. in. wszechstronne omówienie problematyki związanej z ciągami definicyjnymi.

7. 3. Gniazda leksykalne

Badanie gniazd leksykalnych "wyrosło" niejako z analizy ciągów definicyjnych. Nie wspominaliśmy dotąd o tym, że w wypadku leksemów wieloznacznych, wylosowanych w próbie dla języka polskiego, układano ciąg definicyjny tylko dla znaczenia, wymienianego w badanym słowniku jako pierwsze (gdyż układanie ciągów dla pozostałych znaczeń w próbie 1000 leksemów doprowadziłoby do znacznego powiększenia zbioru ciągów, czego chcieliśmy uniknąć w pierwszym etapie pracy). Później jednak okazało się, że układając ciągi definicyjne dla wszystkich znaczeń leksemów polisemicznych otrzymuje się bardzo ciekawą strukturę ciągów tworzących swoiste gniazdo leksykalne, charakterystyczne dla danego leksemu.

Tematyką gniazd leksykalnych zajął się R. Hammerl, którego interesowała przede wszystkim analiza statystyczno-formalna budowy tych gniazd. Wylosował on 100 leksemów ze zbioru 1000 leksemów języka niemieckiego, dla których układał uprzednio ciągi definicyjne, i opracował dla nich (na podstawie tego samego słownika HDG) 100 gniazd leksykalnych, które w wypadku leksemów wieloznacznych składały się z co najmniej dwóch ciągów, zaś gniazda leksemów jednoznacznych pokrywały się z ich ciągami definicyjnymi.

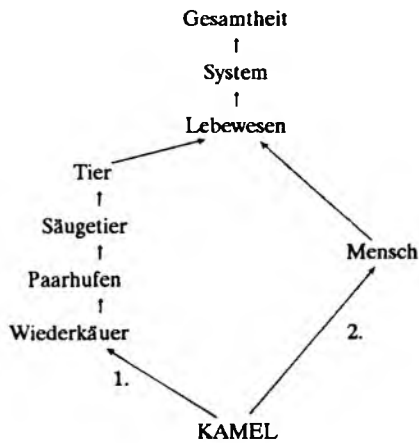
Podane poniżej przykłady gniazd pochodzą z dwóch prac Hammerla (1989 b, 1990); jedynie dwa przykłady cytuję – za zgodą autora – z materiałów jego seminarium. Budowę gniazda niech unaoczní prosty przykład dla leksemu *Schädling* ('szkodnik'), który w słowniku HDG miał dwa znaczenia: 1. *Pflanze* ('roślina'), 2. *Tier* ('zwierzę'):



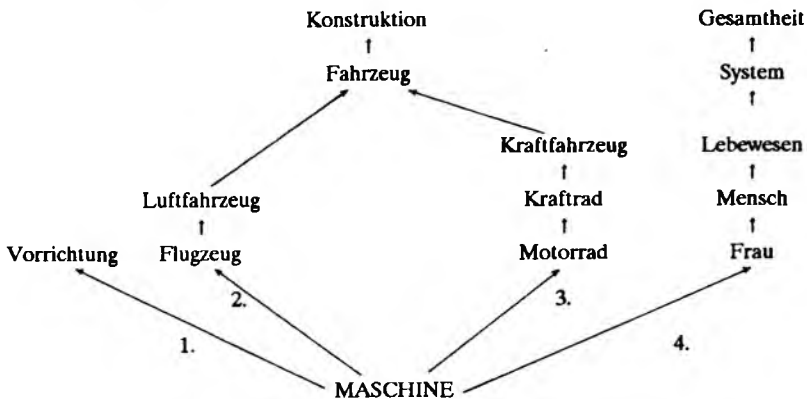
Łatwo zauważyć, że oba ciągi definicyjne schodzą się niejako na odpowiednio wysokim poziomie abstrakcji, a mianowicie wspólny jest im obu hiperonim *Lebewesen*. Ujawnia się tu wewnętrzna relacja między ciągami tworzącymi gniazdo leksykalne – w tym wypadku jednemu hiperonimowi nadrzędnemu są przyporządkowane wszystkie hiperonimy, występujące w ciągach na niższych poziomach. Przy takiej budowie gniazda na najwyższych poziomach abstrakcji występuje już tylko jeden ciąg, zawierający hiperonimy nadrzędne dla wszystkich hiperonimów z niższych poziomów abstrakcji.

Trzeba przy tym pamiętać, że w strukturze gniazda leksykalnego należy uwzględnić w s z y t k i e znaczenia danego leksemu, a więc także znaczenia nacechowane stylistycznie (emocjonalizmy, ekspresywizmy itp.), gdyż one właśnie rozbudowują najczęściej gniazdo leksykalne leksemu. Dobrym przykładem, jaki podaje Hammerl, jest wewnętrzna struktura ciągów dla leksemu *Kamel* (1. 'wielbłąd', 2. 'dureń'), którego gniazdo cytujemy poniżej.

Tak ułożone gniazda leksykalne Hammerl analizował ze względu na 3 cechy liczbowe, jakie można wyróżnić w ich budowie; cechami tymi były: a) liczba hiperonimów końcowych w danym gnieździe, b) liczba gałęzi, prowadzących od leksemu początkowego do hiperonimów końcowych, c) łączna liczba wszystkich hiperonimów w gnieździe (w rzeczywistości badano jeszcze tzw. przeciętną wysokość i szerokość gniazda, ale w dalszych pracach autor z tych dwóch cech zrezygnował).



Aby uprzytomnić czytelnikowi realność wymienionych 3 cech, przytoczymy przykład gniazda bardziej rozbudowanego semantycznie, dotyczącego leksemu *Maschine*, mającego w HDG 4 różne znaczenia:

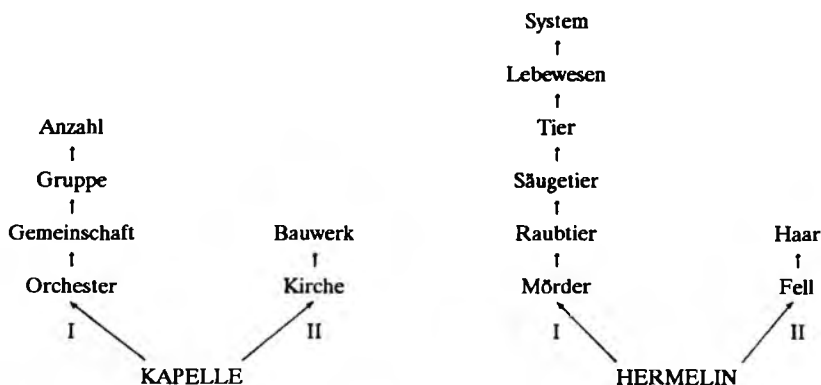


Strukturę statystyczną tego gniazda określają trzy hiperonimy końcowe (*Vorrichtung, Konstruktion, Gesamtheit*), 4 gałęzie oraz 13 hiperonimów składowych.

Okazało się, że w badanym zbiorze 100 gniazd pojęciowych niemal połowę stanowiły gniazda pokrywające się z pojedynczymi ciągami, a więc odpowiadające w zasadzie leksemom jednoznacznym, co jest prostą konsekwencją prawa Kryłowa, nic więc dziwnego, że także większość pozostałych gniazd miała niezbyt rozbudowaną strukturę, podobną do gniazda dla leksemu *Schädling*.

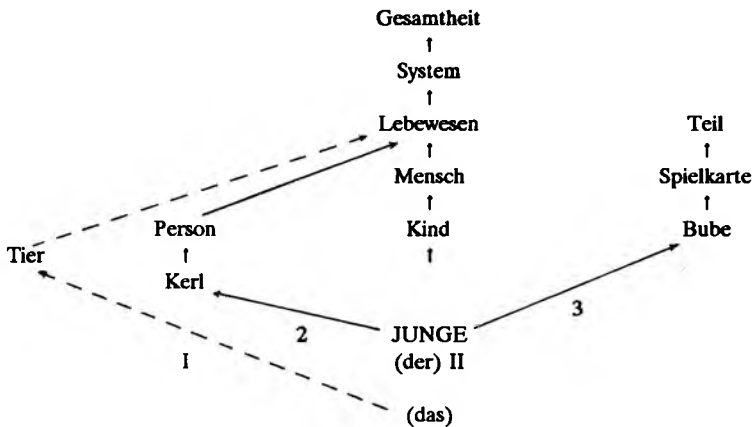
Badając gniazda ze względu na wymienione cechy liczbowe, autor otrzymał bardzo interesujące rozkłady empiryczne tych cech, dla nas jednak ważniejszy jest w tym omówieniu sam materiał, jaki stanowił zbiór owych 100 gniazd, gdyż kształt wielu z nich prowokował wręcz do formułowania pytań ważnych dla leksykologii.

Problemem numer 1 stały się leksemy homonimiczne, które Hammerl uwzględnił także w analizie. Klasyczny przykład gniazd dla homonimów stanowią struktury znaczeń leksemów *Kapelle* i *Hermelin*, traktowanych w HDG jako homonimy (*Kapelle*: I. 'orkiestra', II. 'kaplica'; *Hermelin*: I. 'gronostaj', II. 'futro').



Jak widać, dla każdego z homonimów istnieje odrębny ciąg definicyjny, zatem ciągi te nie mają żadnych leksemów nadrzędnych wspólnych (nie schodzą się one na wyższym poziomie abstrakcji). Taka struktura gniazda leksykalnego jest zgodna z określeniem homonimii, przyjętym przez wielu leksykologów (homonimy – leksemy równokształtne, które nie mają wspólnych nietrywialnych – typu 'rzecz', przyp. J. S. – części znaczenia).

Można byłoby zatem sądzić, że znaczeniom leksemów homonimicznych powinny odpowiadać całkowicie rozbieżne ciągi definicyjne w gnieździe. Okazało się jednak, że w materiale znalazły się leksemy, klasyfikowane w HDG jako homonimy, których gniazda miały budowę nie odpowiadającą powyższej regule. Przykładem może być gniazdo dla leksemu *Junge* (linią przerywaną zaznaczono homonim I, linią ciągłą – homonim II):

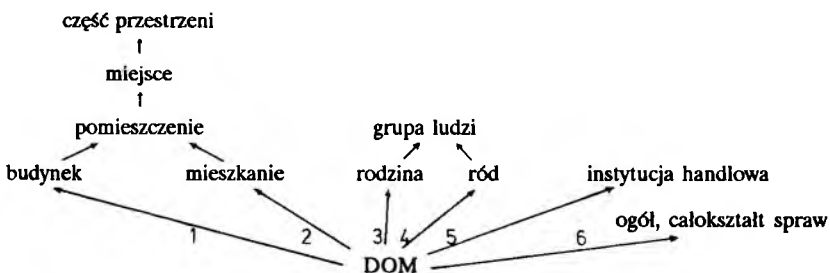


Widać wyraźnie, że ciąg dla znaczenia I, uznanego za homonimiczne (por. polskie *młode* w zn. 'zwierzę niedorosłe'), jest zbieżny z ciągiem dla znaczeń 1. i 2. homonimu II (leksemem wspólnym jest *Lebewesen*), natomiast odrębny ciąg odpowiada znaczeniu 3., traktowanemu w tym opisie jako składowa znaczenia homonimu II, obejmującego znaczenia 1., 2. i 3.

Takich kłopotliwych przykładów było w badanym materiale więcej, co świadczy o tym, że gniazda leksykalne nie dostarczają bezspornego kryterium do rozróżnienia homonimii od polisemii, ale dla jasności obrazu trzeba powiedzieć, że dla zdecydowanej większości przebadanych homonimów ciągi dla znaczeń homonimicznych były niezbieżne z innymi ciągami w danym gnieździe leksykalnym.

Z przytoczonych przykładów można wnosić, iż gniazda leksykalne dzięki swojej strukturze hierarchicznej ujawniają stopień wzajemnej "odległości" poszczególnych znaczeń leksemu polisemicznego. Np. w gnieździe leksemu *Maschine* znaczeniom 2. i 3. odpowiadają ciągi zbieżne z hiperonimem *Fahrrad* 'pojazd', zaś znaczenie 4. 'przezwisko kobiety' jako odległe od pozostałych nie ma wspólnego hiperonimu.

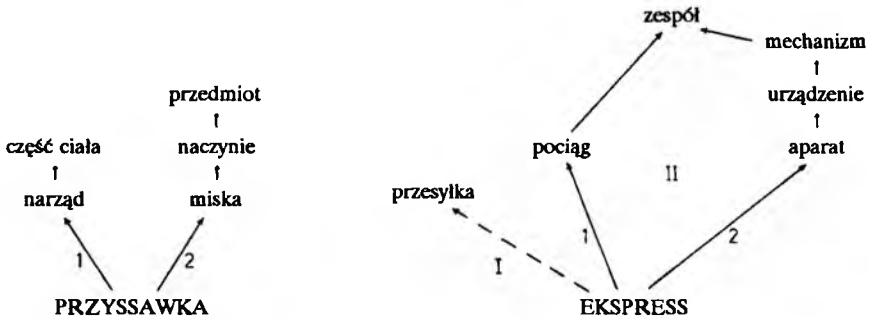
Podobną ilustracją może być ułożone przez nas gniazdo leksykalne dla polskiego leksemu *dom* (opis jego znaczeń, nieco zmodyfikowany, według MSJP):



Ujawniają się tu wyraźnie jako bliskie znaczenia 3. i 4., zbieżne w hiperonimie *grupa ludzi*, natomiast znaczenia 5. i 6. są wzajemnie dość odległe, nie

mając wspólnego hiperonimu (znaczenie 6. to 'ogół spraw związanych z domem', np. w wyrażeniu *mieć cały dom na głowie*).

Powstało pytanie, czy rozbieżność gałęzi gniazda leksykalnego świadczy zawsze o całkowitej rozbieżności jego znaczeń. Przebadano w tym celu kilka zbiorów gniazd leksykalnych, ułożonych dla rzeczowników polskich, wylosowanych ze słowników MSJP i SJP (Gołębiowska, Lewczuk 1992). Okazało się, że w zasadzie kształt gniazda leksykalnego nie może stanowić kryterium rozstrzygającego o stopniu bliskości znaczeń leksemu, a to dlatego, że o bliskości tej decydują nie tylko same hiperonimy jako odpowiadające w definicjach dwudzielnych członowi *genus proximum*, ale również cechy różnicujące, których w strukturze gniazda się nie uwzględnia. Ilustracją mogą być dwa gniazda, ułożone na podstawie SJP dla rzeczowników *przyssawka* i *ekspres*, przy czym drugi przykład potraktowany jest w SJP jako homonim (znaczenia oznaczono kreską przerywaną):

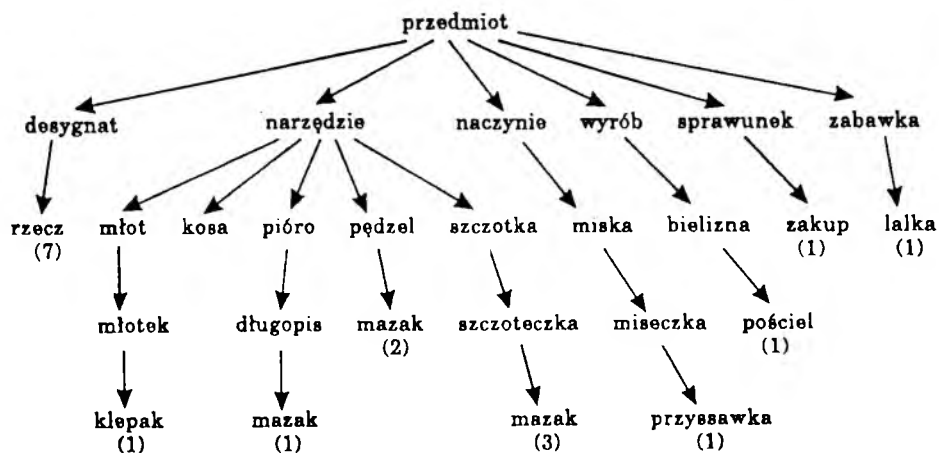


Okazuje się, że w obu wypadkach znaczenia są zbieżne mimo rozbieżności ścieżek grafu, a to ze względu na obecność w nich wspólnej definicyjnej cechy różnicującej: w znaczeniu *przyssawki* chodziło o narząd lub miskę 'przysysające się do podłoża', zaś wszystkie trzy znaczenia (także homonimiczne) leksemu *ekspres* mają wspólny składnik 'szybko' ('nadawana przesyłka', 'jadący pociąg', 'parzący kawę aparat'; Gołębiowska, Lewczuk 1992).

Kolejnym krokiem przy badaniu gniazd leksykalnych było odkrycie przez R. Hammerla tzw. gniazd pojęciowych. O ile gniazdo leksykalne ujmuje hierarchiczną strukturę hiperonimów, odpowiadających znaczeniom danego konkretnego leksemu wieloznacznego, o tyle punktem wyjścia przy budowie gniazda pojęciowego jest określony hiperonim końcowy, a więc o znaczeniu maksymalnie ogólnym, któremu przyporządkowuje się hierarchiczną strukturę podległych mu hiponimów.

Gniazdo pojęciowe zawiera zawsze zbiór ciągów definicyjnych ułożonych niejako "odgórnie", i powstaje na podstawie ułożonego wcześniej zbioru gniazd leksykalnych. Na przykład dla wylosowanego z SJP zbioru 100 rzeczowników ułożono 100 gniazd leksykalnych (opracowanych przez E. Gołębiowską i J. Lewczuk w cytowanej pracy). W gniazdach tych wystąpiło 12 znaczeń, odpowiadają-

cych rzeczownikom konkretnym, które to znaczenia miały ścieżki w gniazdach zakończone hiperonimem *przedmiot*. Przyjąwszy za punkt wyjścia ów hiperonim, przyporządkowano mu owe ścieżki, tworząc gniazdo pojęciowe dla hiperonimu *przedmiot*, które ujmuje hierarchiczną strukturę jego hiponimów (liczby w nawiasie oznaczają numery znaczeń danego leksemu w badanym słowniku):



Powyższe gniazdo opracowano na podstawie małej próby, liczącej tylko 100 gniazd leksykalnych, ale nawet w tak niewielkim zbiorze ujawniają się już hiponimy na pierwszym (licząc od góry) poziomie, które wskazują, jakie to kategorie znaczeniowe są definiowane w SJP jako przedmioty: należą do nich m. in. sprzęty (np. meble), narzędzia, wyroby, naczynia, zabawki.

Gniazda pojęciowe to na razie zaledwie pomysł – ale pozwolą one, być może, opisać explicite hierarchiczną strukturę pojęć, ujawniającą się w słownictwie danego konkretnego słownika, co w przyszłości będzie można wykorzystać do badań porównawczych.

Jak widać, gniazda leksykalne i pojęciowe nasuwają cały szereg problemów o dużych perspektywach badawczych. Przeszkodą jest ogromna czasochłonność układania takich gniazd, jednakże dzięki rozwojowi techniki komputerowej będzie wkrótce możliwe automatyczne układanie ciągów definicyjnych na podstawie danego słownika, co z kolei znacznie przyspieszy układanie gniazd. (Hammerl, Rogalińska 1988). Semantyka leksykalna została tym samym wzbogacona o nowy problem, jakim są gniazda leksykalne i pojęciowe; ich badanie może stanowić początek nowej drogi poszukiwań leksykalno-semantycznych.

8. Na zakończenie przedstawionego przeglądu prac i problemów z zakresu najnowszej semantyki kwantytatywnej chcemy wskazać na dalsze perspektywy badawcze tego nurtu. Wspominano już o nich szczegółowo w paragrafach dotyczących prawa Kryłowa i prawa Beöthy. Praca G. Altmanna, poświęcona dywersyfikacji semantycznej wyrazu, otwiera niezwykle szerokie horyzonty dla dalszych badań, autor rozważa bowiem różnicowanie się znaczeń wyrazu także w odniesieniu do jego asocjacji leksykalnych (znany jest rozkład rangowy wyrazów-asocjatyw, tworzących tzw. asocjacje powszechne, czyli najczęściej występujące dla danego wyrazu-bodźca (por. Kurcz 1976, 184-198).

Innym procesem kwantytatywnym, odwrotnym do procesu powstawania poli-semii, jest tworzenie się synonimów. Zarówno rozkłady rangowe asocjacji leksykalnych, jak i synonimów zdają się sprawdzać prawo Beöthy, jak wskazują na to wyniki nielicznych dotąd prac (Altmann 1991).

Na zakończenie wspomnijmy o sformułowanym w 1986 r. w Bochum międzynarodowym projekcie badań kwantytatywnych leksyki, którego autorami byli G. Altmann i H. Köhler. Ofertę współpracy przyjęło kilkanaście zespołów lingwistów z różnych krajów; spośród języków słowiańskich analizą objęto polski i słowacki, dane dla polszczyzny zostały już opracowane (por. Hammerl 1991, opis polskiej wersji projektu Köhlera – Altmanna w pracach Sambor 1989 a, b).

Dysponując wynikami liczbowymi dotyczącymi różnych języków, będziemy mogli przede wszystkim sprawdzić stopień uniwersalności prawa Kryłowa. Wkrótce – dzięki automatycznemu układaniu ciągów definicyjnych – będzie także możliwe otrzymanie bogatszych materiałów dotyczących hiponimii leksykalnej. Mankamentem tych wszystkich poszukiwań semantyczno-kwantytatywnych jest przyjmowanie jako jednostki badania tego znaczenia leksemu, które jest wyodrębniane w słowniku poprzez intuicję jego autorów. Jest to na pewno kryterium zawodne, czego dowiodły badania ciągów definicyjnych.

Ufamy, że lektura powyższego tekstu pozwoliła Czytelnikowi nabrać przekonania, iż m i m o braku precyzji w wyodrębnianiu znaczeń poszczególnych leksemów przy badaniu dużych zbiorów ujawniają się wyraźnie tendencje statystyczne, dotyczące struktury semantycznej leksykonu. Należy sądzić – wskazują na to nowe poszukiwania G. Altmanna i jego zespołu – że problem dywersyfikacji semantycznej będzie w teoretycznej semantyce kwantytatywnej problemem numer 1 przez najbliższe lata.

Literatura

- A l t m a n n G.: Semantische Diversifikation. "Folia Linguistica", 19:1985 z. 1-2 s. 177-200.

- A l t m a n n G.: Diversifikationprozesse des Wortes. W: Studies in language synergetics Ed. R. Köhler Bochum 1991.
- A l t m a n n G., B e s t K.-H., K i n d B.: Eine Verallgemeinerung des Gesetzes der semantischen Diversifikation. W: Glottometrika 8. Hrsg. I. Fickermann. Bochum. 1987 s. 130-139.
- A l t m a n n G., K i n d B.: Ein semantisches Gesetz. W: Glottometrika 5. Hrsg. R. Köhler Bochum 1983 s. 1-13.
- B e ő t h y E., A l t m a n n G.: The diversifikation of meaning of Hungarian verbal prefixes. I. *me g-* (forthcoming). 1984 a.
- B e ő t h y E., A l t m a n n G.: The diversifikation of meaning of Hungarian verbal prefixes. II. *ki-*. Finnish-Ugrische Mitteilungen 8. 1984 b s. 29-37.
- B e ő t h y E., A l t m a n n G.: Semantic Diversifikation of Hungarian verbal Prefixes. III. *föl-, el-, be-* W: Glottometrika 7. Ed. U. Rothe Bochum 1984 c s. 45-56.
- DW: Deutsche Wortbildung. Typen und Tendenzen in der Gegenwartssprache. 3 Bd. Düsseldorf 1973, 1975, 1978.
- G o l a n o w s k a M.: Funkcje przyimka w w różnych odmianach współczesnej polszczyzny. W: Studia semantyczne. Pod red. R. Grzegorzycykowej i Z. Zaron. Warszawa 1992. (w druku).
- G o ł ę b i e w s k a E., L e w c z u k J.: Stopień bliskości znaczeń w leksemach polisemicznych i w homonimach na materiale gniazd leksykalnych. W: Studia semantyczne. Pod red. R. Grzegorzycykowej i Z. Zaron. Warszawa 1992 (w druku).
- HDG: Handwörterbuch der deutschen Gegenwartssprache, Akademie-Verlag, Berlin 1984.
- H a m m e r l R.: Untersuchungen zur mathematischen Beschreibung des Martingesetzes der Abstraktionsebenen. W: Glottometrika 8. Hrsg. I. Fickermann. Bochum 1987 a s. 113-129.
- H a m m e r l R.: Neue modelltheoretische Untersuchungen im Zusammenhang mit dem Martingesetz der Abstraktionsebenen. W: Glottometrika 9. Hrsg. K.-P. Schultz. Bochum 1987 b s. 105-119.
- H a m m e r l R.: Prawa językowe we współczesnej kwantytatywnej lingwistyce modelowej – na przykładzie tzw. prawa Martina "Poradnik Językowy" 1987 c z. 6 s. 414-428.
- H a m m e r l R.: Cztery etapy rozwoju lingwistyki kwantytatywnej W: Wokół słownika języka polskiego II. Wrocław-Warszawa 1989 a s. 116-126.
- H a m m e r l R.: Untersuchung struktureller Eigenschaften von Begriffsnetzen. W: Glottometrika 10. Hrsg. R. Hammerl. Bochum. 1989 b, s. 141-154. Przekł. polski: Badanie strukturalnych własności gniazd leksykalnych. W: Stu-

- dia semantyczne. Pod red. R. Grzegorzyczkowej i Z. Zaron. Warszawa 1992 (w druku).
- H a m m e r l R.: Überprüfung des Martingeseetzes an deutschem Sprachmaterial. W: S a m b o r J., H a m m e r l R.: Definitionsfolgen und Lexemnetze. Lüchenscheid 1991.
- H a m m e r l R.: Gniazda leksykalne. Przegląd problemów. "Poradnik Językowy" 1990 s. 9-10.
- H a m m e r l R.: Untersuchung zur Struktur der Lexik: Aufbau eines lexikalischen Basismodells, Trier 1991 c.
- H a m m e r l R., R o g a l i ń s k a A.: Erstellung und analyse von Definitionsfolgen und lexemnetzen mit dem PC. Programm NETZE W: S a m b o r J., H a m m e r l R.: Definitionsfolgen und Lexemnetze. Lüdenscheid 1991 s. 157-173.
- H a m m e r l R., S a m b o r J.: Statystyka dla językoznawców, Warszawa 1990.
- H a m m e r l R., S a m b o r J.: O statystycznych prawach językowych, Warszawa 1991.
- K o m u d a M. Analiza semantyczna 300 ciągów definicyjnych języka serbochorwackiego. Warszawa 1987 (praca magisterska mps).
- K ö h l e r R. Zur sprachlichen Synergetik: Struktur und Dynamik der Lexik, Bochum 1986.
- K ö h l e r R., A l t m a n n G.: Synergetic modelling of language phenomena. W: Studies in language synergetics. Ed. R. Köhler. Bochum 1991 (w druku).
- K r y l o v Ju. K.: Eine Untersuchung statistischer Gesetzmässigkeiten auf der paradigmatischen Ebene der Lexik natürlicher Sprachen. W: Studies on Zipf's Law Hrsg. H. Guiter, A. Arapov. Bochum 1982 s. 234-262.
- K r y l o v Ju. K.: Ob odnom paradigme lingvostatističeskich raspredelenij. W: Lingvostatistika i vyčislitel'naja lingvistika. Wyd. J. Soontak. Tartu 1984 s. 80-97.
- K u r c z I.: Psycholingwistyka. Przegląd problemów badawczych. Warszawa 1976.
- MSJP: Mały słownik języka polskiego. Pod red. Skorupki S., Auderskiej, H. Łempickiej Z. Warszawa 1968.
- M a r t i n R.: Syntaxe de la definition lexicographique: étude quantitative des définissants dans le "Dictionnaire fondamental de la langue française". W: Statistique et linguistique. Ed. J. David, R. Martin. Paris 1974 s. 61-71.
- O Ź e g o v S. J.: Slovar' ruskogo jazyka, Moskva 1963.
- P a p p F.: Bearbeitung des ungarischen Wortschatzes auf Lochkartenmaschine. Acta Linguistica Academiae Scientiarum Hungaricae 17. Budapest 1967.

- R o t h e U. Die Semantik des textuellen *et*. Frankfurt a. M. 1986.
- SJP: Słownik języka polskiego. T. 1-3. Wyd. M. Szymczak Warszawa 1978-1981.
- SPP: Słownik poprawnej polszczyzny. Wyd. W. Doroszewski Warszawa 1973.
- S a m b o r J.: Słowa i liczby. Wrocław-Warszawa 1972.
- S a m b o r J.: O budowie tzw. ciągów definicyjnych (na materiale definicji leksykalnych), Biuletyn PTJ 40:1986 s. 151-165.
- S a m b o r J.: Lingwistyka kwantytatywna – stan badań i perspektywy rozwoju. Biuletyn PTJ 41:1988 s. 47-67.
- S a m b o r J.: Polnische Version des Forschungsprojektes "Sprachliche Synergetik. Teil 1. "Quantitative Lexikologie". W: Glottometrika 10. Hrsg. R. Hammerl. Bochum 1989 a s. 171-197.
- S a m b o r J.: O nowym projekcie badań statystycznej struktury słownictwa. W: Wokół słownika współczesnego języka polskiego II. Wrocław-Warszawa 1989 b s. 97-113.
- S a m b o r J.: Znaczenia wyrazów polisemicznych w słownikach a rozkład częstości tych znaczeń w tekście (na przykładzie polskiego przyimka *w*). W: Księga ku czci Prof. dra Władysława Kuraszkiewicza. Poznań 1989 c. (w druku).
- S a m b o r J. Struktura statystyczna wyrazów polisemicznych w słowniku, czyli o tzw. prawie Kryłowa (na materiale języka polskiego i rosyjskiego). "Polonica" 14:1989 d s. 13-31.
- S a m b o r J.: Struktura kwantytatywna wielofunkcyjnego morfemu gramatycznego w tekście (na przykładzie polskiego przyimka *w*.) W: Prace z językoznawstwa ogólnego i kontrastywnego II. Pod red. Sambor J., Huszcza R. Warszawa 1991 s. 89-99.
- S a m b o r J., H a m m e r l R.: Definitionsfolgen und Lexemnetze I. Lüdenschied 1991.
- W i e r z b i c k a A.: Semantic Primitives. Frankfurt 1972.
- Z i p f G. K.: The Psychobiology of Language. Cambridge 1935.
- Z i p f G. K.: Human Behavior and the Principle of Least Effort. Cambridge 1949.

ON THE PROBLEMS OF CONTEMPORARY QUANTITATIVE SEMANTICS

S u m m a r y

The use of computers in contemporary computer linguistics not only made it possible to work out statistically large bodies of texts or lexicons, but also permitted to shift from the stage of material

research to theoretical investigations. Zipf's works gave momentum to these researches. He worked on the relation between frequency and significance, frequency and the length of expressions and etc. All those researches were concerned with the plan of expression, i.e. the linguistic form.

Recently, there have been discovered new statistic regularities which were observed in lexical semantics. This discovery is a complete novelty in quantitative linguistics whose research have not dealt up to now with the semantic level of a language. Kryłow's discovery marked a breakthrough. He proved in an empirical manner that statistic semantic regularities are revealed exclusively thanks to the operation of the law of large numbers.

Three statistic-semantic regularities are the subject-matter of the paper. The first, defined as Kryłow's law, occurs in lexicons; it is a relation between the number of meanings of a lexeme (x) and the contribution of lexemes to a dictionary (W_x) which lexemes bear x meanings. It is an inversely proportional relation: unequivocal lexemes make the largest number of entries in a dictionary, as the degree of polysemy increases the contribution of lexemes decreases. In the paper Kryłow's law has been tested comparatively for given dictionaries of Polish, Russian and English languages, taking into account: a) the whole lexicon, b) nouns and verbs.

The second regularity concerns vocabulary in a text. It is a distribution of meanings in a text of a polysemic word. This distribution reveals itself while investigating into the uses of the words which are highly polysemic, e.g. functional expressions (prepositions conjunctions). Their meaning, grouped in a sequence according to their frequency in the text under investigation, reveals particular regularity called Beothy's law. The paper presents few works devoted to this problem and discusses distributions of the grammatical function of the Polish preposition "w" (in) which was investigated in a passage from a book "Droga wiodła przez Narwik" (The Route Went via Narvik) by K. Pruszyński. The way in which these functions were described in contemporary Polish dictionaries is far from being perfect, that is why A. Weinsberg's test of questions has been taken as the basis for their isolation. Employing this test, the frequencies of the functions of the grammatical preposition "w" (in) have been investigated in three types of contemporary texts (i.e. realistic novel, popularized scientific text on astronomy and causeries). It turned out that the sequence of those functions depends in fact on the kind of text we deal with.

The third problem discussed in the paper is the noun definition sequence of the type: *ammonia* – *gas* – *substance*, made of terms of the genus proximum type on the basis of lexical definitions. In large sets which encounter e.g. 1000 sequences, made for 1000 lexemes randomised from a dictionary, there are revealed also some statistic regularities concerning the length of those sequences, number of hiperonims at particular levels of sequences etc.

The final hiperonims in sequences turned out to be extremely interesting. Their analysis is connected with a philosophical problem of searching after semantic indefinibilia.

In the case of forming sequences for all meanings of one polysemic lexem of the type "dom" – house (1. "building" 2. "home" 3. "family" 4. "house" 5. "trade institution" 6. "the general body of affairs..." – data taken from "The Small Dictionary of the Polish Inaguage" ed. by S. Skorupka et al., Warsaw 1968) there is created the so-called family of words which illustrates the whole hierarchical structure of all hiperonims for the lexeme "house".

Investigation on the family of words was introduced by R. Hammerl on the basis of the German language. Then taking a particular final hiperonim as the starting point in a larger set of family of words which hiperonim was of the type "object" one can form on the basis of all sequences, ended with this hiperonim, the so-called conceptual family seat which grasps the hierarchical structure of all hiponims subordinate to it.

The families of words and concepts make in quantitative semantics a completely new matter for investigation which deals with the hierarchical structure of concepts grasped in the dictionaries of a given language.