

Dwujęzyczna Słowność – możliwości wykorzystania w pracy tłumacza



**Ewa Rudnicka,
Tadeusz Piotrowski**
Politechnika Wroclawska
Katedra Inteligencji
Obliczeniowej
Grupa Naukowa G4.19
ewa.rudnicka@pwr.wroc.pl
2013-04-17

Plan



- Słowność vs. WordNet Princeton:
 - specyfika struktury zasobów
 - różnice w opisie znaczeń jednostek leksykalnych i synsetów
- strategia powiązania zasobów i jej etapy
- system relacji międzyjęzykowych:
 - definicje, hierarchia oraz przykłady relacji
 - relacje specyficzne dla różnych części mowy
 - procedura rzutowania
 - liczności relacji

Słownosieć vs Princeton WordNet



- **Synsety** - zbiory jednostek leksykalnych:

pIWN: wchodzących w te same relacje konstytutywne
(np. hiponimia, hiperonimia, meronimia, antonimia)
(cf. Piasecki et al. 2009, Maziarz et al. 2012)

PWN: będących w relacji synonimii, traktowanej jako relacja
konceptualna ustalana na podstawie intuicji językoznawcy
oraz definicji słownikowych (cf. Fellbaum 1998)

- **Glosy:**

pIWN: glosy dla jednostek;

PWN: definicje dla synsetów

- **Relacje:**

pIWN: bogaty zbiór synsetów i relacji jednostek leksykalnych

PWN: nieduży zbiór relacji synsetów i jednostek leksykalnych

Etapy powiązania Słownosieci z WordNetem Pryncetońskim



- Zdefiniowanie zestawu relacji międzyjęzykowych oraz ustalenie ich hierarchii (cf. Rudnicka et al. 2012)
- Opracowanie procedury rzutowania czyli łączenia synsetów na podstawie podobieństwa znaczeń oraz miejsca w strukturze sieci
- Kierunek rzutowania: Słownosieć > Princeton WordNet od najniższych poziomów sieci stopniowo obejmując coraz wyższe
- Obecnie rzutowane: rzeczowniki i przymiotniki

Relacje międzyjęzykowe

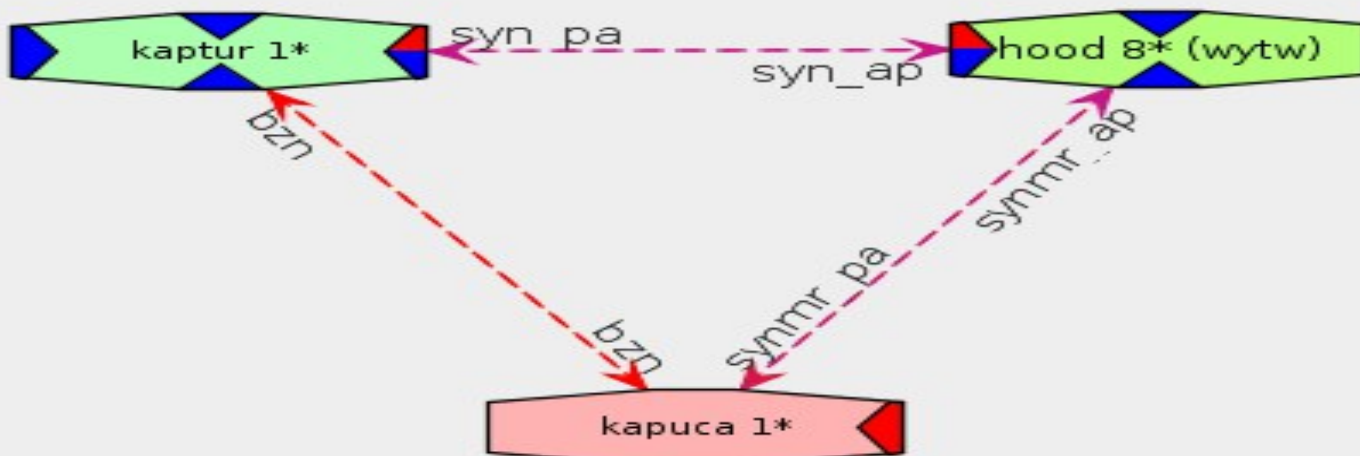


- Zestaw relacji międzyjęzykowych inspirowany:
 - relacjami międzyjęzykowymi z EuroWordNetu (Vossen 2002)
 - relacjami wewnątrzjęzykowymi ze Słownosieci (Maziarz et al. 2012)
- Synonimia (ekwiwalencja)
- Hiponimia i hiperonimia
- Meronimia i holonimia
- Synonimia częściowa
- Synonimia międzyrejestrowa
- Synonimia miedzyparadygmatyczna

Synonimia



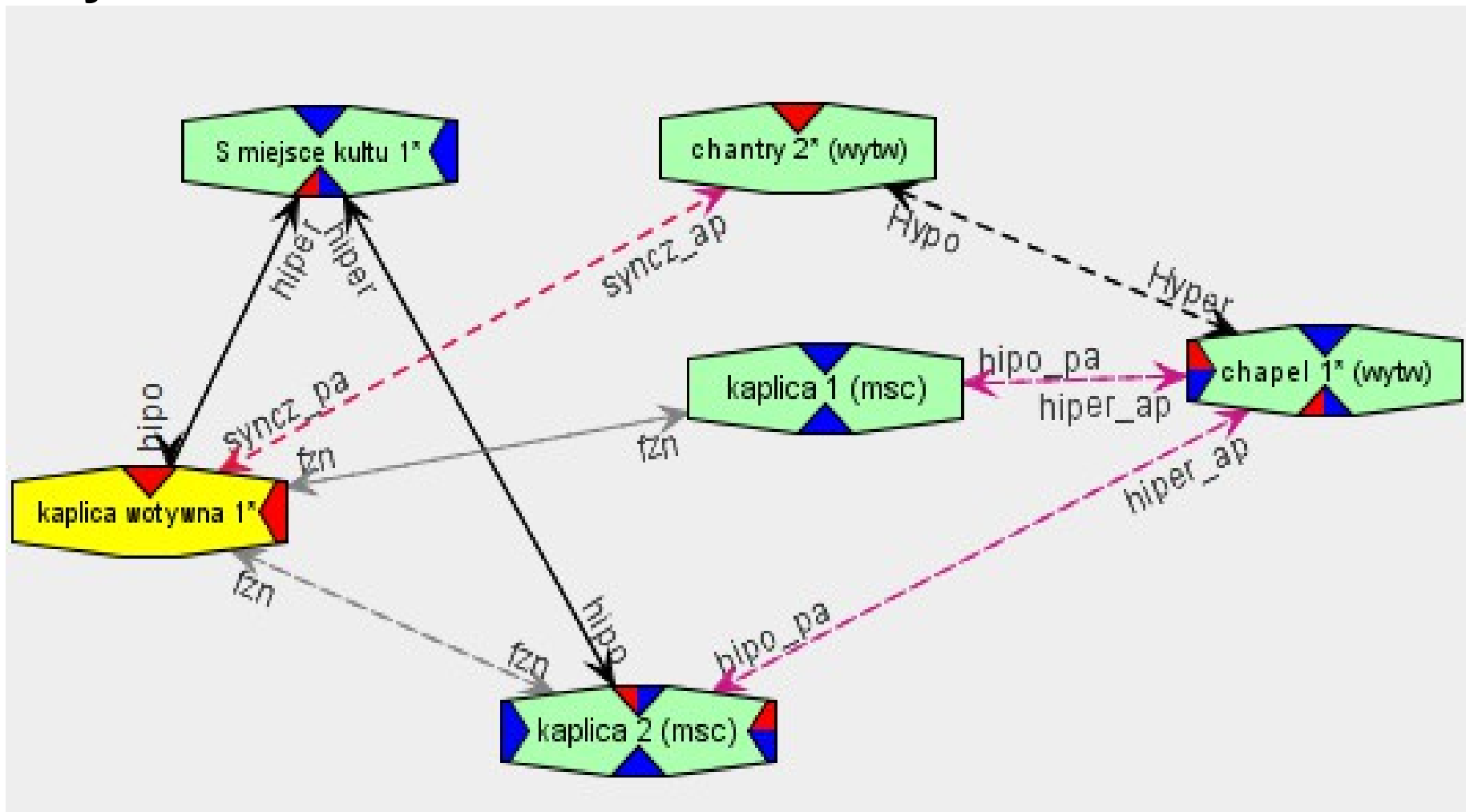
- **Synonimia** - oznacza odpowiedniość znaczeń i struktur; traktowana jako podstawowa i priorytetowa relacja; może być tylko jedna dla jednego synsetu
- **Synonimia częściowa** - stosowana w przypadku częściowej odpowiedniości znaczeń i/lub struktur
- **Synonimia międzyrejestrowa** - dla odpowiedników znaczeniowych różniących się rejestrami stylistycznymi



Synonimia częściowa



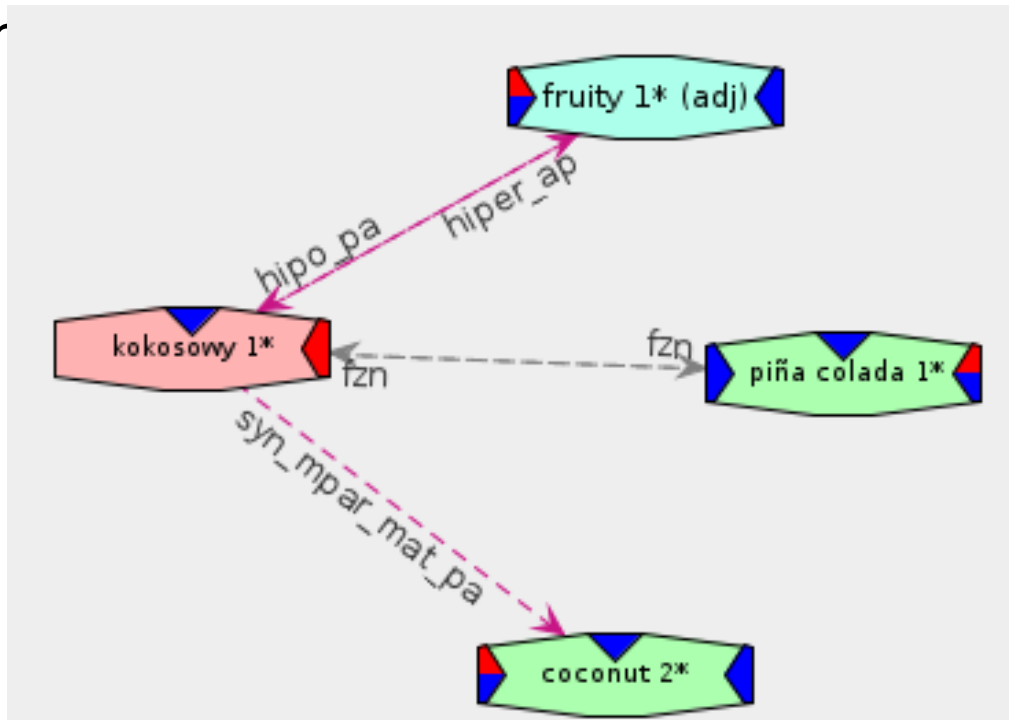
Przykład



Synonimia międzyparadygmatyczna (1)



- 'made of' - 'zrobiony z' (materiał)
<x#> określa materiał denotowany przez <y#>
<x#> łączy relacja synonimii międzyparadygmatycznej lub derywacyjności z N
N i <y#> łączy relacja międzyjęzykowa (najlepiej synonimii, ew. hiponir



Synonimia międzyparadygmatyczna (2)



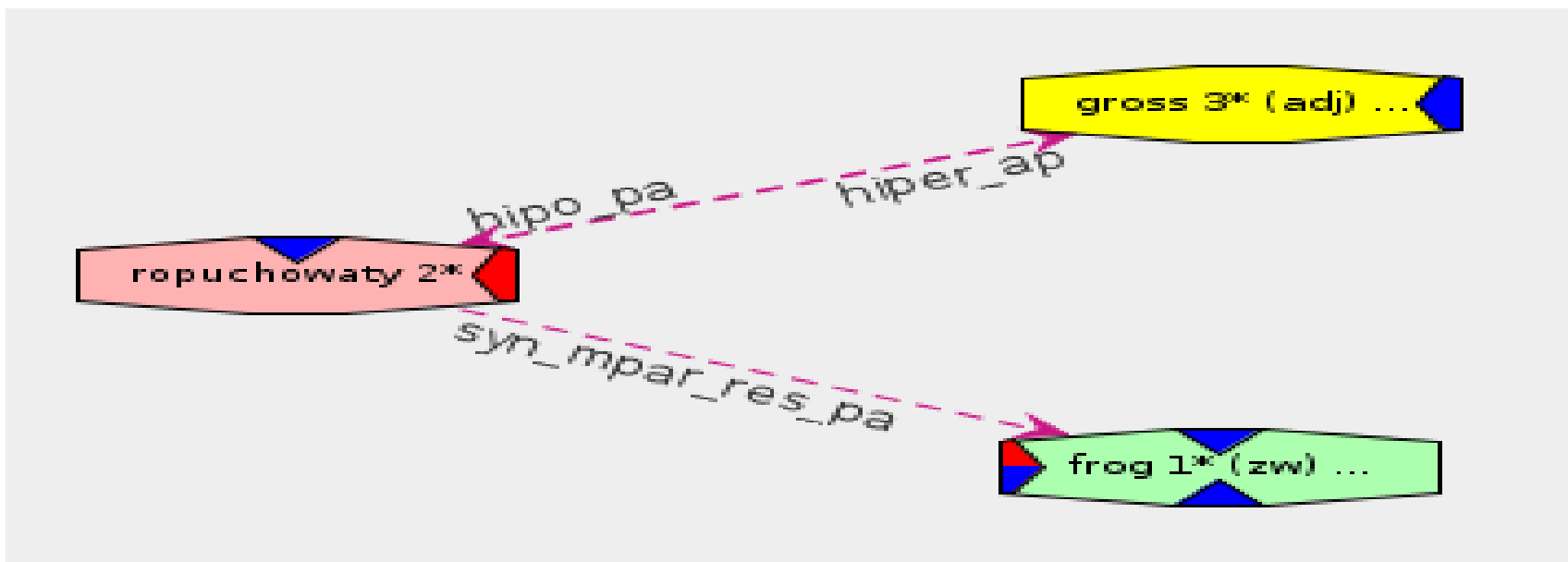
<x#> jest przymiotnikiem w języku polskim,
<y#> jest rzeczownikiem w języku angielskim,
N jest rzeczownikiem w języku polskim

<x#> i <y#> łączy relacja międzyjęzykowej synonimii międzyparadygmatycznej typu "przypominający" jeśli:

<x#> określa właściwość fizyczną <y#>

<x#> łączy relacja synonimii międzyparadygmatycznej lub derywacyjności z N

N i <y#> łączy relacja międzyjęzykowa



Synonimia międzyparadygmatyczna (3)



<x#> jest przymiotnikiem w języku polskim, <y#> jest rzeczownikiem w języku angielskim,

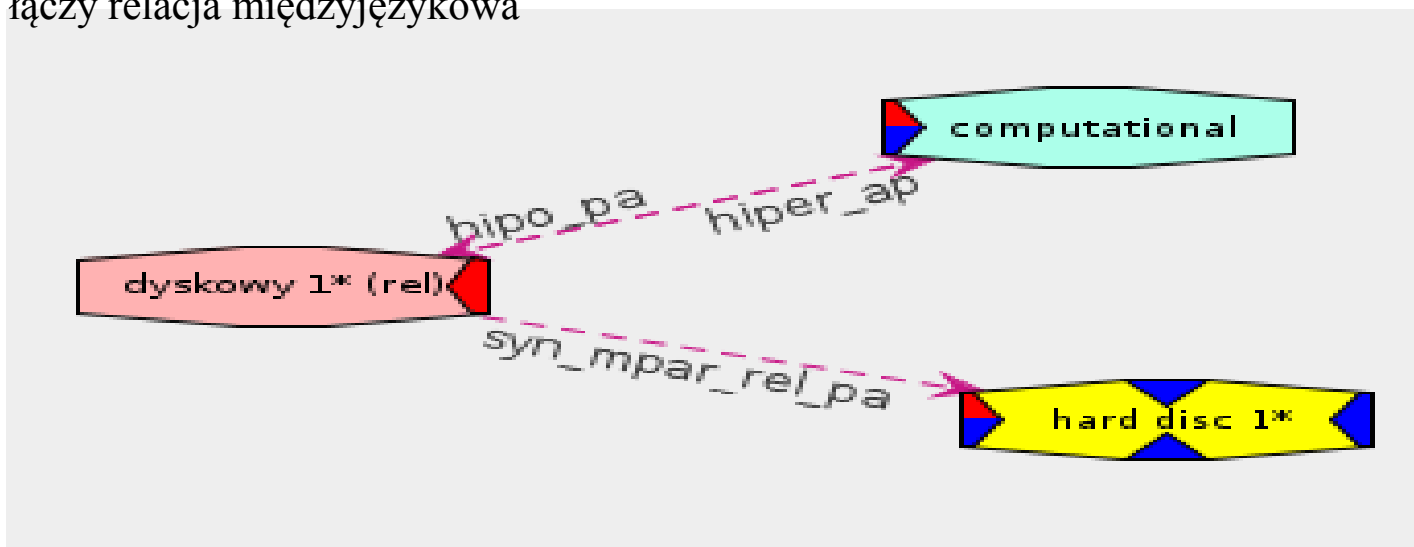
N jest rzeczownikiem w języku polskim

<x#> i <y#> łączy relacja międzyjęzykowej synonimii międzyparadygmatycznej typu "odnoszący się do" jeśli:

<x#> określa właściwości <y#>, które nie są jego cechami fizycznymi lub nie określają materiału, z którego jest wykonany

<x#> łączy relacja synonimii międzyparadygmatycznej lub derywacyjności z N

N i <y#> łączy relacja międzyjęzykowa

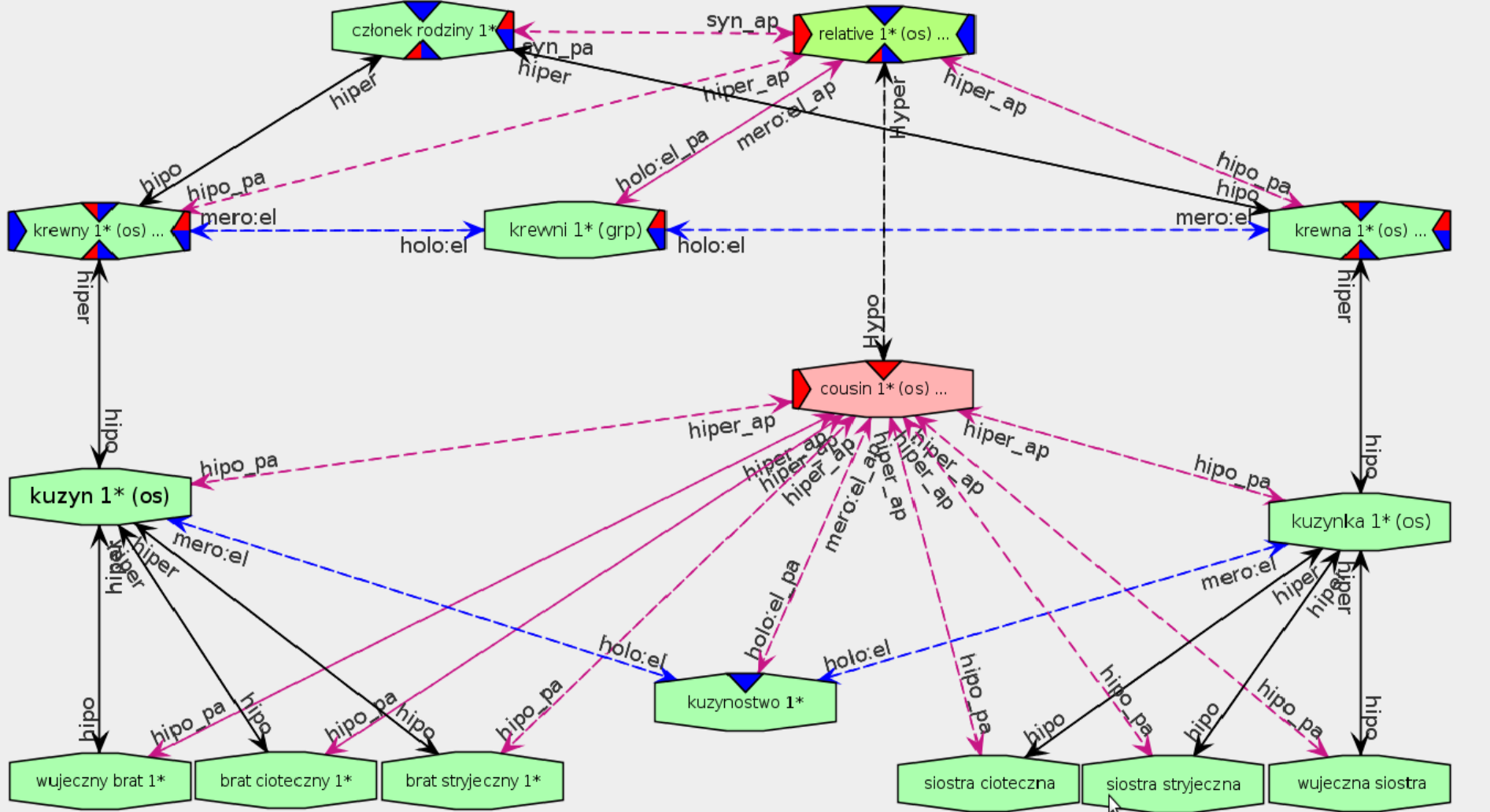


Relacje międzyjęzykowe



- **Hiponimia i hiperonimia** zdefiniowane są w ramach zawierania się denotacji zbiorów: hiponim odnosi się do obiektu, który zawiera się w zbiorze denotatów hiperonimu
- **Meronomia międzyjęzykowa** – oznacza część większej całości
- **Holonimia międzyjęzykowa** – oznacza całość składającą się z mniejszych części

Hyponimia i meronimia



Procedura rzutowania (1)



- **Rozpoznanie znaczenia synsetu źródłowego:**
 - miejsce w strukturze sieci - wszystkie istniejące relacje z naciskiem na hiperonim(y) i hiponimy;
 - definicje, komentarze;
 - porównanie pozostałych synsetów zawierających dany lemat)
- Przykład:
{zagranica 1, obczyzna 1, obce terytorium 1}:
 - jest hiponimem {obszar 1, terytorium 1, obręb 1, strefa 1, zona 1, rejon 3} komentarz: 'ograniczona część przestrzeni, zwykle dużych rozmiarów, określona powierzchnia czegoś (np. obszar państwa, obszary niezalesione)
 - jest meronimem {świat 3, nieznanne 1} komentarz: 'iść w świat'
 - jest w relacji fazynimii z {granica państwa 1}

Procedura rzutowania (2)



- **Poszukiwanie synsetu docelowego:**

wytypowanie kandydatów na synset docelowy

przy wykorzystaniu intuicji, podpowiedzi automatycznych oraz słowników

{foreign country 1} 'any state of which one is not a citizen' –
'każde państwo którego nie jest się obywatelem'

jest hiponimem {state 1, nation 1, country 1, land 9,
commonwealth 2, res publica 1, body politic 1} -

'a politically organized body of people under a single
government' – polityczna organizacja grupy ludzi skupiona
pod jednym rządem'

Procedura rzutowania (3)



- **weryfikacja kandydatów na synset docelowy:**
porównanie struktur hipero i hiponimicznych (i innych jeśli istnieją) z synsetem źródłowym
- sprawdzenie istniejących i/lub potencjalnych relacji międzyjęzykowych;
- definicje, komentarze; słowniki
np. {state 1, ..} jest międzyjęzykowym hiponimem

{państwo 1, kraj 1} - 'zorganizowana politycznie społeczność, zamieszkująca określone terytorium, z niepodległą formą rządów'

Procedura rzutowania (3)



- **Wytypowanie synsetu docelowego oraz relacji międzyjęzykowej:
{foreign country 1}**

Synonimia – nie (inne znaczenie, struktury i relacje)

Hiponimia – nie (znaczenie, struktury i relacje nie kwalifikują jako rodzaj)

Meronimia – tak (znaczenie, struktury i relacje kwalifikują jako część)

Połączenie synsetu źródłowego z synsetem docelowym

Relacje międzyjęzykowe



Relacje międzyjęzykowe (synsety): 143 721

	Rzeczowniki	Przymiotniki
● Synonimia	31 689	3 484
● Hiponimia	65 045	14 330
● Hiperonimia	4 031	49
● Meronimia	6 366	
● Holonimia	2 102	
● Synonimia częściowa	3 240	1 211
● Synonimia międzyrejestrowa	1 603	44
● Synonimia międzyparadygmatyczna		10 484
● Suma	114 076	29 602

Jeszcze niezrzurowane



- **Niezrzurowane synsety Słowosieci:**

- Rzeczowniki: 13 225

- Przymiotniki: 15 533

- Czasowniki: 21 596

- Total: 50 354**

- **Niezrzurowane synsety Princeton WordNet:**

- Rzeczowniki: 44 514

- Przymiotniki: 13 064

- Czasowniki: 13 789

- Total: 71 367**

Praca tłumacza - sytuacje



- Sytuacja 1 – tłumacz styka się z nowym dla siebie słowem w języku źródłowym
- Sytuacja 2 – tłumacz zna znaczenie słowa w języka źródłowym, ale nie zna jego odpowiednika w języku docelowym
- Sytuacja 3 – tłumacz ma kandydatów na odpowiednik w języku docelowym, ale nie potrafi wybrać najlepszego

Przykład 1 (1)



- ‘chapel’

“We have services for all of them. Or maybe you don't believe in any God. In that case you go to the library during **chapel**. But we don't have any Buddhist.”

Eagle Station, Autor: Mark Berent

https://books.google.pl/books?id=_Q-4BQEp3mMC&pg=PT3&pg=PT3&dq=this+is+%22chapel+formation%22&source=bl&ots=KTJBbLI-Mc&sig=PdCWg77aFtIz83Hx0Ynk0U1tFx0&hl=pl&sa=X&ei=kwIzVb7IIYHT7Qa29oCwCQ&ved=0CDUQ6AEwAw#v=onepage&f=false

Przykład 1 (2)



- Właściwy synset PWN:

{chapel service 1, chapel 2} (czy)

'a service conducted in a place of worship that has its own altar;

he was late for chapel'

- Hiperonimia do
{service 3, religious service 1, divine service 1}
- Synonimia międzyjęzykowa do {nabożeństwo 1}

Przykład 2 (1)



- **'fracht'**

“Jeśli polski przewoźnik załaduje **fracht** w Berlinie a miejscem przeznaczenia jest Poznań, to nie mamy do czynienia z kabotażem tylko znów z przewozem międzynarodowym.”

<http://famos.net.pl/kabotaz-w-ue/>

“Stanowiły one bowiem, iż w takim wypadku zleceniodawca uprawniony będzie do obciążenia przewoźnika karą umowną **w** wysokości od 10 do 100 % **frachtu**.”

<http://www.transportoweprawo.pl/2013/kara-umowna-musi-byc-precyzyjnie-okreslona>

“Rzetelnie i kompleksowo obsługujemy wszelkie wyzwania jakie stawia przed nami **fracht morski**.”

Przykład 2 (2)



- Jednostki Słownosieci:

fracht 1 (pos)

fracht 2 (pos)

fracht 3 (zwz)

Przykład 2 (3)



- Synsety Słowosieci:

{**fracht 1**} (pos): hiperonimia do {ładunek 3};

Międzyjęzykowa synonimia częściowa do:

{cargo 1, lading 1, freight 1, load 3, loading 4, payload 2, shipment 1, consignment 1} hyponym {merchandise 1}

- {**fracht 2**} (pos): hiperonimia do {opłata 1},
hiponimia do {fracht dystansowy 1};

Synonimia międzyjęzykowa do {freight 3, freight rate 1, freightage 1}
hiperonimia do {rate 2}

- {**fracht 3**} (zwz): hiperonimia do {żegluga 1};

Synonimia międzyjęzykowa do {freight 2, freightage 2}

hiperonimia do {transportation 5}

Przykład 2 (4)



- “Flights containing **unaccompanied freight** from Somalia will be suspended in the wake of the cargo plane terror plot, Home Secretary Theresa May has said.”

<http://www.telegraph.co.uk/news/uknews/terrorism-in-the-uk/8102883/Al-Qaeda-pl-t-flight-ban-on-freight-from-Somalia.html>

- “Hauling of **weighty freight**, containing means of transport, was accomplished by the world’s greatest operative chopper, the Mil Mi26|Mi-26.”

<https://books.google.pl/books?id=v9oPBwAAQBAJ&pg=PT21&lpg=PT21&d%22freight+containing%22&source=bl&ots=15576nnqhc&sig=GzflXN9AUZY1BWEE7Qo5Ah4qock&hl=en&sa=X&ei=SLInVdv0IIHnsgGQnYP4CQ&ved=0CDQQ6A#v=onepage&q=%22freight%20containing%22&f=false>

- “Demand for container ships has pushed up **freight rates.**”

<http://www.bbc.com/news/business-10999044>

- The parties may however so express the contract that the payment of all or

Przykład 2 (5)



- If a Polish carrier moves **freight** from Berlin and the destination is Poznań, we have to do not with cabotage, but again with **international freight**.
- Popyt na kontenerowce zwiększył **stawki opłat frachtowych**.
-

Przykład 3 (1)



'fafle'

(...) w zamyśleniu, z opuszczoną czarną głową przypominał psa myśliwskiego z **obwisłymi faflami.**” Borelowski Jan Maria Gisges, p. 70

https://books.google.pl/books?id=j-pIAAAAIAAJ&q=fafle+psa+my%C5%9Bliwskiego&dq=fafle+psa+my%C5%9Bliwskiego&hl=en&sa=X&ei=B1Aqf-CCeXcywOIIIBY&redir_esc=y

Jednostki Słowosieci:

fafle 1(czc)

Synsety Słowosieci:

{fafle 1} (czc): hiperonimia do {warga 1} (czc),
meronimia część do {pies myśliwski 1} (zw),

Międzyjęzykowa synonimia do {flews 1} (czc)

'the fleshy hanging upper lip of a bloodhound or similar dog'
hiperonimia do {overlip 1} (czc)

<https://books.google.pl/books?id=TFEiJ7FtJ0sC&pg=PA52&dq=flews+h>

Przykład 3 (2)



- 'His jowls hung in flaps like the **flews** of a hound, and he held the long gun behind him, run down back of his leg.'
Cold Mountain (przykład z wordnik.com)
- “Its eyes are one of the Bloodhound's most well-known characteristics. They are deeply sunk in the orbits, and the lids have a lozenge or diamond shape. This is because the lower lids are dragged down and turned outward by **the heavy flews**. This can cause eye problems if the flews are too heavy, so a good breeder goes for moderation rather than exaggeration.”

Bloodhounds: Everything about Purchase, Care, Nutrition, Breeding, Behavior, and Training, Kim Campbell Thornton, Michele Earle-Bridges, p. 52

Przykład 4 (1)



- 'grzebień'

“Jednym z przyjaciół Frediego był kogut. Był to odważny, dumny ptak o imieniu Klemens. Miał on zawsze pierś dumnie wypiętą do przodu, dziób w górze i **sterczący grzebień.**”

Magiczna bajaderka: Antologia bajek p. 184

https://books.google.pl/books?id=cDdiBgAAQBAJ&dq=magiczna+bajaderka&source=gbs_navlinks_s

- Jednostki Słowosieci:

grzebień 1 (wytw)

grzebień 2 (**czc**)

grzebień 3 (**czc**)

grzebień 4 (wytw)

grzebień 5 (jedz)

grzebień 6 (**czc**)

grzebień 7 (wytw)

grzebień 8 (**czc**)

Przykład 4 (2)



- Synsety Słownosieci:

{grzebień 2} (czc)

'wyrostek na głowie niektórych ptaków i gadów (u gadów może to być szereg wyrostków nie tylko na głowie, ale też wzdłuż grzbietu i ogona)'

hiperonimia do {wyrostek 1} (czc),

meronimia części do {dziób 5} (czc)

{grzebień 3} (czc) hiperonimia do {sierść 1}

{grzebień 6} (czc)

meronimia części do {mostek 3}

{grzebień 8} (czc)

hiperonimia do {element anatomiczny 1}

meronimia części do {kość 1}

Przykład 4 (3)



- {grzebień 2} (czc)

Synonimia międzyjęzykowa do {crest 5}

'a showy growth of e.g. feathers or skin on the head of a bird or other animal'

hiperonimia do {process 5} (czc),

hiponimia do {comb 2} (zw), {topknot 2} (zw), {tuft 2} (zw)

“Now the rooster is an entirely different sort of bird from the hen. It is very proud and has a red **crest** on the top of his head.”

The Benchley Roundup: A Selection by Nathaniel Benchley of His Favourites, Robert C. Benchley, Nathaniel Benchley, p. 95

https://books.google.pl/books?id=YxIA4q6AMNkC&dq=crest+rooster&source=gbs_navlinks_s

Przykład 5 (1)



- **'blistr/blister'**

“W stanach przewlekłych stosować można „wcierki” rozgrzewające, lub **blistry**.”

http://starywww.up.wroc.pl/polish/struktura/wet/kkc/cwiczenia/ortopedia_kop.html

- Jednostki Słownosieci:

blistr 1 (sbst)

blister 1 (wytw)

blister 2 (il)

- Synsety Słownosieci:

{blister 1} (wytw) hiperonimia do {opakowanie 2}

{blister 2} (il) hiperonimia do {zawartość 3}

{blistr 1} (sbst): hiperonimia do {maść 1};

Synonimia międzyjęzykowa do {blister 4}

hiponimia do {liniment 1}

Przykład 5 (2)



- “Harltons Liquid **Blister** is a counter irritant with penetrating action for use in the blistering of horses legs.”

<http://www.allvetsupply.com/halibl12.html>

“If a paint or **blister** has been used, the leg should not be bandaged as the chemicals under the bandages may severely irritate the horse’s skin even more.”

https://books.google.pl/books?id=i_iN_eyz0NMC&pg=PA223&lpg=PA223&dq=apply+blister+horses+legs&source=bl&ots=vlf-LndrwT&sig=4B8w-f9P_IAAQQjZVP3gtJ6BMx4&hl=en&sa=X&ei=38MqVdPAKoPMygPXvIHoAg&ved=0CF4Q6AEwCA#epage&q=apply%20blister%20horses%20legs&f=false

Bibliografia



- Fellbaum, Ch. (ed). (1998). *WordNet: An Electronic Lexical Database*. MIT Press: Cambridge, Massachusetts.
- Maziarz, M., Piasecki, M. and S. Szpakowicz. (2012). 'Approaching plWordNet 2.0.' *Proceedings of the 6th Global Wordnet Conference*, Matsue. pp. 189-196. accepted for publication.
- Piasecki, M., Szpakowicz, S. and B. Broda. (2009). *A WordNet from the Ground Up*. Oficyna Wydawnicza Politechniki Wrocławskiej: Wrocław.
- Princeton WordNet <http://wordnet.princeton.edu/wordnet/>
- Rudnicka, E., Maziarz, M., Piasecki, M., & Szpakowicz, S. (2012) 'A Strategy of Mapping Polish WordNet onto Princeton WordNet'. In *Proceedings of COLING 2012*. ACL.
- SłowoSieć <http://plwordnet.pwr.wroc.pl/wordnet/>

CLARIN

Common Language Resources and Technology Infrastructure



Dziękuję bardzo za uwagę
