

Dwujęzyczna Słownosieć



Ewa Rudnicka
Politechnika Wrocławska
Katedra Inteligencji
Obliczeniowej
Grupa Naukowa G4.19
ewa.rudnicka@pwr.wroc.pl

Dwujęzyczna Słownosieć



- Wielka **polsko-angielska** sieć leksykalno-semantyczna
- Powstaje w wyniku:
 - powiązania **Słownosieci** z **WordNetem princetońskim** i jego rozszerzeniem **enWordNetem**
- Obecnie największy słownik polsko-angielski **dostępny** na całkowicie **otwartej licencji**
- Stanowi **bramę do wielojęzycznych zasobów** europejskich i światowych (Platforma Leksykalna, Open Multilingual WordNet)

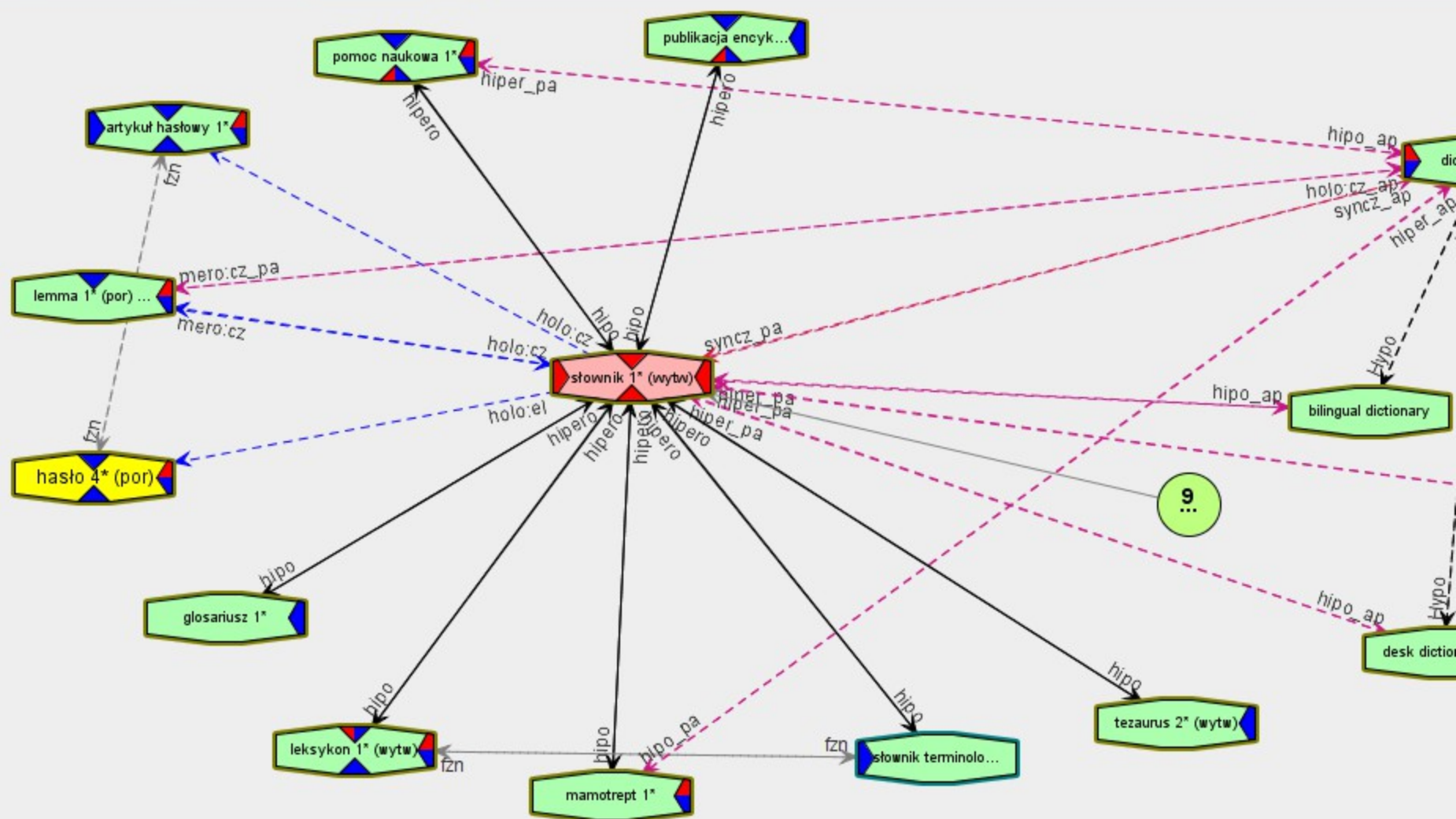
Relacje jednostek

Relacje synsetów

Edycja synsetów

Wizualizacja wordne...

+ hasło #0 x lemat #0 x słownik #0 x



Zastosowania Dwujęzycznej Słownosieci (1)



- nauka i dydaktyka języków obcych
- budowa narzędzi dydaktycznych do nauki języków obcych (np. Cloud Net, dr Monika Madej UŁ)
- budowa słowników dwu i wielojęzycznych: (np. Ling.pl, Glosbe, Miłkowski, BabelNet, Kamusigold.org, PONS)
- wsparcie w przekładzie i tłumaczeniu maszynowym (Google Translate)
- językoznawcze badania porównawcze

Zastosowania Dwujęzycznej Słownosieci (2)



- inspiracja dla rozwoju wordnetów dla innych języków (np. DanNet, Portuguese WordNet)
- Integracja z Open Multilingual WordNet (wielką wielojęzyczną platformą skupiającą wordnety dostępne na otwartej licencji (ponad 100 języków))
- Półautomatyczne rzutowanie ontologii SUMO na Słownosieć
- Półautomatyczne rzutowanie różnych tezaurusów dziedzinowych na Słownosieć (projekt AZON)
- Dwujęzyczne ujednoznacznianie znaczeń (WSD)

06418901-n 'a reference book containing an alphabetical list of words with information about them':

Search WN

| | |
|------------------------------|-----------------------------------------------------------------------------------------------------|
| Bulgarian | <i>речник</i> |
| Catalan | <i>diccionari , lèxic</i> |
| Chinese (simplified) | <i>词典 , 字典</i> |
| Danish | <i>ordbog</i> |
| Greek | <i>λεξικό</i> |
| English | <i>dictionary₅₂ , lexicon (↗)</i> |
| Farsi | <i>واژه نامه , فرهنگ , کتاب لغت , قاموس</i> |
| Finnish | <i>sanasto , sanakirja</i> |
| French | <i>lexique , dictionnaire , tableau associatif</i> |
| Galician | <i>dicionario , lexicón</i> |
| Croatian | <i>rječnik</i> |
| Indonesian | <i>leksikon , kamus , bausastra</i> |
| Icelandic | <i>orðasafn , orðabók</i> |
| Italian | <i>lessico , dizionario , vocabolario , calepino</i> |
| Japanese | <i>辞彙 , 辞典 , 辞林 , 辞書 , レキシコン , 字引き , 字典 , 字引 , 字林 , 字書 , 語典 , 事典 , 辭典 , ディクシ</i> |
| Lithuanian | <i>žodynas</i> |
| Dutch | <i>vertaalwoordenboek , woordenboek , lexicon , dictionaire , de wereld ligt aan iemands voeten</i> |
| Nynorsk | <i>ordbok</i> |
| Bokmål | <i>ordbok</i> |
| Polish | <i>słownik</i> |
| Portuguese | <i>array associativo , vocabulário , léxico , Dicionario , Dicionários , dicionário</i> |
| Chinese (traditional) | <i>詞典</i> |
| Romanian | <i>dicționar , lexicon , vocabular</i> |
| Slovak | <i>lexikón , slovník</i> |
| Slovene | <i>slovar</i> |
| Spanish | <i>léxico , diccionario</i> |
| Swedish | <i>ordbok</i> |
| Thai | <i>ปทานุกรม , ดิก , พจนานุกรม , ดิกชันนารี</i> |
| Malaysian | <i>leksikon , kamus</i> |

Open source components and licenses.

We use the following licenses for each language listed below.

| Language | Wordnet | License |
|----------------|----------------------------------|-----------------|
| Albanian | Albanet | CC-BY 3.0/GPL 3 |
| Arabic | Arabic Wordnet | CC-BY-SA 3 |
| Chinese | Chinese Wordnet (Taiwan) | Wordnet |
| Danish | DanNet | Wordnet |
| English | Princeton Wordnet | Wordnet |
| Farsi/Persian | Persian Wordnet | Free to use |
| Finnish | FinnWordNet | Wordnet |
| French | WOLF (Wordnet Libre du Francais) | CeCILL-C |
| Hebrew | Hebrew Wordnet | Wordnet |
| Italian | MultiWordNet | CC-BY-3.0 |
| Japanese | Japanese Wordnet | Wordnet |
| Catalan | Multilingual Central Repository | CC-BY-3.0 |
| Galecian | Multilingual Central Repository | CC-BY-3.0 |
| Spanish | Multilingual Central Repository | CC-BY-3.0 |
| Indonesian | Wordnet Bahasa | MIT |
| Malaysian | Wordnet Bahasa | MIT |
| Norwegian | Norwegian Wordnet | Wordnet |
| Polish | plWordNet | Wordnet |
| Portuguese | OpenWNET | CC-BY-SA-3.0 |
| Thai | Thai Wordnet | Wordnet |
| Italian Creole | MIT-Hati Initiative | CC-BY 4.0 |

Irish language data from Feras na Gaoidhe's New English-Irish Dictionary (English database designed and developed for Feras na Gaoidhe by Lexicography MasterClass Ltd.)

Welsh language data from Gwladur by Geiric.

Certain content is copyright Oxford University Press USA, some phrase translations come from Wikitravel.

Typy powiązań



- **Powiązania pomiędzy synsetami** (zbiorami synonimów) Słownosieci i WordNetu princetońskiego oraz enWordNetu
- Tylko czasem powiązanie 1-1 („hasło do hasła”):
 - {kompilator (os) 2} I-syn {compiler (os) 1}
- Często powiązania:
 - kilka – kilka:
 {glosariusz 1, glosarium 1} I-syn {glossary 1, gloss 2}
 - 1 – kilka: {hasło 4} I-syn {lexical entry 1, dictionary entry 1}
 - Kilka -1: {lemma 1, forma podstawowa} I-syn {lemat 1}

Proces konstrukcji

- **Rzutowanie** tj. powiązanie dwóch sieci na podstawie podobieństw w ich strukturze poprzez łączenie odpowiadających sobie fragmentów grafów
- Wszystkie **powiązania** tworzone **ręcznie**
- Przy wsparciu systemów automatycznych podpowiedzi
- Szczegółowe **procedury rzutowania**, dedykowane dla różnych **części mowy**
- Bogaty zestaw **relacji międzyjęzykowych**

Relacje międzyjęzykowe (1)



- Inspirowane:
 - relacjami **międzyjęzykowymi** z **EuroWordNetu** (ang. *Equivalence Relations*, Vossen 2002)
 - **wewnętrznymi** relacjami **Słownosieci** (Maziarz et al. 2013)
- opracowano **definicje, testy podstawieniowe i hierarchię** wprowadzania (cf. Rudnicka et al. 2012)

Relacje międzyjęzykowe (2)

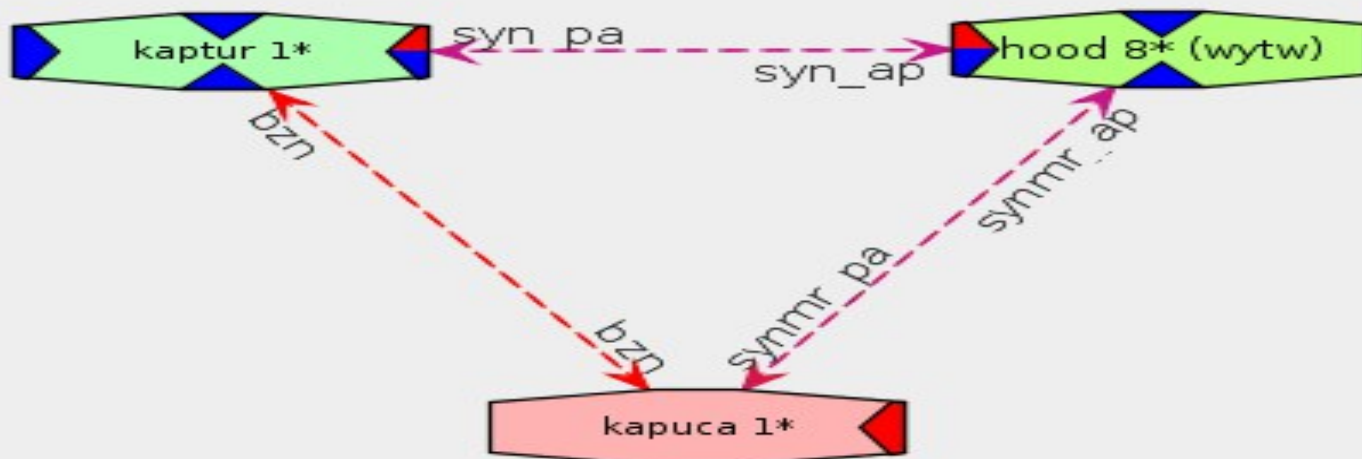


- Synonimia (rozumiana jako ekwiwalencja)
- Synonimia częściowa
- Synonimia międzyrejestrowa
- Hiponimia
- Hiperonimia
- Typ
- Egzemplarz
- Meronimia
- Holonimia
- Synonimia międzyparadygmatyczna
- Szczegółowe relacje czasownikowe

Synonimia



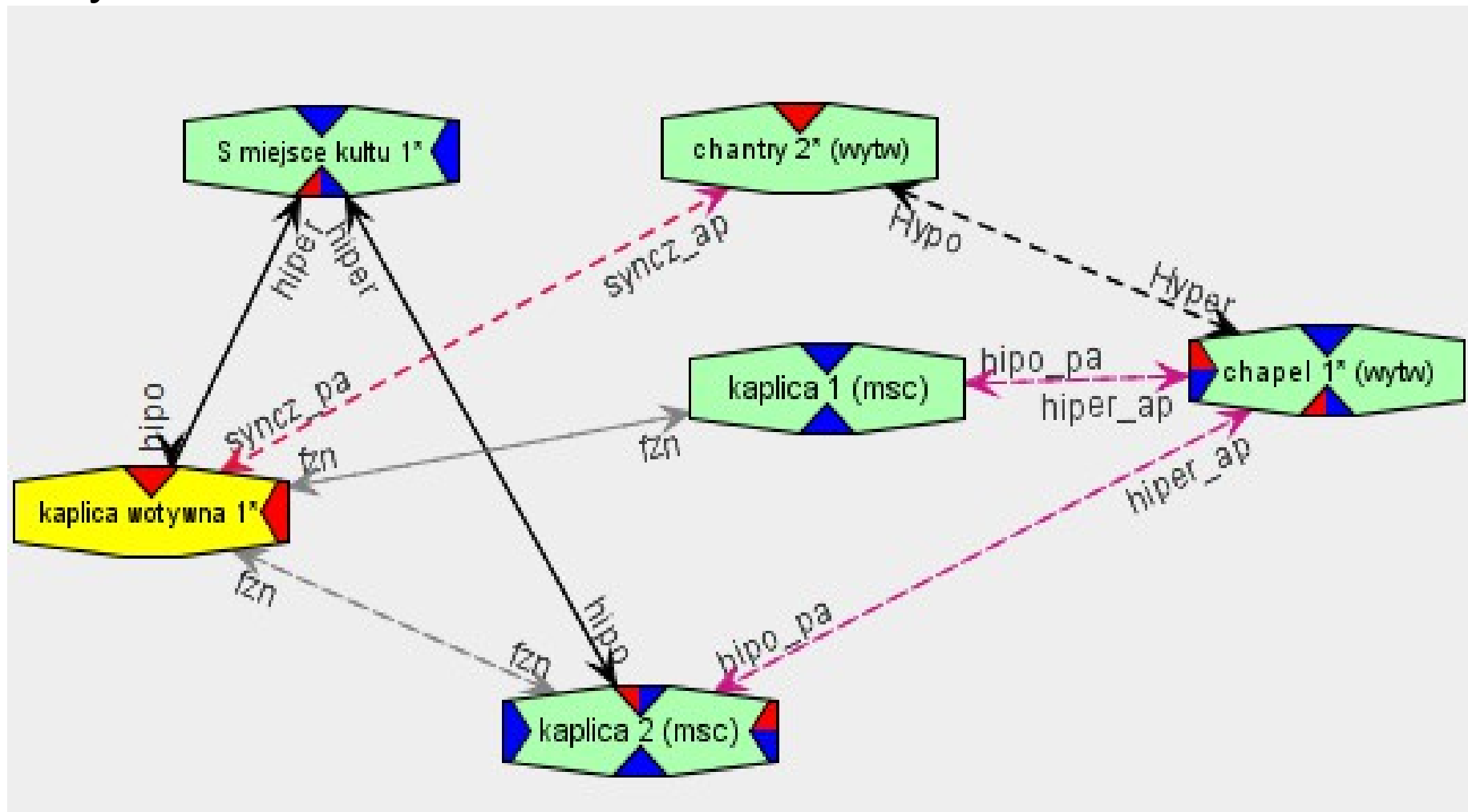
- **Synonimia** - oznacza dużą odpowiedniość znaczeń i struktur; traktowana jako podstawowa i priorytetowa relacja; może być tylko jedna dla jednego synsetu
- **Synonimia częściowa** - stosowana w przypadku częściowej odpowiedniości znaczeń i/lub struktur
- **Synonimia międzyrejestrowa** - dla odpowiedników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



Synonimia częściowa



Przykład

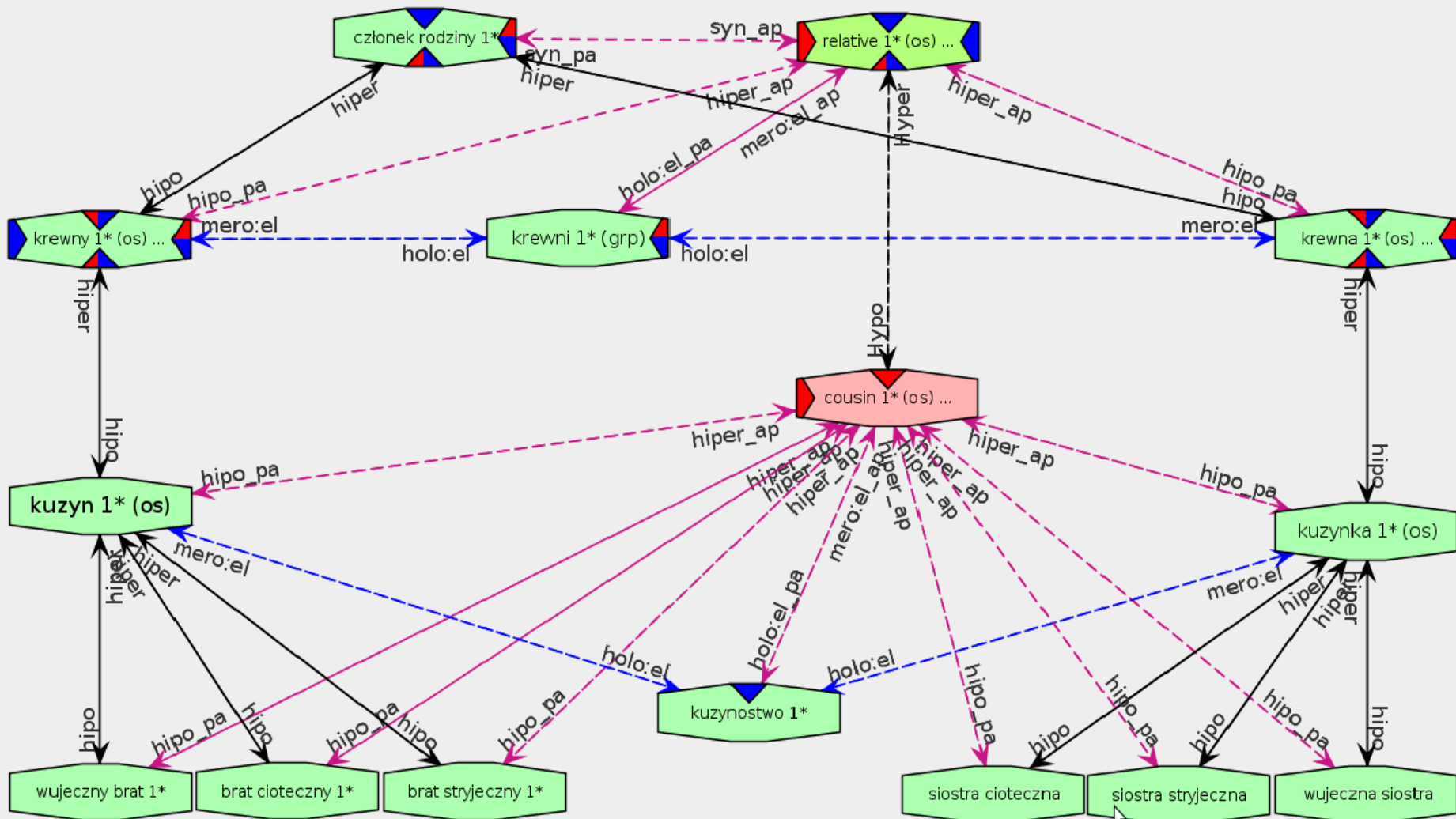


Relacje międzyjęzykowe (2)



- **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

Hiponimia i meronimia



Wnioski



- Wyraźny prymat synonimii i hiponimii nad pozostałymi relacjami międzyjęzykowymi
- Wyraźna, dwukrotna przewaga w liczności relacji hiponimii międzyjęzykowej nad synonimią międzyjęzykową
- Duża liczba synonimii międzyparadygmatycznej dla przymiotnika

- Kontynuacja uzupełniania rzutowania w kierunku ANG-PL oraz na bieżąco PL-ANG (wraz z rozbudową Słowsieci)
- Rozszerzenie rzutowania do poziomu jednostek leksykalnych (przy wykorzystaniu wyników pilotażowego projektu 'Ekwiwalencja w leksykalnych bazach danych' w ramach konkursu NCN Harmonia 7)
- Stworzenie siatki silnych i regularnych odpowiedników tłumaczeniowych
- Kontynuowanie budowy enWordNetu

Przykład 1 (1)



- **'grzebień'** – znane słowo, nieznanymi odpowiednikami

“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

- Jednostki Słowosieci:

grzebień 1 (wytw)

grzebień 5 (jedz)

grzebień 2 (czc)

grzebień 6 (czc)

grzebień 3 (czc)

grzebień 7 (wytw)

grzebień 4 (wytw)

grzebień 8 (czc)

Przykład 1 (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),

{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 1 (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, cockscomb 4, coxcomb 3} (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

Przykład 2 (1)



- Znane słowo (lub lemat) w nieznanym kontekście:

“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

- Jednostki leksykalne Princeton WordNet:

chapel 1 (wytw)

chapel 2 (czy)

Przykład 2 (2)



- Właściwy synset PWN:
{chapel service 1, chapel 2} (czy)
- Glosa i przykład:
'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}
'the act of public worship following prescribed rules'
Synonimia międzyjęzykowa do {nabożeństwo 1}

Przykład 3 (1)



- Nie znamy odpowiednika lub chcemy zweryfikować znany odpowiednik
- “W stanach przewlekłych stosować można „wcierki” rozgrzewające, lub blistry”
http://starywww.up.wroc.pl/polish/struktura/wet/kkc/cwiczenia/ortopedia_kop.html
- Jaką formę bazową wyszukamy?

Przykład 3 (2)



- Synsety Słowosieci:

- {blister 1} (wytw) hiperonimia do {opakowanie 2}
synonimia międzyjęzykowa do {blister pack 1, bubble pack 1}

- {blister 2} (il) hiperonimia do {zawartość 3}
hiponimia międzyjęzykowa do {containerful 1}

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

synonimia międzyjęzykowa do {blister 4}

hiponimia do {liniment 1}

Przykład 3 (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?](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=38MqVdPAKoPMygPXvIH0Ag&ved=0CF4Q6AEwCA#v=onepage&q=apply%20blister%20horses%20legs&f=false)

[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=38MqVdPAKoPMygPXvIH0Ag&ved=0CF4Q6AEwCA#v=onepage&q=apply%20blister%20horses%20legs&f=false](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=38MqVdPAKoPMygPXvIH0Ag&ved=0CF4Q6AEwCA#v=onepage&q=apply%20blister%20horses%20legs&f=false)



- Bond, F., Fellbaum, Ch., Hsieh, S., Huang, Ch., Pease, A. and P. Vossen (2014). A Multilingual Lexico-Semantic Database and Ontology. In *Towards the Multilingual Semantic Web* Paul Buitelaar and Philipp Cimiano (eds), Springer pp. 243–258. (Publisher's page: <http://compling.hss.ntu.edu.sg/who/bond/pdf/2014-msw-omw.pdf>)
- Fellbaum, Ch. (ed). (1998). *WordNet: An Electronic Lexical Database*. MIT Press: Cambridge, Massachusetts.
- Kędzia, P., Piasecki, M., Rudnicka, E., Przybycień, K. (2013). Automatic Prompt System in the Process of Mapping plWordNet on Princeton WordNet. *Cognitive Studies* 13: 123-141.
- 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.
- Rudnicka, E., Witkowski, W. & Kaliński, M. (2015). Towards the Methodology for Extending Princeton WordNet. *Cognitive Studies* 15.
- SłowoSieć <http://plwordnet.pwr.wroc.pl/wordnet/>
- Vossen, P. (ed). (2002). *EuroWordNet. General Document*. Amsterdam.

Najnowsze publikacje



- Rudnicka, E., Piasecki, M., Piotrowski, T. Grabowski, Ł., Bond, F. 2017. 'Mapping wordnets from the perspective of inter-lingual equivalence.' *Cognitive Studies* 17.
- Rudnicka, E., Bond, F., Grabowski, Ł., Piasecki, M., Piotrowski, T. 2017. 'Towards Equivalence Links between Senses in pIWordNet and Princeton WordNet. *Lodz Papers in Pragmatics*. Vol. 13, No 1.
- Rudnicka, E., Bond, F., Grabowski, Ł., Piasecki, M., Piotrowski, T. 2018. 'Lexical Perspective on Wordnet to Wordnet Mapping.' IX Global WordNet Conference, Singapur.
- W przygotowaniu:
- Rudnicka, E., Witkowski, W., Piasecki, M. 2018. 'enWordnet - a mapping based extension of Princeton WordNet'. *Linguistic Issues in Language and Technology*.

CLARIN

Common Language Resources and Technology Infrastructure



Dziękuję bardzo za uwagę
