

Marta Dobrowolska*
Uniwersytet Wrocławski, Wrocław
martadobr@gmail.com

HIPONIMY MIEJSCA W POLSKIEJ SŁOWOSIECI I AMERYKAŃSKIEJ BAZIE WORDNET¹

Słowa kluczze: wordnet, jednostka leksykalna, synset, rzutowanie wordnetów, relacja międzyjęzykowa

Keywords: wordnet, lexical unit, synset, wordnet-to-wordnet mapping, inter-lingual relation

1. Cel i charakter badania

Celem artykułu jest analiza podobieństw i różnic między siatką hiponimiczną leksemu *miejsce* w polskiej bazie relacji semantyczno-leksykalnych oraz analogiczną strukturą w amerykańskim WordNecie. Została ona przeprowadzona przede wszystkim podczas rzutowania fragmentu polskiej Słownosieci na amerykański WordNet (Princeton WordNet, dalej: PWN), czyli łączenia jednostek polskich z angielskimi za pomocą synonimii międzyjęzykowej lub innej relacji międzyjęzykowej, zdefiniowanej w polskim wordnecie (Słownosieci)². Rzutowanie trwa od marca 2012, a więc niespełna rok; w tym czasie zostało wprowadzonych 44 979 relacji międzyjęzykowych, w tym 4 225 relacji dla miejsc³. Obserwacje dotychczasowe mogą mieć charakter jedynie

* Autorka pozostaje pod opieką naukową dr hab. Agnieszki Libury.

1 Wielką literą pisane są nazwy własne baz, natomiast pisownia małą literą sygnalizuje użycie słowa w funkcji nazwy wspólnej.

2 Rzutowanie odbywa się w ramach projektu współfinansowanego ze środków Europejskiego Funduszu Rozwoju Regionalnego *Adaptacyjny system wspomagający rozwiązywanie problemów w oparciu o analizę treści dostępnych źródeł elektronicznych*: POIG.01.01.02-14-013/09.

3 Nie są to dane uwzględniające wszystkie hiponimy miejsca, te bowiem niekiedy mają inną dziedzinę (np. *obiekt naturalny, wytwór człowieka*), czyli pomocniczą etykietę służącą nie definiowaniu jednostek,

wstępnych ustaleń, pozwalających jednak określić pewne różnice między praktyką leksykograficzną a konstruowaniem wordnetów, jak również pomocnych na dalszych etapach rzutowania.

Podstawę porównania stanowią narzędzia semantyki strukturalnej, relacje semantyczno-leksykalne (Lyons 1983; Apresjan 2000), będące osią budowy zarówno PWN, jak i Słowosieci. Przy identyfikacji różnic między wordnetami wykorzystano też wiedzę z zakresu psychologii poznawczej (Rosch 2007), a także stosowane w semantyce kognitywnej pojęcie konceptualizacji. Niniejsza analiza nie opisuje języka, raczej jego model, a dokładnie model jednego z jego podsystemów. Szczególnie ważne było znalezienie odpowiedzi na pytanie, w jakim stopniu rzutowanie jest w stanie wyjaśnić strukturę semantyczną jednej siatki leksykalnej za pomocą odniesień do jednostek drugiej siatki.

Sygnałem różnic między wordnetami może być brak synonimu międzyjęzykowego lub też odmienny sposób zdefiniowania synonimicznych synsetów, a przyczyn tych różnic należy szukać zarówno w czynnikach językowych – przede wszystkim w różnej konceptualizacji tych samych zjawisk – jak i w sposobach budowania i rozwijania wordnetów. Typologizacja różnic między wordnetami, sporządzona na podstawie wyników pierwszego etapu rzutowania kilkunastu dziedzin, takich jak ludzie, miejsca i umiejscowienia, jednostki czasu itd., została przedstawiona w artykule dotyczącym strategii rzutowania polskiego wordnetu na amerykański (Rudnicka et al. 2012). W niniejszym artykule znajdują się odniesienia do tej typologizacji, skupiono się w nim jednak na problemach związanych z rzutowaniem hiponimów miejsca, czyli konkretnej gałęzi Słowosieci.

2. Wprowadzenie

Wordnet jest rodzajem tezaurusu, bazą jednostek leksykalnych uporządkowanych według kryterium semantycznego. Baza ta może być wykorzystana w różnoraki sposób – jako słownik synonimów, a także jako narzędzie bądź składnik narzędzi do przetwarzania języka naturalnego (Broda, Piasecki, Szpakowicz 2009: 11–13). Najstarszym i, jak do tej pory, największym wordnetem jest PWN. Podstawowym założeniem przy jego budowie było utworzenie struktury odzwierciedlającej proces akwizycji znaczeń leksykalnych u dzieci (ibid.: 7). Wordnet jest zatem u swych podstaw strukturą porządkującą sensy (pojęcia), nie zaś, jak ma to miejsce w przypadku tradycyjnych słowników, jednostek leksykalnych (Żmigrodzki 2003: 48–52).

Najmniejszym samodzielnym elementem WordNetu jest synset – zbiór słów odnoszących się do tego samego pojęcia. Podstawowymi relacjami znaczeniowymi są synonimia,

ale ułatwieniu tworzenia i edytowania Słowosieci. Obecność relacji międzyjęzykowych została zaś określona właśnie w ramach dziedzin.

hiperonimia i hiponimia oraz meronimia i holonimia. Relacja synonimii jest wyrażona przynależnością do jednego synsetu – pakietu, w którym zgrupowane są jednostki o takim samym (a najczęściej bardzo zbliżonym) znaczeniu. Pozostałe wymienione relacje semantyczne są nie tyle relacjami między jednostkami leksykalnymi⁴, co relacjami synsetów, i to przede wszystkim one służą definiowaniu znaczeń. Kolejnym po PWN, a zdecydowanie poprzedzającym Słowosiec projektem jest European WordNet (EWN), wielojęzyczna baza semantycznych relacji leksykalnych. Tworzą ją wordnety kilku języków europejskich, zaś kompatybilność ich struktur opiera się na wspólnie przyjętym indeksie międzyjęzykowym (Inter-Lingual-Index), czyli liście pojęć stanowiących podstawę tworzenia powiązań między znaczeniami leksykalnymi różnych języków (Peters et al. 1998: 222). Powiązana z indeksem tzw. ontologia⁵ wysokiego poziomu została zastosowana w EWN jako szkielet konceptualny, zbudowany z najistotniejszych pojęć (znaczeń) obecnych we wszystkich wordnetach. Struktura EWN daje możliwość porównywania leksyki różnych języków, a co za tym idzie, rozwijania technik translatologicznych.

3. Różnice w budowie Słowosieci i WordNetu

Jednym z czynników decydujących o różnicach w strukturze hiponimów miejsca w omawianych bazach leksykalnych są różnice w ich całościowej organizacji. Twórcy polskiego wordnetu postawili sobie za cel stworzenie struktury niezależnej od siatki pojęciowej zastosowanej w EWN, w zamian dokładnie charakteryzującej polską leksykę (Broda, Piasecki, Szpakowicz 2009: 15) – wraz z różnorodnymi niuansami stanowiącymi o specyfice polszczyzny (nie tylko w wymiarze samej leksyki, ale i morfologii). Dość istotna różnica między wordnetami polega na tym, że w polskiej bazie podstawowym samodzielnym elementem jest nie synset, ale jednostka leksykalna (Rudnicka et al. 2012: 1040). W Słowosieci, oprócz relacji synsetów, znajdują się zatem tzw. relacje jednostek. Są to relacje derywacyjne, które zachodzą w obrębie jednej lub różnych części mowy. Zestaw relacji WordNetu nie zawiera relacji derywacyjnych, bogactwo systemu słowotwórczego jest bowiem domeną języków słowiańskich (Derwojedowa, Zawisławska 2007). Podstawowe relacje synsetów są jednak w obu bazach jednakowe.

Polski i amerykański wordnet różnią się w sposób istotny nadrzędną organizacją struktury. Wyższe piętra hierarchii PWN bazują na tzw. ontologii SUMO (Suggested

4 Przyjmuje się, że jednostkę stanowi lemat oraz jego znaczenie. Przykładowo lematowi *dom* w polskim wordnecie odpowiada sześć jednostek, np. *dom* 1 to ‘budynek’, *dom* 2 – ‘miejsce zamieszkania’, *dom* 3 ‘instytucja, np. dom kultury, dom publiczny’.

5 Termin *ontologia*, będący zapożyczeniem semantycznym z języka angielskiego, jest tutaj używany zgodnie ze znaczeniem stosowanym w informatyce oraz badaniach nad sztuczną inteligencją. Oznacza on schemat pojęć i łączących je relacji, stanowiący reprezentację pewnego zasobu wiedzy (zob.: <http://www.ontologyportal.org/>).

Upper Merged Ontology), włączonej do wordnetu na potrzeby projektu utworzenia globalnej struktury skorelowanych ze sobą słowosieci. Zastosowana siatka składa się z 5000 ogólnych i w założeniu uniwersalnych pojęć, takich jak ‘corpuscular object’, ‘artifact’, ‘sentient agent’, ‘cognitive agent’, ‘human’ itd., przyporządkowanych poszczególnym synsetom zajmującym najwyższe piętra PWN. Polska baza nie włączyła SUMO do swojej struktury, stąd niemożliwe jest porównanie odpowiednich fragmentów polskiej i amerykańskiej bazy leksykalnej w odniesieniu do wspólnej siatki pojęciowej. Co więcej, zastosowana ontologia determinuje kształt relacji między synsetami na niższych poziomach hierarchii, a więc jest źródłem różnic między tymi hiponimami w bazie angielskiej i polskiej: pewne odpowiedniki polskich hiponimów *miejsca* 1 (np. *garage* 1 w PWN) będą miały jako daleki hiperonim np. *artifact*. Hipoteza dotycząca związku między różnymi poziomami hierarchii znajduje swoje uzasadnienie w konkretnych przypadkach obserwowanych w trakcie rzutowania.

4. Proces rzutowania

Procedura rzutowania polega na łączeniu synsetu Słowosieci z synsetem PWN za pomocą relacji synonimii międzyjęzykowej lub – przy braku synonimu – jedną z innych relacji międzyjęzykowych: synonimii międzyrejestrowej, hiponimii, meronimii, hiperonimii, holonimii oraz synonimii częściowej. Synonimia częściowa jest relacją stworzoną na potrzeby rzutowania, stosuje się ją wówczas, gdy zakresy znaczeniowe synsetów częściowo na siebie zachodzą. Dotychczas rzutowaniu poddano kilka tysięcy⁶ synsetów będących hiponimami *miejsca* 1, definiowanego jako ‘fragment przestrzeni lub powierzchni charakteryzowany ze względu na przeznaczenie, funkcję, budowę, cechę; odnosi się zarówno do obiektów naturalnych, jak i wytworzonych przez człowieka’, a także pozostałych siedmiu, znacznie mniej licznych pod względem zbioru hiponimów, jednostek o tym lemacie. Na obecnym etapie rzutowanie odbywa się w kierunku od bazy polskiej do amerykańskiej, zatem analiza materiału ma charakter jednokierunkowy. Stopień zaawansowania prac nad rzutowaniem – umożliwiający raczej poczynienie pewnych obserwacji i wskazanie prawidłowości niż sformułowanie ogólnych wniosków – oraz rozmiary niniejszej analizy nie pozwalają na symetryczne porównanie struktur wordnetów. Zastosowaną przez zespół edytujący Słowosieć metodę można nazwać rzutowaniem głębokim bądź pionowym: kolejne jednostki są opracowywane wraz ze wszystkimi jej hiponimami, aż do samego dołu drzewa. Jest to o tyle uzasadnione, że hiponimy stanowią podstawę weryfikowania zakresu znaczeniowego jednostki polskiej względem angielskiego kandydata na synonim. Odwrotny sposób polegałby na łączeniu relacjami międzyjęzykowymi synsetów należących do

6 Stan rzutowania w styczniu 2012 r.

kolejnych pięter hierarchii: po kolei, poczynawszy od wszystkich synsetów najwyższego piętra, do najniższego.

W dalszej części artykułu rozpatrzonych zostanie kilka problemów, które pojawiają się w trakcie rzutowania. Są one wynikiem, a zarazem sygnałem różnic między gałęzią hiponimów miejsca w Słownosieci a potencjalnymi odpowiednikami w WordNecie. Będą to: brak synonimu, brak hiperonimu, krzyżowanie się znaczeń synsetów oraz brak korelacji między poszczególnymi gałęziami polskiego i amerykańskiego drzewa semantycznego.

5. Synonimia a różne poziomy kategoryzacji

Pierwszym sygnałem różnic między fragmentami siatek relacji leksykalno-semantycznych jest brak w bazie amerykańskiej odpowiednika polskiego znaczenia, czyli tzw. synonimu międzyjęzykowego. Jego przyczyna może mieć charakter językowy lub pozajęzykowy – wtedy jest wynikiem zastosowania odmiennych założeń przy budowie wordnetów (Rudnicka et al. 2012). W wypadku braku synonimu szuka się dla polskiego synsetu jak najbliższego angielskiego hiperonimu. Często, żeby sprecyzować znaczenie, oprócz hiperonimu potrzebny jest holonim lub meronim. Tak się dzieje w przypadku nazw pomieszczeń technicznych, np. *emaliernia* to hiponim *room* ('an area within a building enclosed by walls and floor and ceiling') oraz meronim *factory*.

Jedną z podstawowych obserwacji nasuwających się w trakcie rzutowania dotyczy specyfiki różnych poziomów hierarchii systemu leksykalnego. Mianowicie, szukanie synonimu znaczenia ogólnego (znajdującego się wysoko w hierarchii) znacznie odbiega od analogicznej procedury u dołu drzewa semantycznego. Na szczegółowym poziomie kategoryzacji⁷ ustalenie synonimu lub stwierdzenie jego braku jest znacznie prostsze, dotyczy bowiem znaczeń odnoszących się do obiektów mocno skonkretyzowanych; leksemy je nazywające mogą mieć charakter specjalistyczny i być rzadziej używane w codziennej komunikacji. W naszym zbiorze będą to przykładowo nazwy pomieszczeń: *antykwiariat*, *ciemnia*, *suszarnia*, *królikarnia*, *pompownia*, *ptaszarnia*, *szlifiernia*, *suszarnia*, a także *budka suflera*, *pasmanteria*, *ogród zimowy*, *mykwa*, *epicentrum*, *perycentrum*, *gabinet luster*⁸. Dla części z nich istnieje w PWN synonimiczny synset,

7 Eleanor Rosch (Rosch 2007: 412–417) wyróżniła trzy poziomy kategoryzacji. Podstawowy (środkowy) to taki, dla którego ludzki umysł jest już w stanie wytworzyć i przechowywać reprezentację w postaci modelu (prototypu), np. *pies*, *fotel* (na poziomie najniższym znajdują się np. *jamnik*, *fotel bujany*, a na najwyższym – *zwierzę*, *mebel*). Poziomów kategoryzacji w wordnecie jest znacznie więcej niż wyróżnionych przez Rosch, zatem odniesienie danego synsetu do jednego z trzech poziomów ma charakter jedynie orientacyjny.

8 Pojedyncze jednostki leksykalne reprezentują tutaj cały synset. Jednostki będą numerowane wtedy, gdy będzie to niezbędne dla jasności wyводу, na przykład przy omawianiu co najmniej dwóch jednostek o tym samym lemacie.

np.: *pompownia* – *pump house*, *ptaszarnia* – *aviary*, *epicentrum* – *epicenter*, *perycentrum* – *periapsis*. Natomiast dla wielu jednostek z tego poziomu synonim się nie znalazł, np.: *antykwarjat*, *pasmanteria*, *królikarnia*, *szlifiernia*, *mykwa*. Nie są to nazwy, które nazywają obiekty obce kulturze i realiom angielskim bądź amerykańskim, odpowiedniki angielskie np. *mykwy* czy *pasmanterii* notują słowniki jednojęzyczne.

Nie widać tutaj jakichś cech semantycznych, które przesądzałyby o włączeniu danego leksemu do wordnetu, przy wyborze decydująca była frekwencja danego wyrazu w korpusach językowych. Czynnikiem decydującym o tego rodzaju rozbieżnościach jest nieuwzględnianie przez twórców wordnetów niektórych jednostek leksykalnych, podyktowane ograniczonymi rozmiarami tezaurusów.

Jak już wspomniano, o różnicach między wordnetami decydują również czynniki językowe. Należą do nich tzw. różnice leksykalne: kiedy dane pojęcie istnieje w obu językach, ale nie jest w jednym z nich wyrażone konkretnym leksemem, oraz tzw. różnice kulturowe: gdy brak leksemu odsyłającego do pewnego pojęcia jest spowodowany nieobecnością pojęcia w rzeczywistości pozajęzykowej (Rudnicka et al. 2012). Do rozbieżności pierwszego typu można zaliczyć brak angielskich odpowiedników dla słów *cerkiew* i *zbór*. W WordNecie świątynie różnych wyznań chrześcijańskich określa wspólny synset *church*, natomiast w tekstach pojawia się ono z odpowiednią przydawką (np. *Orthodox*, *Protestant*), sygnalizującą, o jaki desygnat chodzi. Przykładem rozbieżności drugiego typu jest brak w PWN synonimów takich jednostek jak: *uroczysko* ‘miejsce związane z kultem bóstwa w społeczności dawnych Słowian’, *województwo* ‘jednostka podziału administracyjnego w Polsce’, *kuczka* ‘liche pomieszczenie, szałas’. W przypadku *kuczki* trudne jest nie tylko znalezienie synonimu językowego, ale i połączenie jej z PWN za pomocą innej relacji międzyjęzykowej. Decydują o tym dwa czynniki. Po pierwsze, słowo *kuczka* pochodzi z niemalże wymarłej gwary lwowskiej, więc jest ono nacechowane regionalnie, a zarazem skostniałe. Po drugie, znaczenie podane w USJP odnosi się zarówno do pomieszczenia, jak i szałas, co wymaga rozdzielenia na dwa znaczenia i połączenia ich hiponimią międzyjęzykową odpowiednio z synsetami *room* oraz *hovel*, ponieważ w Słowosieci nie może pojawiać się tzw. hiponimia ‘lub’.

Da się również wskazać pewne różnice wynikające z odmiennej budowy morfologicznej polszczyzny i angielszczyzny. Widać to na przykładzie całych klas miejsc, które mają bardzo liczną reprezentację w Słowosieci, lecz nie znajdziemy dla nich odpowiedników w bazie amerykańskiej. Są nimi przede wszystkim hiponimy dwóch sztucznych synsetów⁹: *pomieszczenie techniczne* (tutaj 53 synsety, m.in.: *aparatownia*, *dojrzewalnia*, *emaliernia*, *greplarnia*, *kompresorownia*, *narzędziownia*) oraz *pomieszczenie, w którym żyją zwierzęta* (łącznie 23 hiponimy, np.: *ciełętnik*, *fokarium*, *królikarnia*, *lwiarnia*, *małpiarnia*, *owczarnia*). Mogłoby się wydawać, że do Słowosieci z jakichś

9 Sztuczne synsety nie są jednostkami języka naturalnego, a kategoriami pojęciowymi pozwalającymi grupować naturalne synsety w pewne klasy.

powodów włączono większą część specjalistycznych nazw pomieszczeń niż do PWN. W tym wypadku jednak rzut oka na definicje słownikowe pokazuje, że język angielski po prostu nie ma wielu odpowiedników polskich nazw pomieszczeń (np. *królikarni*). Natomiast niektóre z polskich jednostek, mimo że nie zostały zanotowane przez słowniki, są używane bądź w języku środowiskowym, bądź w polszczyźnie ogólnej. Ten stan rzeczy wynika wprost z właściwości polskiego systemu słowotwórczego. *Nomina loci* są kategorią semantyczną, którą uznajemy za odrębną kategorię słowotwórczą głównie dlatego, że ma ona swoje wykładniki formalne: sufiksy *-arnia*, *-alnia*, *-ownia*, *-isko*. Są to nazwy miejsc charakteryzowanych ze względu na wykonywaną w nich czynność, ich cechę fizyczną, a także znajdujące się w nim przedmioty, rośliny, zwierzęta (Grzegorzczkowska 1979: 47–49). Żywoćność i regularność tego modelu derywacji sprawia, że nazwy miejsc są w języku polskim często tworzone, nawet *ad hoc*.

Oczywiście, niewystarczalność systemu leksykalnego jest zawsze tylko częściowa, rzeczywistość możemy bowiem nazywać opisowo, za pomocą używanych często, zazwyczaj dwuwyrazowych zwrotów (np. według *Słownika polsko-angielskiego* PWN–Oxford *lakiernia* to *paint shop*, *antykwiariat* – *antique shop*), które jednak nie zostały włączone do WordNetu jako samodzielne jednostki leksykalne.

Na podstawowym poziomie kategoryzacji problem doboru synonimu przedstawia się nieco inaczej. Rzadko zdarza się tutaj, by w PWN brakowało leksykalnych wykładników jakichś pojęć uwzględnionych w Słownosieci. Znacznie częściej mamy do czynienia z ustrukturyzowaniem tych samych treści w różny sposób. Pewne pojęcie wyrażone jednym leksemem w języku polskim może być objęte znaczeniem angielskiego wyrazu o szerszym zakresie (denotacji), przy czym nie ma ono hiponimu odpowiadającego owemu polskiemu słowu. Taka sytuacja ma miejsce w przypadku jednostek *sklep* oraz *shop*, które w dydaktyce języków obcych oraz przez słowniki dwujęzyczne są przedstawiane jako synonimy. Tymczasem już glosa synsetu {*shop* 1, *store* 1} ('a mercantile establishment for the retail sale of goods or services') wskazuje na szersze znaczenie angielskiego leksemu w stosunku do polskiego, a potwierdzają to jego hiponimy. Oprócz takich synsetów jak *bookshop*, *boutique* oraz *butcher shop*, znajdują się wśród nich *repair shop* (dalszy hiponim to *garage*) czy *salon* ('a shop where hairdressers and beauticians work'). Mimo że wiele kohiponimów *sklepu* to synonimy kohiponimów synsetu {*shop* 1, *store* 1}, sam *sklep* może być połączony z nim tylko relacją hiponimii międzyjęzykowej.

Tego rodzaju sytuacje mają miejsce również na podrzędnym poziomie kategoryzacji, tak jak w przypadku *piekarni* oraz *bakery* ('a workplace where baked goods – breads and cakes and pastries – are produced or sold'). Budowa słowotwórcza tych wyrazów sugeruje, że oznaczają one takie same obiekty, tymczasem Polacy przez *piekarnię* rozumieją miejsce, gdzie piecze lub sprzedaje się przede wszystkim pieczywo. Warto zauważyć, że znaczenie *bakery* jako miejsca, gdzie wytwarza się lub sprzedaje pieczywo i innego rodzaju wypieki, wpasowuje się w zakres znaczeniowy wyznaczany przez jego hiperonim – *shop*. Tymczasem w Słownosieci muszą istnieć dwie jednostki – jedna

ma za swój hiperonim *sklep*, druga zaś *wytwórnię* – ponieważ jednostki Słowsieci, w przeciwieństwie do jednostek WordNetu, nie mogą być definiowane za pomocą tzw. hiponimii ‘lub’, a taką należałoby zastosować, gdyby wprowadzono jedną jednostkę dla lematu *piekarnia* – oznaczającą albo miejsce, gdzie się sprzedaje pieczywo, albo takie, gdzie się je piecze. Gdyby nasz *sklep* miał tak szerokie znaczenie jak angielskie *shop*, wystarczyłoby ono jako *genus proximum* dla znaczenia słowa *piekarnia*. Widać tutaj wyraźnie, jak wyższe piętra struktury determinują budowę niższych.

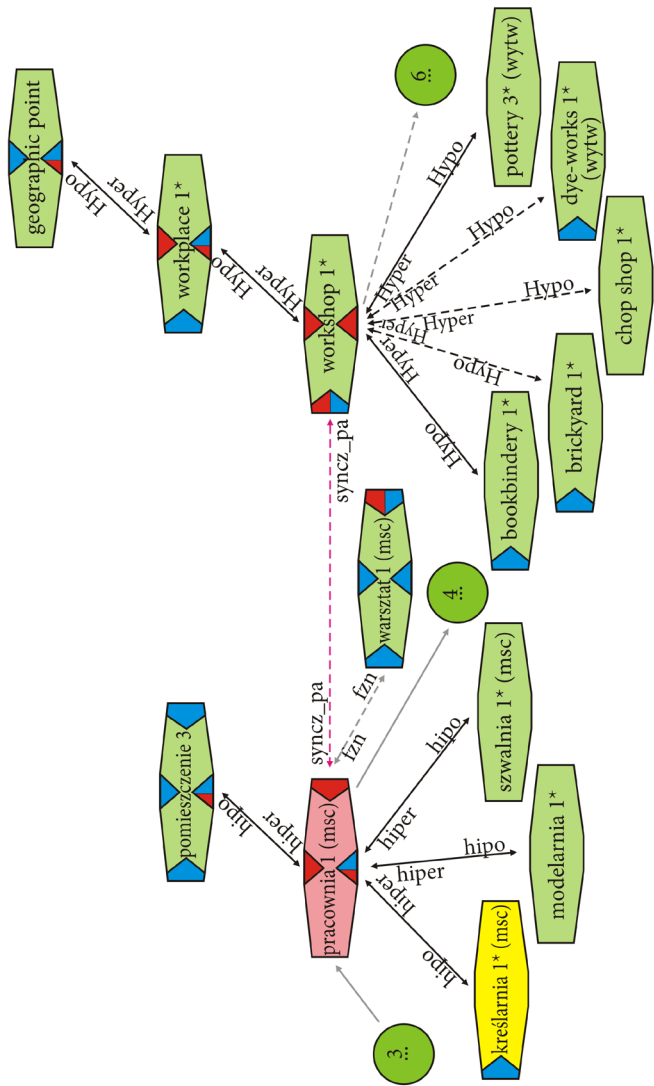
Mimo zawężającego pole widzenia badacza kierunku rzutowania udaje się zaobserwować pewne prawidłowości dotyczące braku w Słowsieci synonimów niektórych synsetów PWN, jak choćby odpowiednika angielskich *storage space* oraz *storeroom* o ogólnym znaczeniu ‘miejsca, pomieszczenia służącego do przechowywania’. Synsety te okazały się odpowiednimi hiperonimami dla jednostek niższego rzędu, dla których nie znalazł się synonim międzyjęzykowy (*dojrzewalnia*, *leżakownia*, *przechowalnia*, *składzik*, *tunel chłodniczy*, *zamrażalnia*) – pozwalającymi uniknąć bardzo ogólnych hiperonimów, takich jak *room*. Z drugiej strony – próżno szukać w angielskim leksykonie odpowiednika polskiego *pokoju* (‘część budynku, mieszkania, w której się mieszka lub pracuje’), znaczenie sugerowanego przez słowniki *room* jest znacznie ogólniejsze, bliższe *pomieszczeniu* – mamy więc tutaj do czynienia z relacją hiperonimii–hiponimii.

6. Krzyżowanie się znaczeń – synonimia częściowa

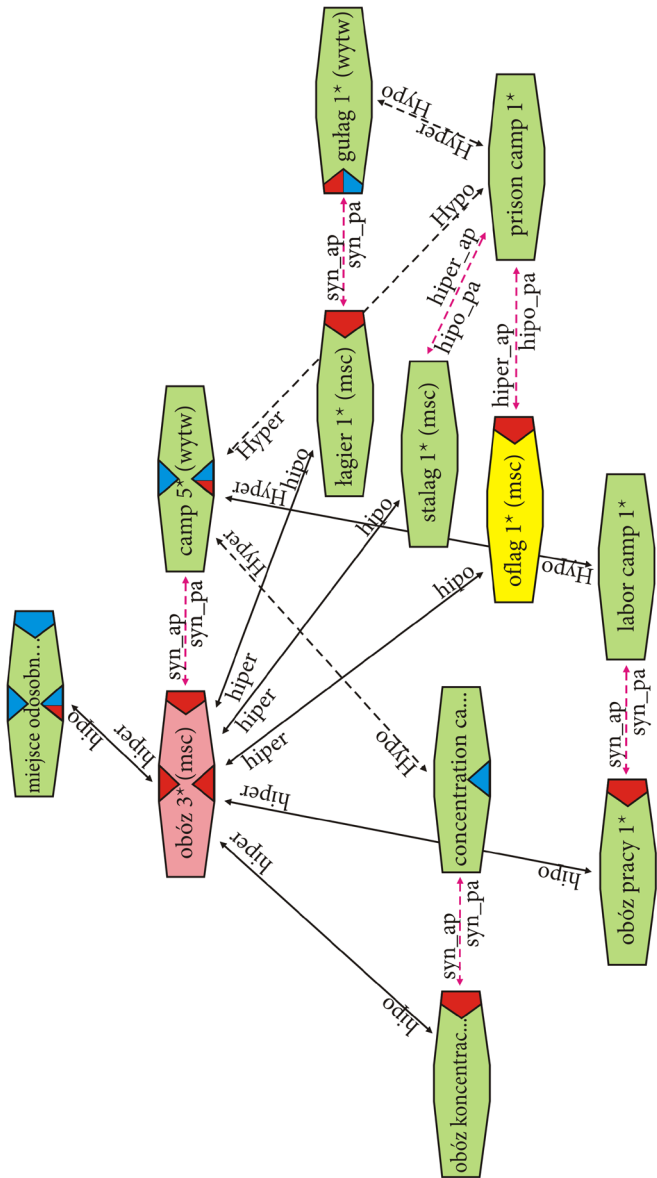
Z sytuacją krzyżowania się znaczeń spotykamy się wówczas, gdy niemożność połączenia synsetów synonimią międzyjęzykową spowodowana jest nie zbyt szerokim bądź zbyt wąskim znaczeniem jednego z synsetów, ale krzyżowaniem się zakresów ich treści. Wskaźnikiem takiego krzyżowania się – przy częstej schematyczności głoś do synsetów PWN – jest zestaw hiponimów obu synsetów. Wyjaśnia to, dlaczego identyfikujemy je najczęściej na wyższych poziomach kategoryzacji. Istotne jest tutaj zastrzeżenie, że nawet jeśli pozornie mamy do czynienia z synonimią częściową, istnienie relacji hiperonimii–hiponimii ją wyklucza, tak jak ma to miejsce w przypadku *drugstore* oraz *apteki* i *drogerii*. *Drugstore* (wyraz niegdyś oznaczający połączenie tych różnych dziś rodzajów sklepu) jest hiperonimem międzyjęzykowym *apteki* i *drogerii*, a więc nie może być częściowym synonimem.

Omawiana relacja zachodzi natomiast w przypadku synsetów *pracownia* i *workshop* (zob. rys. 1). Znaczenie *pracowni* obejmuje miejsce wytwarzania wyrobów rzemieślniczych oraz pracy artysty, jej hiponimy to: *cerownia*, *hamownia*, *kreślarnia*, *szwalnia*, *modelarnia*, *studio fotograficzne* i *wzorcownia* – trzy ostatnie nie nadają się na hiponimy dla *workshop*. Również *workshop* (‘small workplace where handcrafts or manufacturing are done’) posiada hiponimy, których znaczenie zawiera się częściowo w treści *pracowni* (*print shop*, *bookbindery*, *dye-works*), a częściowo poza nią wykracza (np.

Rys. 1. Synonimia częściowa między synsetami *pracownia 1* i *workshop 1*



Rys. 2. Zgodność synsetów *obóz* 3 i *camp* 5 ze względu na zbiór ich hiponimów



chop shop, brickyard, saddlery). Zazębiające się zakresy znaczeniowe (przy wyraźnej rozłączności części zbiorów hiponimów) wskazują na relację synonimii częściowej. Jak wspomniano, relacja ta zdarza się stosunkowo rzadko, zazwyczaj bowiem przy braku synonimii mamy do czynienia z hiponimią bądź hiperonimią (w zależności od kierunku rzutowania), natomiast istnieją przypadki, kiedy żadna inna relacja nie jest w stanie posłużyć do określenia związku semantycznego między synsetami z dwóch różnych języków.

7. Podobieństwa i różnice między strukturami

Między poszczególnymi miejscami struktury wordnetów rysują się pewne odpowiedniości. Mimo to gałęzie, których głównymi węzłami są synsety będące bardzo dokładnymi synonimami, niekoniecznie muszą sobie odpowiadać w tym sensie, że istnieje między nimi zgodność co do hiperonimu oraz zestawu hiponimów. Jeśli zaś różnią się one przyporządkowaniem do właściwego hiperonimu, to znaczy, że odmienny jest sposób ich zdefiniowania, ponieważ hiponimia jest główną relacją definiującą znaczenie synsetów. Owo zróżnicowanie jest skutkiem odmiennych strategii, jakie przyjęli twórcy obu baz leksykalnych, konstruując najwyższe piętra hierarchii (zastosowanie SUMO w PWN), a także dobierając oraz porządkując zasób leksyki. Zastosowanie w Słownosieci sztucznych synsetów – niebędących jednostkami języka, porządkujących jedynie jednostki leksykalne w pewne klasy, a co za tym idzie, nieodgrywających roli w definiowaniu znaczeń – spowodowało, że potencjalne hiponimy naturalnych synsetów znalazły się w innym miejscu Słownosieci. Przyczyna takiego a nie innego umieszczania poszczególnych synsetów w hierarchii może mieć też uzasadnienie wewnątrzjęzykowe: te same obiekty mogą być przez użytkowników różnych języków konceptualizowane (czyli pojęciowo strukturyzowane) w odmienny sposób. W języku polskim wszelkie pomieszczenia, jako wydzielone przez człowieka fragmenty przestrzeni, będą traktowane jako rodzaje miejsca (*topographic point* w PWN), natomiast w języku angielskim większość pomieszczeń to hiponimy *structure* ('a thing constructed; a complex entity constructed of many parts'), a patrząc w górę hierarchii – *artifact* ('a man-made object taken as a whole'). Wspólnym hiperonimem *place* 1 (czyli *topographic point*) oraz *structure* jest dopiero *object* ('a tangible and visible entity'), synset o bardzo ogólnym znaczeniu.

Zilustrujmy nakreślony problem przykładami. Polski synset *obóz* (hiperonim: *miejsce odosobnienia* – 'miejsce, w którym ktoś jest więziony lub internowany' (ISJP)) łączy się synonimią międzyjęzykową z synsetem *camp* 6 (hiperonim: *penal institution*). Co bardzo istotne dla spójności rzutowania, wszystkie synonimy międzyjęzykowe hiponimów *obozu* (*obóz koncentracyjny, obóz pracy, łagier*) są hiponimami *camp* (*oflag* i *stalag* są hiponimami *prison camp*). Również hiperonimy *obozu* oraz *camp* są ze sobą powiązane: *miejsce odosobnienia* jako nieco szczegółowsze od *penal institution* będzie jego hiponimem (wyznaczające zakres znaczeniowy zbioru hiponimów są jednak w przypadku obu

nazw bardzo zbliżone, co przesuwając relację w kierunku synonimii), różnice pojawiają się dopiero na wyższych piętrach hierarchii: hiperonimem *miejsca odosobnienia* jest *miejsce*, natomiast hiperonim *penal institution*, *institution* jest zaklasyfikowany jako wytwór człowieka (*artifact*). Pomimo rozbieżności na wyższych poziomach klasyfikacji można uznać, że struktura tej gałęzi hiponimów miejsca znajduje całkiem wierne odbicie w PWN (zob. rys. 2).

Znacznie bardziej skomplikowany jest przypadek synsetów *miejsce pracy* i *workplace*. Synsety te są synonimiczne, mimo że zbiory hiponimów wyznaczające zakres treści obu nadrzędnych pojęć w dość znacznym stopniu się różnią. Różnice te mają trojaki charakter: synonim danego hiponimu *miejsca pracy* nie jest hiponimem *workplace*, bliski (tzn. o szczegółowym znaczeniu) hiperonim nie jest hiponimem *workplace* albo wreszcie dany polski hiponim *miejsca pracy* nie ma synonimu bądź bliskiego hiperonimu w PWN. Do synsetów pierwszego rodzaju należą np.: *fabryka* (synonim *factory* – hiperonim *building*), *oczyszczalnia* (*sewage disposal* – *building*), *młyn* (*mill* – *machinery*), *hurtownia* (*warehouse* – *depository*), *kibuc* (*kibutz* – *collective*), *reżyserka* (*control room* – *room*)¹⁰. Do drugiej grupy należą: *gabinet kosmetyczny* (hiperonim *salon* – *shop*), *resort* (*department* – *division*), a do trzeciej: *ksero*, *myjnia*, *wypożyczalnia*. Sytuacje te mają miejsce oczywiście przy rzutowaniu również w odwrotnym kierunku, w analizie skupiono się jednak na aspektach różniących stronę amerykańską od polskiej. Najistotniejsze z punktu widzenia braku odpowiedniości między strukturami wordnetów są synsety należące do pierwszej grupy. Przyczyną tych różnic jest wydobywanie w trakcie definiowania danego pojęcia różnych dominant semantycznych, np. po stronie polskiej ‘miejsce pracy’, po amerykańskiej zaś ‘budynek’. Efektem różnych sposobów definiowania poszczególnych synsetów są znaczące różnice w zestawach hiponimów dwóch synsetów nadrzędnych, synonimicznych względem siebie.

Z nieco odmienną sytuacją mamy do czynienia w przypadku *miejsca kultu* oraz *place of worship*. W tym wypadku nadrzędny poziom hierarchii determinuje budowę dolnego piętra. Hiperonimem *place of worship* jest *building* – ‘a structure that has a roof and walls and stands more or less permanently in one place’ – co powoduje, że nic, co nie jest budynkiem, np. *altar*, synonim *ołtarza*, będącego w Słownosieci rodzajem miejsca kultu, nie może być hiponimem *place of worship*. Zatem niemające angielskiego odpowiednika *uroczysko* 1 – ‘miejsce, zwykle w głębi puszczy, m.in. związane z kultem bóstwa lub uważane za siedzibę złych duchów’ (na podstawie USJP) – nie może być zdefiniowane jako międzyjęzykowy hiponim *place of worship*. *Miejsce kultu* i *place of worship* wydają się synonimami, nie jest jednak łatwo to sprawdzić: wyrażenia *place of worship* nie notują bowiem słowniki jednojęzyczne. Natomiast hiperonim oraz hiponimy tego synsetu w PWN wskazują na wyraźną różnicę znaczeniową między nim a polskim *miejscem kultu*.

¹⁰ W celu podkreślenia różnic w definiowaniu synonimicznych synsetów podawane tutaj hiperonimy nie są hiperonimami bezpośrednimi, lecz wyższego rzędu.

Przyczyną różnic w definiowaniu synonimów językowych są też różnice w strategiach rozbudowy wordnetów, czyli w sposobie organizowania siatek jako całości. Jak już wspomniano, polska baza różni się od PWN zastosowaniem sztucznych synsetów, niebędących jednostkami języka naturalnego. Skutki użycia ich w strukturze Słownosieci można zaobserwować na przykładzie hiponimów *pomieszczenia*, spośród których ponad jedna trzecia została przyporządkowana jednemu z czterech sztucznych synsetów, eksponujących funkcję pomieszczeń, nie zaś na przykład ich budowę. Najwięcej hiponimów zgromadził synset *pomieszczenie techniczne*. Angielskie odpowiedniki poszczególnych jego hiponimów mają różnorodne hiperonimy: *maszynownia* (synonim *engineering* – hiponim *room*), *lokomotywnia* (*roundhouse* – *workplace*), *newsroom* (*newsroom* – *office*), *pompownia* (*pump house* – *house*), *kotłownia* (*stokehold* – *chamber*)¹¹. Mają one bogatszą treść niż *pomieszczenie* lub należą do zupełnie innej gałęzi wordnetu (np. *building*).

Twórcy wordnetów przyjęli również odmienne sposoby definiowania niektórych relacji synsetów. Różnice te dotyczą chociażby synonimii: w PWN jest ona określana mniej restrykcyjnie niż w Słownosieci. Przykładowo, w jednym synsecie PWN mogą znajdować się jednostki różniące się kategorią liczby, np. {*Pyramid* 3, *Great Pyramid*, *Pyramids of Egypt*}. Podobnie jest w przypadku hiponimii. Dopuszczanie w PWN hiponimii ‘lub’ powoduje, że ma ona charakter tylko częściowo przechodni, tzn., że nie wszystkie hiponimy hiponimów danej jednostki są jej hiponimami. Widać to na przykładzie synsetów *side chapel* i *lady chapel*, które dwa piętra ponad sobą mają hiperonim *place of worship*, zdefiniowany jako rodzaj budynku. Te dwa rodzaje kaplicy są częściami kościoła, nie mogą więc być hiponimami budynku. Żeby uniknąć takiej sytuacji, edytorzy Słownosieci rozdzielają znaczenia słów, które mają w definicji słownikowej spójnik *lub*. Mamy zatem *kaplicę* 1, która jest samodzielną budowlą, oraz *kaplicę* 2, będącą częścią kościoła. Są one zatem hiponimami międzyjęzykowymi angielskiego *chapel* (Rudnicka et al. 2012: 1045–1046).

Różnice między wordnetami wynikają również z zastosowania w opisie tej samej zależności semantycznej innych relacji (ibid.: 1045). Synset {*jezdnia* 1, *ulica* 2} ‘część ulicy przeznaczona dla ruchu pojazdów’ jest połączony z *ulicą* 1 – rozumianą jako ‘wytyczona droga w mieście, na osiedlu itp., składająca się z jezdni, przeznaczonej dla pojazdów, i z chodników dla pieszych’ (USJP) – relacją meronimii, podczas gdy jego angielski odpowiednik *street* 2 ‘the part of a thoroughfare between the sidewalks, on which vehicles travel’ – niejako wbrew głosie – łączy się z synsetem *thoroughfare* za pomocą hiponimii.

Wszystkie wymienione różnice powodują dwukrotną przewagę hiponimii międzyjęzykowej nad synonimią językową wśród relacji poprowadzonych od hiponimów miejsca do synsetów PWN. Obrazuje to stosunkowo duży stopień niezgodności fragmentów wordnetów odnoszących się do miejsc, bowiem w przypadku wszystkich zrzuconych synsetów hiponimii międzyjęzykowej jest jedynie o 13% więcej niż synonimii (22 071 relacji do 19 556). Dla porównania: wśród nazw roślin zdecydowanie przeważa synonimia.

¹¹ Hiponimom *pomieszczenia* przyporządkowano ich hiperonimy bezpośrednie – czyli synsety stanowiące dla ich *genus proximum*.

Może być to spowodowane tym, że ich językowe znaczenia oparte są na taksonomii stosowanej w biologii, a ta jest wspólna całemu zachodniemu kręgowi cywilizacji, podczas gdy nazwy miejsc bądź ludzi w większym stopniu bazują na pojęciach potocznych, uwikłanych w powiązania kulturowe i w rzeczywistości bardziej złożonych niż pojęcia powiązane z taksonomią naukową.

Tabela 1. Liczba poszczególnych relacji międzyjęzykowych dla nazw miejsc i roślin

Relacja międzyjęzykowa	Synonimia	Hiponimia	Hiperonimia	Meronimia	Holonomia	Synonimia częściowa	Synonimia międzyrejestrowa
Miejsca	1614	2299	76	77	5	65	89
Rośliny	315	118	2	5	2	1	1

8. Podsumowanie

Analiza różnic między strukturami hiponimów miejsca Słownosieci i WordNetu, przeprowadzona na podstawie rzutowania pierwszej bazy na drugą, potwierdza wnioski wyciągane na podstawie wiedzy o sposobie budowania i rozwijania PWN oraz Słownosieci: organizacja wyższych pięter hierarchii leksykalnej wywiera znaczący wpływ na definiowanie znaczeń jednostek niższego rzędu, a co za tym idzie, na powstawanie rozbieżności między wordnetami. Niekompatybilność porównywanych fragmentów baz objawia się dwójako: w braku synonimów jednostek poziomu szczegółowego oraz w odmiennym zdefiniowaniu – za pomocą relacji semantyczno-leksykalnych – synsetów poziomu podstawowego. Określenie wszystkich możliwych przyczyn tych rozbieżności wymaga znacznie szczegółowszych badań porównawczych. Przypuszcza się, że do takich należą: czynniki wewnątrzjęzykowe, odmienne założenia przyjęte przed budową baz leksykalnych oraz różne sposoby rozwijania wordnetów. Analizując proces rzutowania w obrębie hiponimów miejsca, znajdujemy przykłady większości typów przyczyn dyferencji zaproponowanych przez Rudnicką et al. (2012). Do czynników językowych zaliczamy: brak danego pojęcia w kulturze i języku, brak leksemu nazywającego pojęcie, odmiennosc morfologii języków. Wśród czynników pozajęzykowych znajdują się: różnice w definiowaniu relacji synsetów, odmienne strategie wydzielania znaczeń, stosowanie innych relacji do wyrażenia tych samych zależności semantycznych, zastosowanie ontologii SUMO w WordNecie, obecność sztucznych synsetów w Słownosieci oraz nieuwzględnianie przez twórców wordnetów pewnych znaczeń podyktowane ograniczonymi rozmiarami tezaurusów. Warto również pamiętać, że stopień niezgodności strukturalnej między bazami leksykalnymi nie przedstawia się symetrycznie: miejsca są dziedziną, w której jest on stosunkowo wysoki, co objawia się znaczną przewagą relacji hiponimii językowej nad synonimią.

Brak korelacji między podstrukturami drzew leksykalnych uniemożliwia oddanie w trakcie rzutowania pełnej struktury semantycznej danego fragmentu wordnetu – będzie to wyzwaniem na dalszych etapach rzutowania, a jeszcze trudniejszym zadaniem uczyni próby zastosowania odwzorowanej struktury do automatycznych translatorów. Stworzenie pogłębionej systematyki różnic pomoże w dobraniu odpowiednich procedur rzutowania w przypadku zachodzenia konkretnych typów rozbieżności między wordnetami.

Literatura

- APRESJAN J., 2000, *Semantyka leksykalna. Synonimiczne środki języka*, Warszawa.
- BRODA B., PIASECKI M., SZPAKOWICZ S., 2009, *A Wordnet from The Ground Up*, Wrocław.
- CAMBRIDGE DICTIONARIES ONLINE: dictionary.cambridge.org/dictionary, 20 IV 2012.
- DERWOJEDOWA M., ZAWISŁAWSKA M., 2007, *Relacje leksykalne w polskiej i czeskiej bazie WordNet*, [w:] H. Popowska-Taborska, Z. Rudnik-Karwatowa (red.), *Językoznawstwo: prace na XIV Międzynarodowy Kongres Slawistów w Ochrydzie 2008*, Warszawa, s. 15–23.
- ENGLISH-POLISH DICTIONARY. *Wielki słownik angielsko-polski PWN–Oxford*, Warszawa.
- GRZEGORCZYKOWA R., 1979, *Zarys słowotwórstwa polskiego. Słowotwórstwo opisowe*, Warszawa.
- [HTTP://WWW.ONTOLOGYPORTAL.ORG/](http://www.ontologyportal.org/), 24 kwietnia 2012.
- ISJP: M. Bańko (red.), *Inny słownik języka polskiego*, Warszawa 2002.
- LYONS J., 1983, *Semantyka*, t. 1–2, Warszawa.
- OXFORD DICTIONARIES: <http://oxforddictionaries.com/>, 20 IV 2012.
- PETERS W., VOSSEN P., DÍEZ-ORZAS P., ADRIAENS G., 1998, *Cross-linguistic Alignment of Wordnets with an Inter-Lingual-Index*, [w:] P. Vossen (red.), *EuroWordNet. A Multilingual Database with Lexical Semantic Networks*, Dordrecht, s. 149–179.
- ROSCH E.H., 2007, *Zasady kategoryzacji*, [w:] Z. Chlewiński (red.), *Psychologia poznawcza w trzech ostatnich dekadach XX wieku*, Gdańsk, s. 409–430.
- RUDNICKA E., MAZIARZ M., PIASECKI M., SZPAKOWICZ S., 2012, *A strategy of Mapping Polish WordNet onto Princeton WordNet*, [w:] M. Kay, C. Boitet (red.), *Proceeding of COLING 2012: Posters Volume 3*, Mumbai, s. 1039–1048.
- SŁOWNIK POLSKO-ANGIELSKI: www.angool.com, 16 V 2012.
- SŁOWNIK POLSKO-ANGIELSKI: www.ling.pl, 20 V 2012.
- USJP: S. Dubisz (red.), *Uniwersalny słownik języka polskiego*, Warszawa 2003.
- ŻMIGRODZKI P., 2003, *Wprowadzenie do leksykografii polskiej*, Katowice.

Hyponyms of place in Polish and American wordnets

Summary

The aim of the paper is to compare the sets of hyponyms of *place* in Polish and American wordnets. Polish “Słownosiec” and American WordNet are thesauri which define the meaning of an individual lexeme by the lexical-semantic relations existing between this and other lexemes. The researched material consists of several thousand of synsets (sets of lexemes referring to a common concept) which have

been linked by interlingual, lexical-semantic relations to the synsets of English lexemes in a process of mapping of a fragment of the Polish wordnet to the American database. An analysis of particular cases illustrates the basic issues concerning the procedure of mapping, first of all the problem of choosing a foreign synonym for the original word. It also shows the incompatibility of the corresponding fragments of the Polish and English lexical systems, which is usually not reflected in dictionaries and other glottodidactical publications. Differences between the structures of semantic trees of the two languages result partly from different assumptions made by the authors of both wordnets, and partly from language-specific factors, such as different word-formation systems or different ways in which some objects from the real world are conceptualized.