

I. ROZPRAWY I ANALIZY

DOI: 10.17951/et.2016.28.117

Izabela G a t k o w s k a
(Kraków, UJ)

DOM W EMPIRYCZNYCH SIECIACH LEKSYKALNYCH

W artykule wyjaśniono, czym są leksykalne sieci empiryczne. Porównano znaczenie wyrazu *dom* w dwóch sieciach: sieci korpusowej Wortschatz (Uniwersytet w Lipsku) i sieci eksperymentalnej Katedry Lingwistyki Komputerowej UJ. Obydwie sieci jako najistotniejsze znaczenia, czyli najmocniej powiązane z wyrazem *dom*, wskazują: ‘dom – miejsce rodziny’ (*dom – rodzinny*) oraz ‘dom – miejsce do mieszkania’ (*mieszkanie*). Obie sieci wyróżniają też znaczenia ‘dom – budynek’. Znaczenie ‘dom – instytucja’ jest lepiej reprezentowane w sieci korpusowej. Sieć eksperymentalna wyróżnia dodatkowo znaczenie ‘dom – moje miejsce schronienia’ (*schronienie, azyl, twierdza*). Sieć eksperymentalna w wyniku tego, że występujące w niej powiązania mają kierunek, organizuje podsieci, wyjaśniające znaczenia, np. podsieć znaczenia ‘dom – miejsce rodziny’ (*dom rodzinny*), wykazuje szczególną rolę ‘matki’, która organizuje podsieć reprezentującą to znaczenie. Porównanie znaczeń wyróżnianych i wyjaśnianych przez podsieci ze słownikowymi definicjami znaczeń pozwala przyjąć, że badania nad sieciową reprezentacją znaczeń mogą być przydatne w leksykografii.

SŁOWA KLUCZOWE: lingwistyka, sieć empiryczna, sieć korpusowa, semantyka leksykalna, sieciowa reprezentacja znaczenia

Pomysł reprezentowania leksykonu języka naturalnego w formie sieci leksykalnej narodził się w latach 70. XX wieku w trakcie badań psychologicznych (Kiss, Armstrong, Millroy, Piper 1973) i prowadzonych równolegle badań nad sztuczną inteligencją (Schank 1975; Minsky 1975). Sieć to struktura zbudowana z jednostek leksykalnych i relacji zachodzących pomiędzy jednostkami (Minsky 1975). Przyjmuje się, że relacje (powiązania), w jakie

wchodzi konkretna jednostka leksykalna z innymi jednostkami leksykalnymi w sieci reprezentują zależności pomiędzy znaczeniami jednostek wchodzących w powiązanie (Clark 1970).

1. Sieci empiryczne

Empiryczną sieć leksykalną można zbudować za pomocą eksperymentu prowadzonego na odpowiednio licznej grupie osób, które podają w odpowiedzi pierwszy wyraz, jaki przychodzi na myśl po zrozumieniu wyrazu bodźca (Kent–Rosanoff 1910) lub za pomocą algorytmu zliczającego i interpretującego współwystąpienia wyrazów w odpowiednio dużym zbiorze tekstów (Biemann 2007). Za pomocą obu metod uzyskujemy empirycznie umotywowany zbiór par: jednostka definiowana – jednostka definiująca, przyjmujemy bowiem, że znaczenia obu jednostek leksykalnych wchodzi w powiązanie, którego siłę określa liczba współwystąpień obu członów pary w zbiorze wszystkich odpowiedzi udzielonych dla jednostki definiowanej lub zbiorze wszystkich zdań zawierających jednostkę definiowaną. Postaramy się pokazać, że występująca w obu sieciach siła powiązania zachodzącego pomiędzy jednostką definiowaną a jednostką definiującą pozwala wyróżnić znaczenia jednostek definiowanych, a także wskazuje na hierarchię znaczeń w słowniku – funkcje te pełnią tzw. powiązania najsilniejsze.

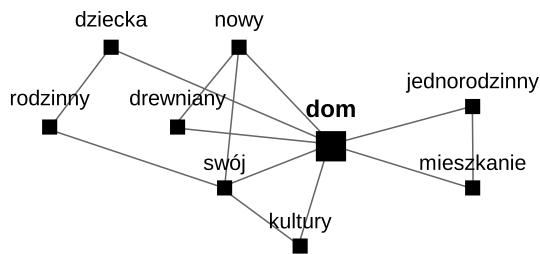
Zbiór par, tj. powiązań zachodzących pomiędzy jednostkami leksykalnymi (węzłami sieci), tworzy sieć leksykalną. Sieć zbudowana eksperymentalnie ma pewne właściwości formalne, których brak w sieci generowanej z korpusu tekstów: powiązania w sieci eksperymentalnej mają kierunek, zawsze od bodźca do odpowiedzi. Analiza kierunku powiązań w sieci eksperymentalnej pozwala zaobserwować, że sieć ta zawiera podsieci wyjaśniające znaczenia wyróżniane za pomocą siły powiązania.

Sposób definiowania znaczenia w sieciach empirycznych pokażemy na przykładzie znaczeń jednostki leksykalnej *dom*. Na koniec porównamy znaczenia wyróżnione w sieci ze znaczeniami wyróżnianymi przez słownik budowany zgodnie z zasadami ukształtowanymi przez wielowiekową tradycję leksykograficzną.

1.1. Sieć leksykalna budowana z korpusu tekstów

Do rozważań o sieciowym ujęciu znaczenia wyrazu *dom* wybraliśmy polskojęzyczne korpusy tekstów i algorytmy generujące sieć leksykalną zbudowane w projekcie Wortchatz realizowanym od kilkunastu lat na uniwersytecie w Lipsku (Biemann i inni 2011). Wortschatz ma pięć korpusów polskojęzycznych, tj. korpusy wiadomości prasowych z lat 2007, 2008 i 2011 oraz

korpusy ekstrahowane z Wikipedii z 2007 i 2010 roku. Do generowania sieci leksykalnej użyliśmy korpusu wiadomości prasowych z 2011 roku, zawierającego 6 494 575 zdań i liczącego 96 476 260 form fleksyjnych, reprezentujących w tekście jednostki leksykalne języka polskiego. Algorytm generujący sieć leksykalną działa w dwu fazach. Najpierw – operując na zdaniach – tworzy pary: wyraz definiowany – wyraz współwystępujący, np. *dom* – *rodzinny* i oblicza częstość występowania danej pary (częstość współwystępowania). Algorytm generujący pary zakłada, że wyrazem współwystępującym jest tylko taki wyraz, który łączy się bezpośrednio w zdaniu z wyrazem definiowanym. Wyraz (jednostka leksykalna) *dom* występuje w korpusie 8897 razy, a lista wyrazów współwystępujących odpowiednio często z wyrazem *dom* liczy 814 pozycji – po 407 wyrazów, wstępujących w zdaniu po lewej i prawej strony wyrazu *dom*. Istnieją wyrazy, które występują jednocześnie po prawej i po lewej stronie wyrazu definiowanego, co sprawia, że ich powiązanie z wyrazem definiowanym nie ma kierunku. Z wyrazem *dom* łączy się w ten sposób 59 wyrazów – będziemy je nazywać definiującymi. Najczęstsze z nich to: *rodzinny*, *mieszkanie*, *swój*, *nowy*, *drewniany*. Algorytm generujący sieć leksykalną wyrazu definiowanego oprócz wyrazów definiujących bierze także pod uwagę wyrazy występujące jednocześnie po obu stronach wyrazów definiujących. W konsekwencji sieć leksykalna zbudowana wokół wyrazu *dom* liczy kilkaset wyrazów (węzłów), między którymi zachodzi kilkakrotnie większa liczba powiązań (zależności). Graf reprezentujący sieć o takiej strukturze byłby nieczytelny. Dlatego dla ilustracji pokażemy sieć zbudowaną wyłącznie z wyrazów najczęściej współwystępujących z wyrazem *dom* (rys. 1).



Rys. 1. *Dom* w sieci korpusowej

Dla potrzeb przedstawionego w dalszej części tekstu porównania powiązań sieci korpusowej z powiązaniem sieci eksperymentalnej musimy przekształcić wyrażaną liczbowo częstość współwystępowania w korpusie na tzw. siłę powiązania wyrażaną w procentach. Przekształcenia dokonujemy, dzieląc częstość współwystępowania konkretnego wyrazu definiującego przez

liczbę wszystkich wystąpień wyrazu definiowanego. Przykładowo, jeśli wyraz *rodziny* współwystępuje z wyrazem *dom* 1680 razy, a częstość wystąpienia wyrazu *dom* w całym korpusie wynosi 8897, to siła powiązania *dom* – *rodziny* równa się 18,19 procent (1680/8897 po zaokrągleniu). Możemy więc powiedzieć, że przedstawiona wyżej sieć jest zbudowana z wyrazów najmocniej powiązanych z wyrazem *dom*.

1.2. Sieć eksperymentalna KLK UJ

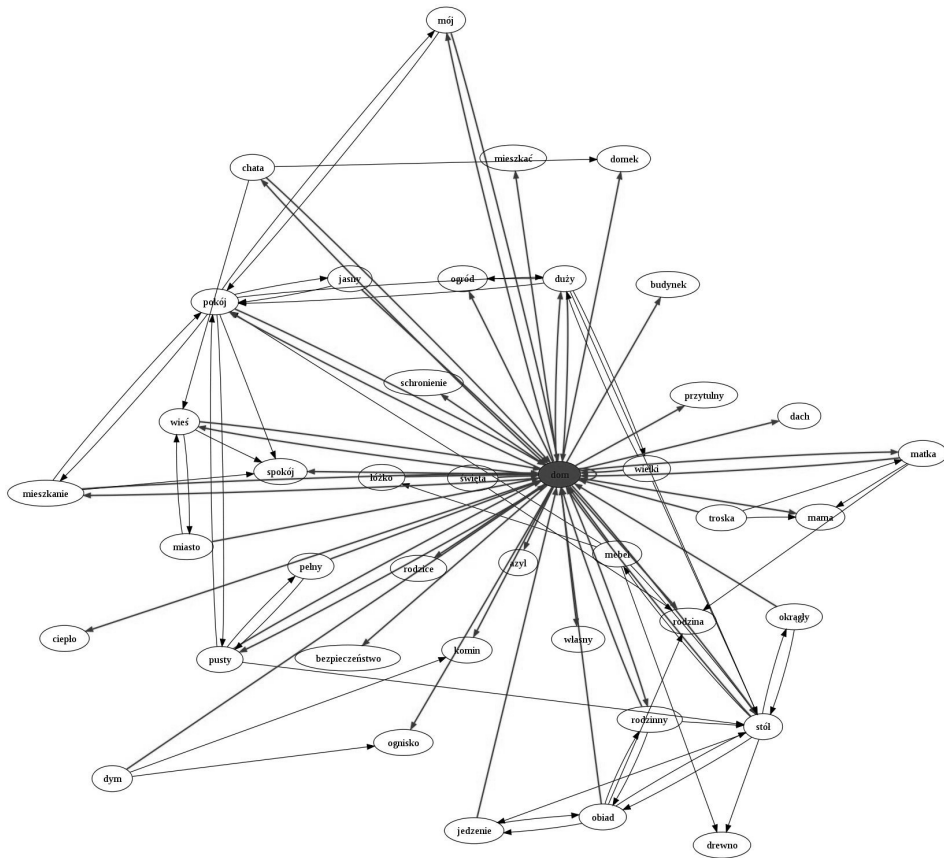
Eksperymentalna sieć powiązań leksykalnych jest wynikiem eksperymentu wykorzystującego ludzki mechanizm skojarzeniowy – badany podaje pierwszy wyraz, jaki przychodzi na myśl po zrozumieniu wyrazu bodźca. Jeśli przebadamy w ten sposób odpowiednią liczbę osób, uzyskamy listę wyrazów powiązanych z wyrazem bodźcem, mówiąc technicznie, listę par: bodziec – odpowiedź. Każda z par ma eksperymentalnie ustaloną siłę powiązania bodźca i odpowiedzi. Siłę powiązania wyraża liczba badanych, którzy udzielili konkretnej odpowiedzi. Zgodnie z tradycją opisywania wyników eksperymentu (Kiss, Armstrong, Millroy, Piper 1973) będziemy siłę powiązania wyrażać w procentach, dzieląc liczbę badanych, którzy udzielili konkretnej odpowiedzi przez liczbę wszystkich badanych. Przykładowo, gdy liczba wszystkich odpowiedzi dla bodźca A wynosi 95, z czego 49 przypada na odpowiedź B, to siła powiązania $A \rightarrow B$ wynosi 52 procent (49/95 po zaokrągleniu)¹.

Jeśli jednak rozszerzymy eksperyment i przeprowadzimy go cyklicznie², co oznacza, że wyrazy – odpowiedzi uzyskane w cyklu pierwszym zostaną użyte jako bodźce w cyklu drugim, to w wyniku uzyskamy sieć leksykalną o strukturze innej niż struktura sieci budowanej z korpusów tekstowych. Różnicę pokazuje graf reprezentujący eksperymentalną sieć zbudowaną wokół wyrazu *dom* (zob. rys. 2. Graf powstał w wyniku eksperymentu autorского opisanego w Dodatku).

Przyczyną różnic jest fakt, że wszystkie powiązania w sieci eksperymentalnej mają kierunek, zawsze od bodźca do odpowiedzi. W konsekwencji wyraz będący węzłem w sieci leksykalnej zbudowanej z par bodziec – odpowiedź może mieć powiązania wychodzące, np.: *dom* → *rodziny* (*dom* był bodźcem) oraz przychodzące np. *drewniany* → *dom* (*dom* był odpowiedzią). W rezultacie możemy wyróżniać w sieci węzły pełne, tj. takie wyrazy, które mają powiązania wychodzące i przychodzące (np. *dom*) oraz

¹ Tak zbudowana lista par była stosowana w psychiatrii jako wzorzec diagnostyczny (Kent–Rosanoff 1910).

² Jak pierwsi zastosowali ten mechanizm twórcy *The Edinburgh Associative Thesaurus* (Kiss i inni 1973).

Rys. 2. *Dom* w sieci eksperymentalnej

węzły zredukowane, tj. takie wyrazy, które mają tylko powiązania przychodzące (np.: *rodzina*). Możliwa jest także sytuacja, w której pomiędzy dwoma wyrazami (węzłami) sieci zachodzi powiązanie zwrotne, np. *dom* → *mieszkanie* i *mieszkanie* → *dom*. Powiązania zwrotne nie występują często, np. w eksperymentalnej sieci (Gatkowska 2014) na 50 849 wszystkich powiązań mamy tylko 1181 powiązań zwrotnych. Pokażemy, że powiązania zwrotne zazwyczaj służą do organizowania podsieci.

2. *Dom* – znaczenia wyróżniane przez najsilniejsze powiązania

Pokazaliśmy, że sieć korpusowa i sieć eksperymentalna mają różne struktury. Jednak to, że obie sieci są zbudowane z par wyrazów, którym została

przyporządkowana siła powiązania umożliwia częściowe porównanie sieci. W tabeli zestawiamy bezpośrednie powiązania wyrazu *dom* w sieci generowanej z korpusu, której powiązania nie mają kierunku (stąd pojedyncza kolumna korpus) z powiązaniem *domu* w sieci eksperymentalnej, gdzie wyróżniamy powiązania wychodzące i przychodzące. Wyrazy w tabeli są uporządkowane malejąco według siły powiązania z wyrazem *dom*. Lista wyrazów powiązanych z wyrazem *dom* w każdej z sieci liczy po kilkaset pozycji, więc do porównania wybraliśmy tylko wyrazy najmocniej powiązane.

Tabela 1.

Korpus		Eksperyment KLK			
		przychodzące		wychodzące	
rodzinny	18,19	rodzinny	16,86	17,67	rodzinny
mieszkanie	14,86	mieszkanie	15,25	12,25	mieszkanie
swój	10,44	rodzina	12,16	10,55	mój
nowy	9,28	spokój	2,98	7,49	drewniany
drewniany	8,74	ciepło	2,64	6,11	pusty
jednorodzinny	8,04	ogród	2,41	5,45	wielki
dziecka	7,78	mój	2,18	4,18	chata
Mój dom	7,74	dach	1,95	3,45	duży
kultury	6,92	mama	1,61	2,58	pokój
kw	6,87	bezpieczeństwo	1,49	2,18	wysoki
pow	5,84	duży	1,49	2,00	ulica
domu	5,68	pokój	1,38	1,64	piękny
którym	5,24	pusty	1,38	1,49	obiad
kupił	5,00	dom	1,15	1,42	mały
mój	4,90	zły	1,15	1,31	spokojny
mieszkalny	4,89	budynek	1,03	1,27	zły
własny	4,62	chata	0,92	1,24	biały
wart	4,48	własny	0,80	1,24	ładny
rodzinę	4,45	matka	0,69	1,24	mebel
m	4,35	mieszkać	0,57	1,16	dym
uderzył	4,00	rodzice	0,57	1,09	okno
dzieci	3,97	azyl	0,46	0,87	wieś
nasz	3,96	domek	0,46	0,87	wuja Toma
wybudować	3,84	komin	0,46	0,73	matka
samochód	3,79	ognisko	0,46	0,73	podłoga
spłonął	3,78	przytulny	0,46	0,69	pełny
uderzyła	3,73	schronienie	0,46	0,65	praca
wybudował	3,72	stół	0,46	0,51	ziemia
kupić	3,52	wieś	0,46	0,47	za wsią
towarowy	3,46	ciepły	0,34	0,40	święta
tys	3,45	daleko	0,34	0,36	dom
gdzie	3,42	własny	0,80	0,36	polski
TIR	3,42	dziecka	0,34	0,36	żelazny
stoi	3,34	miłość	0,34	0,33	kolorowy

Korpus		Eksperyment KLK			
		przychodzące		wychodzące	
działkę	3,32	otwarty	0,34	0,29	dywan
ogród	3,28	Tom	0,34	0,29	jasny
mody	3,23	wielki	0,34	0,29	łóżko
pod	3,14	cegła	0,23	0,29	miasto
wjechał	3,14	chałupa	0,23		
maklerski	2,99	chatka	0,23		
spalił	2,96	drewno	0,23		
Mój	2,85	drzwi	0,23		
ogrodem	2,80	jednorodzinny	0,23		
handlowy	2,60	miejsce	0,23		
gospodarstwo	2,45	obiad	0,23		
i	2,43	osiedle	0,23		
powierzchni	2,42	ostoja	0,23		
Stał	2,37	podwórko	0,23		
zbudować	2,29	powrót	0,23		
Aukcyjny	2,24	radość	0,23		
Rodzinny dom dziecka	2,24	stan	0,23		
zniszczył	2,20	szczęście	0,23		
Stary	2,18	ściany	0,23		
Ciężarówka	2,18	twierdza	0,23		
Ma	2,15				
Duży	2,15				
Kupili	2,11				
Zakonny	2,11				
Drewniany	2,10				

W tabeli obserwujemy uderzającą zbieżność: wyrazy *rodzinny* i *mieszkanie* są najsilniej powiązane z wyrazem *dom* w obu sieciach³. Dodatkowo, w sieci eksperymentalnej oba wyrazy wchodzą w powiązania zwrotne. Można więc przyjąć, że najsilniejsze powiązania wyróżniają dwa znaczenia wyrazu *dom*. Pierwsze to ‘dom rodzinny – miejsce rodziny’, które sygnalizują także powiązane słabiej wyrazy: *rodzinę*, *dzieci* w sieci korpusowej oraz wychodzące: *rodzina*, *mama*, *matka*, *rodzice* i przychodzące *matka* w sieci eksperymentalnej. Drugie z wyróżnionych znaczeń wyrazu *dom* to ‘miejsce do mieszkania; miejsce, w którym się mieszka’, które oprócz *mieszkanie* sygnalizują także słabiej powiązane: *kupił*, *własny*, *kupić*, *ma* w sieci korpusowej oraz w sieci eksperymentalnej wychodzące: *pokój*, *własny*, *miejsce*, *pusty* i przychodzące: *pusty*, *pokój*, *mebel*, *podłoga*, *dywan*, *łóżko*. Jedynym

³ W badaniach wykorzystujących ludzki mechanizm skojarzeniowy uzyskano podobne wyniki, por. Kurcz 1967 (1000 osób badanych): *dom*: *rodzinny* (48), *mieszkanie* (45), *mój* (45); Gawarkiewicz 2008 (500 osób badanych): *dom*: *rodzina* (164), *rodzinny* (11), *mieszkanie* (22).

czasownikiem, który występuje w sieci eksperymentalnej w węźle *dom*, jest czasownik *mieszkać* (to zbieżne z etnolinwistycznymi obserwacjami, Bartmiński 2015). Pozostałe znaczenia wyrazu *dom* nie są sygnalizowane poprzez pojedynczy wyraz mocno powiązany z wyrazem *dom*. Można je jednak zidentyfikować poprzez grupę wyrazów słabiej powiązanych. Uwagę zwraca pojawiający się w obu sieciach zaimek *mój*, który w sieci eksperymentalnej wchodzi z *domem* w powiązanie zwrotne. *Mój* można powiązać z występującymi tylko w sieci eksperymentalnej powiązaniem wychodzącymi *spokój*, *ciepło*, *bezpieczeństwo*, *własny*, *azyl*, *ostoja*, *schronienie*, *szczęście*, *twierdza* i *przyjąć*, że wyrazy te sygnalizują kolejne znaczenie wyrazu ‘dom – moje miejsce schronienia’. Kolejne znaczenie *dom*, tj. ‘dom – budynek’ jest reprezentowane w obu sieciach. W sieci korpusowej znaczenie to reprezentują przede wszystkim powiązane z *domem* czasowniki: *kupił*, *uderzył*, *wybudować*, *splonął*, *uderzyła*, *wybudował*, *kupić*, *stoi*, *wjechał*, *spalił*, *stał*, *zbudować*, *zniszczył*, *ma*, *kupili*, a także *drewniany*, *jednorodzinny*, *mieszkalny*, *samochód*, *TIR*, *działkę*, *ogród*, *pod*, *ogrodem*, *gospodarstwo*, *ciężarówka*. W sieci eksperymentalnej znaczenie ‘dom – budynek’ reprezentują powiązania wychodzące: *ogród*, *dach*, *budynek*, *domek*, *komin*, *wieś*, *cegła*, *chałupa*, *chatka*, *drewno*, *jednorodzinny*, *drzwi*, *okno*, *osiedle*, *podwórko* oraz przychodzące: *drewniany*, *wielki*, *dym*, *chata*, *wysoki*, *ulica*, *mały*, *wieś*, *za wsią*, *miasto*. Ostatnim ze znaczeń, które pojawia się w sieciach leksykalnych, jest ‘dom – instytucja’, które w sieci korpusowej reprezentują powiązania: *kultury*, *towarowy*, *mody*, *maklerski*, *rodzinny dom dziecka*, *zakonny*, a w sieci eksperymentalnej tylko powiązanie wychodzące: *dziecka*.

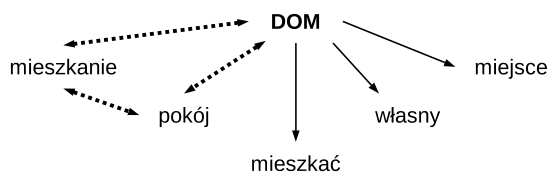
Wymienione znaczenia wyrazu *dom* wyróżniliśmy intuicyjnie. Uzasadnienie dla intuicji znajdziemy w strukturze sieci empirycznej, gdzie wyróżnione znaczenia tworzą wyraźne podsieci. Ponieważ elementy podsieci nie są widoczne w zredukowanym grafie *dom*, przedstawimy graf podsieci każdego ze znaczeń, odwołując się, gdy trzeba, do innych, powiązanych z omawianą strukturą sieci.

3. Podsieci znaczeń wyrazu *dom* w sieci eksperymentalnej

W sieci eksperymentalnej poszczególne znaczenia wyrazu *dom* przyjmują strukturę podsieci zorganizowanych w sposób naturalny wokół węzła *dom*. Szkieletem podsieci są zazwyczaj powiązane zwrotnie (wzajemne) pomiędzy węzłami. Przyjmujemy, że podsieci wyjaśniają poszczególne znaczenia wyrazu definiowanego, a siła powiązań tworzących podsieć pokazuje rangę poszczególnych znaczeń.

3.2. Znaczenie: ‘dom – miejsce do mieszkania’

Szkieletem podsięci reprezentującej znaczenie ‘dom – miejsce do mieszkania’ są powiązania zwrotne: *dom* ↔ *mieszkanie* (*dom* → *mieszkanie* 15,25; *mieszkanie* → *dom* 12,25), *dom* ↔ *pokój* (*dom* → *pokój* 1,38; *pokój* → *dom* 2,58); *mieszkanie* ↔ *pokój* (*mieszkanie* → *pokój* 3,09; *pokój* → *mieszkanie* 8,14) oraz powiązania wychodzące węzła *dom*: *dom* → *mieszkać* 0,57; *dom* → *miejsce* 0,23; *dom* → *własny* 0,8.

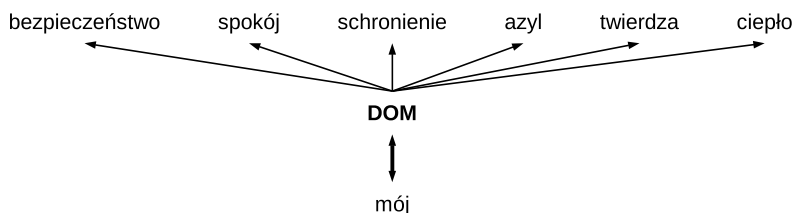


Rys. 4. Dom rodzinny – miejsce do mieszkania

Węzłem oddającym istotę znaczenia: ‘dom – miejsce do mieszkania’ jest węzeł *mieszkanie* wraz ze swoimi powiązaniem. ‘Miejsce do mieszkania’ ma cechy obiektu fizycznego, a więc – czego ze względów technicznych nie ma na rysunku – *mieszkanie* jest *własne, przestronne, wygodne, nowe, swoje, małe, ciasne ale własne, duże*, może się znajdować w *bloku*, ma też swój rodzaj: *apartament*, ma wreszcie cechę swoistą: *dom* ↔ *pusty* (*pusty*: *dom* → *pusty* 1,38; *pusty* → *dom* 6,11). Najistotniejszym elementem ‘miejsca do mieszkania’ jest *pokój*, co sygnalizuje powiązanie zwrotne: *mieszkanie* ↔ *pokój*, *dom* ↔ *pokój*. *Pokój* ma także cechy obiektu fizycznego wymienione wyżej.

3.3. Znaczenie: ‘dom – moje miejsce schronienia’

Znaczenie ‘dom – moje miejsce schronienia’ wyrażają powiązanie zwrotne: *dom* ↔ *mój* (*dom* → *mój* 2,18; *mój* → *dom* 10,55) oraz powiązania wychodzące z węzła *dom* do węzłów: *bezpieczeństwo* 1,49; *schronienie* 0,46; *azyl* 0,46; *ostoja* 0,23; *twierdza* 0,23.

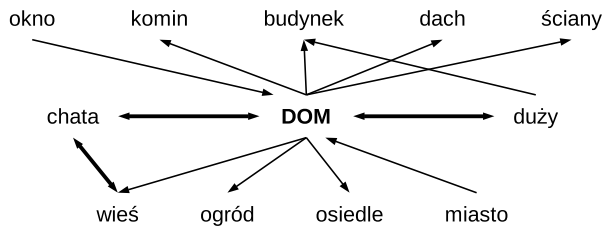


Rys. 5. Dom – moje miejsce schronienia

Znaczenie ‘dom – moje miejsce schronienia’ łączy się poprzez powiązanie zwrotne *pokój* ↔ *mój* (*mój* → *pokój* 3,3; *pokój* → *mój* 2,26) z węzłem *pokój* współtworzącym szkielet znaczenia podsieci reprezentującej znaczenie ‘dom – miejsce do mieszkania’.

3.4. Znaczenie: ‘dom – budynek’

Znaczenie ‘dom – budynek’ reprezentuje podsieć, w której brak powiązań zwrotnych tworzących szkielet znaczenia. Powiązania zwrotne pełnią tu takie same funkcje jak powiązania wychodzące i przychodzące, tj. tworzą podsieć. Powiązania węzła *dom* charakteryzują obiekt fizyczny, a więc składniki (części): *komin*, *dach*, *ściany* okno, nadrzędnik (hiperonim) *budynek*, współpodrzędnik (kohiponim) *chata*, rozmiar *duży*, *wielki*, i lokalizacje: *wieś*, *ogród*, *osiedle*, *miasto*.



Rys. 6. Dom – budynek

Powiązanie zwrotne: *duży*⁴ ↔ *dom* (*dom* → *duży* 1,49%; *duży* → *dom* 3,45%) wskazuje na tę cechę *domu*, która jest też przypisana *budynkowi* – powiązaniem przychodzącym *budynek* ← *duży* 1,13%.

Powiązania, które pominęliśmy na rysunku, jak np. *kościół*, *wysoki*, *cegła*, *drewno*, *drewniany*, tylko potwierdzają przedstawione wyżej obserwacje.

3.5. Znaczenie: ‘dom – instytucja’

Znaczenie ‘dom – instytucja’ jest stosunkowo dobrze reprezentowane w sieci korpusowej Wortschatz. W empirycznej sieci eksperymentalnej nie ma podsieci reprezentującej to znaczenie. Być może dlatego, że zestawienia, takie jak: *dom* – *dziecka*, *kultury*, *towarowy*, *handlowy*, *mody*, *aukcyjny*, *marklerski*, *zakonny* mają status terminów, tj. wyrażen, których znaczenie jest definiowane poza systemem językowym, np. za pomocą definicji prawnych.

⁴ Wyraz *duży* w eksperymencie należał do tych, który miał dużo pustych odpowiedzi. Problematyce pustych odpowiedzi poświęcony jest artykuł *Empty answers in an experiment of free word association* (Gatkowska 2015a).

4. *Dom* – znaczenia sieciowe a znaczenia wyróżniane przez słownik

*Słownik języka polskiego PWN*⁵ podaje 6 znaczeń wyrazu *dom* – przyjmujemy, że numeracja oznacza hierarchię znaczeń:

1. 'budynek przeznaczony na mieszkania, na pomieszczenia dla zakładów pracy, instytucji, itp.';
2. 'mieszkanie, pomieszczenie mieszkalne, miejsce stałego zamieszkania';
3. 'rodzina, domownicy; też: mieszkanie wraz z jego mieszkańcami';
4. 'ogół spraw rodzinnych, domowych; gospodarstwo domowe';
5. 'ród, rodzina, dynastia';
6. 'instytucja państwowa, społeczna handlowa, itp., mieszcząca się zwykle w oddzielnym lokalu lub budynku; budynek, w którym się ona znajduje'.

Przeprowadzona analiza podsieci objaśniających znaczenia wyrazu *dom* pozwala zauważyć, że sieć eksperymentalna inaczej buduje hierarchię znaczeń, np. znaczenie 'budynek' umieszczone w słowniku na pierwszym miejscu, w sieci zajmuje miejsce przedostatnie. Widać też wyraźne różnice w zestawie wyróżnianych znaczeń wyrazu *dom*. Wspólne są niewątpliwie znaczenia 1. 'budynek' i 2. 'mieszkanie'. Znaczenie 3. 'rodzina, domownicy'; jest w sieci organizowane przez 'matkę' i ma inną strukturę niż sugeruje definicja słownikowa 'też: mieszkanie wraz z jego mieszkańcami'. Natomiast wyróżniane przez słownik znaczenie 6. 'instytucja' nie ma podsieci objaśniającej znaczenie. Są jednak istotne różnice w zestawach wyróżnianych znaczeń. W sieci nie ma wyróżnianych przez słownik znaczeń: 4. 'ogół spraw rodzinnych' i 5. 'ród, dynastia'. I odwrotnie, słownik nie wyróżnia istotnego w sieci eksperymentalnej 'dom – moje miejsce schronienia'. Poprzestaniemy na samym wskazaniu różnic, bowiem próba ich wyjaśniania wykraczałaby zdecydowanie poza ramy krótkiego tekstu.

Podsumowanie

Przedstawiliśmy znaczenie wyrazu *dom* w sieciach empirycznych. Każda z tych sieci powstała w inny sposób. Jednak obie budują podobną hierarchię wyróżnianych znaczeń: 'dom – miejsce rodziny', 'dom – miejsce do mieszkania', 'dom – budynek', 'dom – instytucja'. Sieć budowana eksperymentalnie wnosi dodatkową jakość, gdyż zawiera podsieci wyjaśniające każde ze

⁵ *Słownik języka polskiego* pod red. M. Szymczaka (1978), t. 1, s. 422. Podobnie *dom* definiuje słownik internetowy SJP PWN on-line: <http://sjp.pwn.pl/sjp/dom;2453213.html> (data dostępu: 15.03.2016), wyróżniając dodatkowo: 'miejsce, z którego ktoś pochodzi'.

znaczeń. Porównanie sieciowego opisu znaczeń z definicjami umieszczonymi w słowniku pokazuje, że sieć eksperymentalna potrafi wykryć znaczenie nieobecne w słowniku. Można więc uznać, że badania znaczeń w empirycznych sieciach leksykalnych może być pomocne w leksykografii.

Trzeba przy tym dodać, że analiza znaczeń odwołująca się do strukturalnych właściwości sieci eksperymentalnej nie wyczerpuje opisu znaczenia w sieci. Wiadomo bowiem (Clark 1970), że zależność pomiędzy bodźcem a odpowiedzią jest zależnością semantyczną, np. powiązanie ‘dom’ → ‘dach’ można objaśnić jako powiązanie zachodzące pomiędzy całością a częścią (metonimia). Jednak analiza wszystkich zależności semantycznych zachodzących pomiędzy węzłami sieci wymagałaby odrębnego (Gatkowska 2015b) i znacznie obszerniejszego opracowania.

Dodatek.

Test swobodnych skojarzeń słownych jako narzędzie budowy sieci leksykalnej

1. Krótki szkic historii testu skojarzeniowego

Test skojarzeniowy jest metodą pozyskiwania zależności leksykalnych przechowywanych w ludzkim umyśle. Przebieg testu jest prosty – badany podaje pierwszy wyraz, jaki przyjdzie mu na myśl po zrozumieniu wyrazu bodźca przedstawionego przez badającego. Pierwszy raz do celów naukowych metodę skojarzeniową wykorzystano w 1910 roku. Para amerykańskich lekarzy psychiatrów – G. Kent i A. Rosanoff (1910) – opracowała i przeprowadziła test swobodnych skojarzeń słownych dla 100 wyrazów, w którym uczestniczyło 1000 zdrowych osób. Celem testu było zbudowanie wzorca diagnostycznego (*association norms*), z którym porównywano zestawy skojarzeń uzyskane od osób z zaburzeniami psychicznymi.

Badania Kent–Rosanoffa przeprowadzono ponownie, tym razem na grupie studentów amerykańskich (Schellenberg 1930, niepublikowana rozprawa doktorska, Uniwersytet Minnesota, za: Postman, Keppel 1970). Kolejny raz badania powtórzono po ponad dwudziestu latach – W. A. Russell i J. Jenkins (1952) – w obydwu przypadkach badano studentów uczelni amerykańskich. Badania powtórzone przyniosły wyniki zaskakująco zbieżne z wynikami tekstu oryginalnego, co prowadziło do wniosku, że powiązania leksykalne są stałe i niezależne od badanej grupy osób. W wyniku tej obserwacji w kolejnych badaniach zdecydowano się na zmniejszenie liczby uczestników eksperymentu, a zwiększenie liczby wyrazów-bodźców (Palermo, Jenkins

1964 – 500 badanych i 200 wyrazów; Postman, Keppel 1970). Idąc dalej tym tokiem myślenia, badacze brytyjscy, Kiss, Armstrong, Milroy, Piper (1973), zredukowali liczbę uczestników do 100 badanych, zwiększając liczbę badanych wyrazów do 8400. Badacze brytyjscy zmodyfikowali zarazem przebieg eksperymentu, gdyż jako pierwsi przeprowadzili eksperyment fazowo, co znaczy, że początkowy zbiór bodźców liczył nieco ponad 2000 wyrazów, które poddano testowi. Najczęstsze odpowiedzi do bodźców początkowych stały się bodźcami w drugiej fazie eksperymentu. Opisany zabieg umożliwił zbudowanie teaurusu dla języka angielskiego – The Edinburgh Associative Thesaurus (EAT), to jest pierwszej w dziejach eksperymentalnej sieci powiązań leksykalnych.

W drugiej połowie XX w. przetłumaczono listę Kent–Rosanoffa na szereg języków, otwierając tym samym nowy rozdział w badaniach. Badania skojarzeniowe przeprowadzono między innymi dla języków: francuskiego – Rosenzweig (1957), niemieckiego – Russell i Meseck (1959), włoskiego – Levi, Rosenzweig (1961). Dla języka polskiego badanie swobodnych skojarzeń słownych za pomocą listy Kent–Rosanoffa przeprowadziła Kurcz (1967), która przebadła 1000 studentów uczelni warszawskich.

Od momentu opracowania EAT wyraźnie wyróżniają się dwa nurty badań posługujących się testem skojarzeń słownych. Nurt pierwszy to prowadzone badania mechanizmów psycholingwistycznych, posługujące się zazwyczaj stosunkowo nielicznym zbiorem bodźców i prowadzone na stosunkowo nielicznych grupach osób. Badania te są prowadzone głównie dla potrzeb psychologii, więc nie będziemy tu o nich mówić. Nurt drugi to badania służące poznaniu struktury słownictwa i łączliwości znaczeń, prowadzone dla licznych zbiorów bodźców i na stosunkowo dużych grupach osób, np. De Deyne Storms (2008). Badania te często rezygnują z rygorów eksperymentu psycholingwistycznego, tj. są prowadzone przez Internet i badani mają podać więcej niż jeden wyraz kojarzący się z bodźcem, np. Schulte im Valde (2012). Gdy chodzi o badania nad słownictwem polskim, to jedyne duże badanie (110 wyrazów i 500 badanych) przeprowadził zespół badaczy szczebińskich – Gawarkiewicz, Pietrzyk, Rodziewicz (2008). Celem badania była budowa słownika asocjacyjnego, który opracowano dla potrzeb polsko-rosyjskich badań kontrastywnych. Słownik asocjacyjny zbudowano za pomocą testu podstawowego – jeden cykl dla 110 bodźców. W efekcie słownik asocjacyjny jest zbliżony do list skojarzeniowych (*association norms*) uzyskanych w wyniku eksperymentu Kurcz. Oba badania różnią się jedynie sposobem prezentacji wyników. Listy asocjacyjne Kurcz zawierają bodziec i zbiór odpowiedzi, z których każda jest opatrzona liczbą określającą, ilu badanych podało tę właśnie odpowiedź. W słowniku asocjacyjnym mamy dwa rodzaje

list. Pierwszy typ listy – jak u Kurcz – podaje bodziec i listę odpowiedzi dla tego bodźca, przy czym każda odpowiedź jest opatrzona liczbą odpowiadających. Drugi typ listy zawiera konkretną odpowiedź i listę bodźców, które wywołały daną odpowiedź, przy czym liczba odpowiadających, wiążących bodziec z odpowiedzią, jest umieszczona przy bodźcu – Kurcz nie opublikowała drugiego typu list, gdyż nie zalicza się ich do *association norms*. Patrząc na słownik asocjacyjny z sieciowego punktu widzenia, możemy powiedzieć, że listy pierwszego typu są zbudowane z powiązań wychodzących od konkretnego bodźca, a listy drugiego typu z powiązań przychodzących do konkretnej odpowiedzi. Węzły sieci leksykalnej zbudowane wokół bodźca są strukturami znacznie bogatszymi niż listy, nawet jeśli pominiemy informację zawartą w ścieżkach węzła leksykalnego (pewną orientację odnośnie do zawartości ścieżek daje graf węzła *dom*). Różnice pomiędzy listą skojarzeniową a węzłem leksykalnym sieci wystarczająco dobrze pokazują powiązania bezpośrednie wyrazu-bodźca. Lista skojarzeniowa bodźca w słowniku asocjacyjnym ma tylko powiązania wychodzące. Natomiast w sieci węzeł leksykalny bodźca ma powiązania wychodzące od i przychodzące do bodźca, co czyni węzeł strukturą semantycznie bogatszą. Pokażemy to na zbiorze najsilniejszych powiązań przymiotnika *biały*, który był bodźcem w obu badaniach. Dla potrzeb porównania w danych KLK UJ pominiemy siłę powiązania, podając tylko liczbę odpowiedzi.

Tabela 2.

Słownik asocjacyjny		Sieć leksykalna KLK UJ			
wychodzące		wychodzące		przychodzące	
czarny	128	śnieg	181	137	orzeł
kolor	78	kolor	111	131	czarny
śnieg	61	czarny	79	66	fartuch
czystość	28	kruk	43	47	ser
czysty	26	dom	34	39	cukier
niewinność	18	miś	28	30	mleko
bałwan	7	obrus	26	21	dym
anioł	6	koń	17	20	bielinek
kruk	6	ser	17	13	jasny
mleko	6	kieł	15	9	baran
orzeł	6	orzeł	15	5	owca
dobry	5	dzień	13	4	doktor
gołąb	5	papier	11	4	lekarz
flaga	4	kot	10	4	motyl

Porównanie powiązań wychodzących, uzyskanych w obu badaniach, pokazuje wspomnianą wyżej powtarzalność wyników testu skojarzeniowego – różnice w kolejności powiązań mogą wynikać ze sposobu przeprowadzania

testu (zob. niżej) i różnej liczby badanych. Natomiast powiązania przychodzące obecne w węzle sieci tworzą z wychodzącymi powiązania zwrotne, *biały* → *czarny* i *biały* ← *czarny* oraz wzbogacającą charakterystykę znaczeniową bodźca, np. *biały* ← *fartuch*, *mleko*, *cukier*, itd.

2. Eksperyment cykliczny KKK UJ

a. Dobór bodźców

Celem badania było zbudowanie eksperymentalnej sieci leksykalnej, więc eksperyment przeprowadzony przez KKK UJ został zaprojektowany jako badanie **cykliczne**. W pierwszym cyklu badania wykorzystano 63 wyrazy bodźce, tj. 63 rzeczowniki z listy Kent–Rosanoffa (w polskiej wersji językowej za: Kurcz 1967) – są to tzw. bodźce prymarne. W drugim cyklu jako bodźce wykorzystano najsilniejsze skojarzenia do bodźców prymarnych (po 5 skojarzeń dla każdego bodźca prymarnego) – są to tzw. bodźce sekundarne w liczbie 259. Zredukowana liczba bodźców sekundarnych wynika z tego, że jeśli jakiś wyraz pojawiał się w zbiorze pięciu najczęstszych odpowiedzi dla kilku bodźców prymarnych, to wyraz taki pojawiał się w zbiorze bodźców sekundarnych tylko raz. Ogółem w pierwszym i w drugim cyklu badania wykorzystano 322 wyrazy-bodźce. Cykliczność eksperymentu jest nowością w badaniach polskiego słownictwa i odróżnia nasze badania od badań Kurcz i Gawarkiewicza.

b. Osoby badane

W trakcie eksperymentu prowadzonego w latach 2011–2014 przebadano 900 studentów UJ i AGH. Badanie było anonimowe, uczestnik zaznaczał tylko wiek i płeć. Wśród badanych nie było studentów kierunków filologicznych ani studentów psychologii, gdyż wiedza i umiejętności wyniesione z tego typu studiów mogłyby zniekształcić wyniki testu.

c. Sposób przeprowadzenia eksperymentu

Eksperyment przeprowadziliśmy w warunkach kontrolowanych. Test odbywał się w laboratorium komputerowym, wykorzystano specjalnie zaprojektowany system komputerowy, co jest nowością w stosunku do wymienionych wcześniej badań wykorzystujących metodę formularzową (Kurcz 1967; Gawarkiewicz 2008). Eksperyment przebiegał następująco. Każdy uczestnik miał do dyspozycji komputer, na którego ekranie system wyświetlał wyrazy-bodźce. Polecenie brzmiało zapisz pierwszy wyraz, który przychodzi na myśl po przeczytaniu i zrozumieniu wyrazu bodźca. Badany miał ustalony empirycznie czas (5 sekund), by rozpocząć wpisywanie skojarzenia. Jeśli badany nie rozpoczął wpisywania odpowiedzi w ciągu 5 sekund, to

system zapisywał odpowiedź pustą, a na ekranie wyświetlał się kolejny wyraz bodziec. Dzięki takiej organizacji eksperymentu uczestnik testu nie miał możliwości powrotu do wyrazu, który wyświetlił się wcześniej. Nie mógł więc skorygować raz udzielonej odpowiedzi ani też uzupełnić odpowiedzi brakujących, co było możliwe w eksperymencie tradycyjnym prowadzonym za pomocą formularzy papierowych. Testy Kurcz i Gawarkiewicza były prowadzone metodą tradycyjną, co odróżnia je od naszego eksperymentu.

Podsumowując, można powiedzieć, że prowadzenie testu za pomocą systemu komputerowego wymusza spontaniczność skojarzeń, dostarczając zarazem danych istotnych dla psychologii. Przykładowo rejestracja czasu reakcji i odpowiedzi pozwala budować profile poszczególnych badanych.

3. Charakterystyka eksperymentalnej sieci leksykalnej

W wyniku eksperymentu uzyskaliśmy sieć, w której mamy: 50 849 powiązań oraz 11 224 węzłów leksykalnych, w tym 322 węzły pełne. Pozostałe to węzły zredukowane, tj. zbudowane z powiązań przychodzących. W sieci występuje 1181 par wyrazów (węzłów) wchodzących w powiązania zwrotne.

Zbiór węzłów sieci leksykalnej KLIK UJ tworzy:

- 7757 rzeczowników – 69,1%
- 2459 przymiotników – 21,9%
- 744 czasowników – 6,6%
- 264 inne części mowy – 2,4%

Zestawienie to pokazuje duże podobieństwo eksperymentalnie zbudowanej sieci języka polskiego do budowanej w ten sam sposób sieci języka niderlandzkiego (De Deyne, Storms 2008), gdzie węzłami są:

- rzeczowniki – 72%
- przymiotniki – 18%
- czasowniki – 9%
- inne – 1%

Liczba uczestników badania ma wpływ na strukturę węzła leksykalnego sieci. Bodziec w sieci KLIK UJ (900 badanych) ma średnio 150 powiązań wychodzących, podczas gdy w EAT (100 badanych) tylko 50 powiązań bezpośrednich. W rezultacie sieć KLIK ma bogatszą strukturę, co pozwala znaleźć w węzle leksykalnym ścieżki, wyjaśniające powiązania odległe, np. *baranina* – *sweter*. Węzeł leksykalny EAT ma strukturę uboższą, a więc strukturę, która często nie pozwala na wyjaśnienie powiązań odległych.

Literatura

- Bartmiński Jerzy, 2015, *DOM – koncept uniwersalny i specyficzny kulturowo*, [w:] *Leksykon aksjologiczny Słowian i ich sąsiadów*, t. 1, *Dom*, red. Jerzy Bartmiński, Iwona Bielińska-Gardziel, Beata Żywicka, Lublin: Wydawnictwo UMCS, s. 15–33.
- Biemann Chris, Heyer Gerhard, Quasthoff Uwe, Richter Matthias, 2007, *The Leipzig Corpora Collection – Monolingual corpora of standard size*, [w:] *Proceedings of Corpus Linguistics 2007*, Birmingham, UK, <http://ucrel.lancs.ac.uk/publications/CL2007/paper/190.Paper.pdf> (dostęp: 24.08.2016).
- Clark Herbert H., 1970, *Word Associations and Linguistic Theory*, [w:] John Lyons (red.), *New Horizon in Linguistics*, Middlesex: Penguin Books Ltd, Harmondsworth, s. 271–286.
- De Deyne Simon, Storms Gert, 2008, *Word associations: Network and semantic properties*, „Behavior Research Methods” 40 (1), s. 213–231.
- Gatkowska Izabela, 2014, *Word Associations as a Linguistic Data*, [w:] Piotr Chruszczewski, John R. Rickford, Katarzyna Buczek, Aleksandra Knapik, Jacek Mianowski (red.), *Languages in Contact 2012*, t. 1, Wrocław, s. 79–92.
- Gatkowska Izabela, 2015a, *Empty answers in an experiment of free word association*, „Academic Journal of Modern Philology”, t. 4, s. 41–49.
- Gatkowska Izabela, 2015b, *Empiryczna sieć powiązań leksykalnych*, „Polonica”, t. 35, s. 155–178.
- Gawarkiewicz Roman, Pietrzyk Izabela, Rodziewicz Barbara, 2008, *Polski słownik asocjacyjny z suplementem*, Szczecin: Print Group.
- Gawarkiewicz Roman, 2008, *Dom w językowym obrazie świata młodzieży polskiej i rosyjskiej*, [w:] Roman Gawarkiewicz, Izabela Pietrzyk, Barbara Rodziewicz, *Polski słownik asocjacyjny z suplementem*, Szczecin: Print Group, s. 235–246.
- Kent Grace H., Rosanoff Aaron J., 1910, *A study of association in insanity*, „American Journal of Insanity” 67 (37–96), s. 317–390.
- Kiss George R., Armstrong Christine, Milroy Robert, Piper James, 1973, *An associative thesaurus of English and its computer analysis*, [w:] Aitken A.J., Bailey R.W. (red.), *The Computer and Literary Studies*, Edinburgh: University Press, s. 153–165.
- Kurcz Ida, 1967, *Polskie normy powszechności skojarzeń swobodnych na 100 słów z listy Kent – Rosanoffa*, „Studia Psychologiczne”, t. 8, Wrocław–Warszawa–Kraków, s. 122–255.
- Minsky Marvin, 1975, *A Framework for Representing Knowledge*, [w:] Patrick H. Winston, McGraw-Hill (red.), *The Psychology of Computer Vision*, New York: McGraw-Hill, s. 211–277.
- Palermo David S., Jenkins James J., 1964, *Word Association Norms: Grade School through College*, Minneapolis: University of Minnesota Press.
- Postman Leo Joseph, Keppel Geoffrey, 1970, *Norms of word association*, New York: Academic Press.
- Rosenzweig Mark R., 1961, *Comparisons among word-association responses in English, French, German, and Italian*, „American Journal Psychology”, t. 64, s. 347–360.
- Rosenzweig Mark R., 1957, *Etudes sur l’association des mots*, „L’Année Psychologique”, t. 57, s. 23–32.

- Russell Wallace A., Meseck O.R., 1959, *Der Einfluss der Association auf das Erinnern von Worten in der deutschen, französischen und englischen Sprache*, „Zeitschrift für Experimentelle und Angewandte Psychologie”, t. 6, s. 191–211.
- Schank Roger C., 1975, *Conceptual Information Processing*, Amsterdam: North-Holland.
- Schulte im Walde Sabine, Borgwaldt Susanne, Jauch Ronny, 2012, *Association Norms of German Noun Compounds*, [w:] *Proceedings of the 8th International Conference on Language Resources and Evaluation*, Istanbul.
- Szymczak, Mieczysław (red.), 1978, *Słownik języka polskiego*, Warszawa: PWN, t. 1, s. 422.

Źródła elektroniczne

- Biemann Chris, 2011, <http://corpora.informatik.uni-leipzig.de/> (dostęp, 20.11.2015).

DOM ‘HOUSE/HOME’ IN EMPIRICAL LEXICAL NETWORKS

The article explains the concept of empirical lexical networks. A comparison is then made between the meanings of the Polish lexeme *dom* ‘house/home’ as represented in the corpus network Wortschatz (University of Leipzig) and the experimental network of the Department of Computational Linguistics of Jagiellonian University in Kraków. In both networks, the most vital meanings are ‘the place where the family is’ and ‘the place to live’. Both networks also contain the meaning ‘building’, whereas the meaning ‘institution’ is better represented in the corpus network. The experimental network additionally contains the meaning ‘my shelter’. That network, being characterised by directional internal links, organises subnetworks that explain meanings, e.g. the subnetwork of the meaning ‘the place where the family is’ points to the special role of ‘mother’, which organises the subnetwork that represents that meaning. A comparison of the meanings being identified and explained in subnetworks with the dictionary definitions suggests that research on the network representation of meaning may be useful in lexicography.

KEY WORDS: linguistics, empirical network, lexical semantics, network representation of meaning, house/home