Wordnet as a Relational Semantic Dictionary Built on Corpus Data

Maciej Piasecki, Agnieszka Dziob
Wrocław University of Science and Technology
G4.19 Research Group
maciej.piasecki@pwr.edu.pl
2019-09-03

Politechnika Wrocławska

CLARIN-PL
Common Language Resources and Technology Infrastructure

Plan

- The original idea of WordNet and its adoption
- Relational semantic dictionary as inspired by a wordnet
- Relational model for a wordnet (plWordNet)
  - synset definition
  - constitutive relations and features
- Non-relational elements of the wordnet structure
- Corpus-based wordnet development process
  - supporting tools: editing and semantic exploration of corpora
  - as a way to free the wordnet
- Wordnet is not enough – a system of resources
- plWordNet lesson – the model in practice
- plWordNet in use

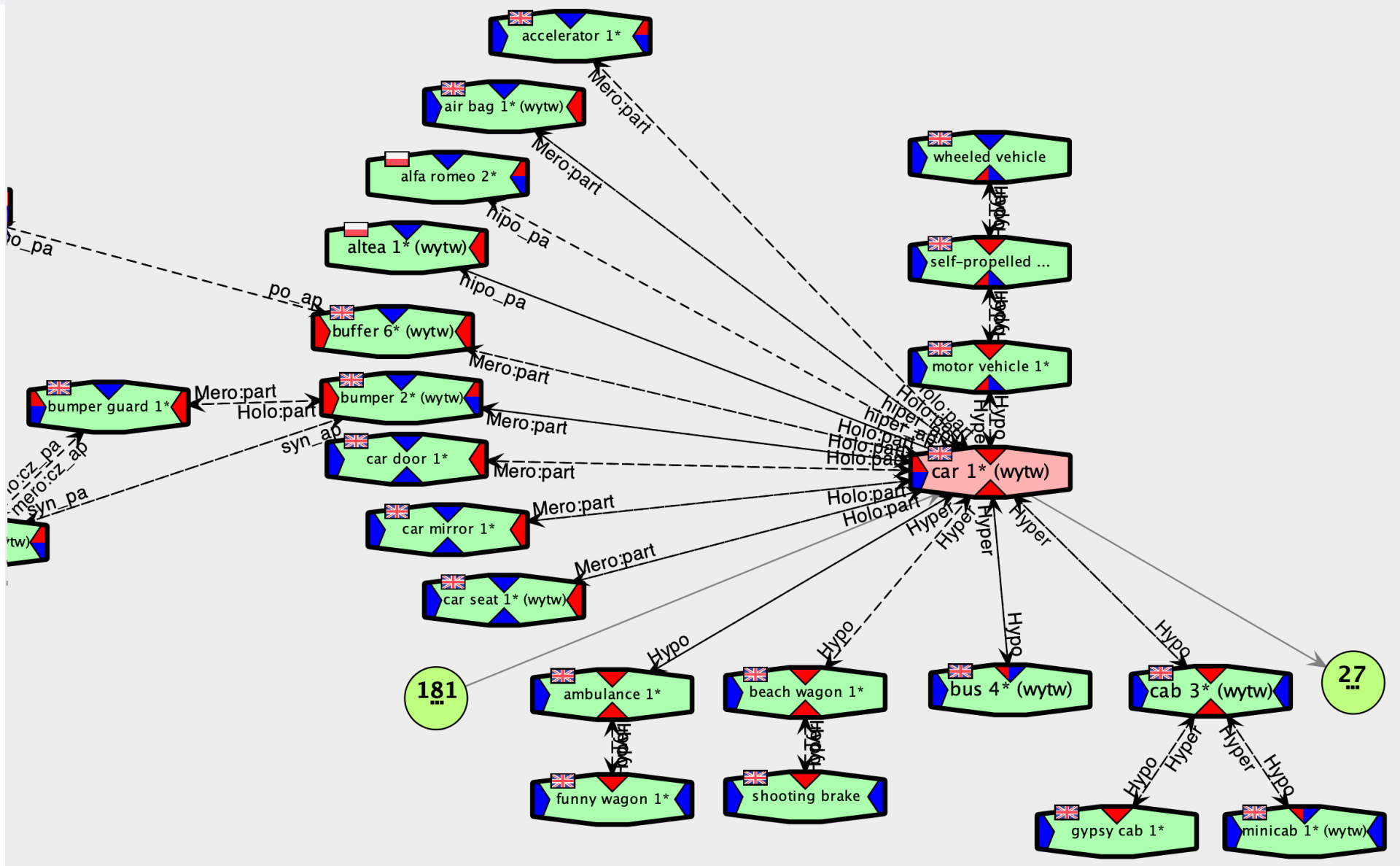
The idea of WordNet

- Princeton WordNet – the prototypical wordnet
- A psycholinguistic experiment on language acquisition by children
  - aimed to explain how lexical meaning is stored in the mind,
  - “WordNet is an on-line lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory.” Miller et al. (1993, p. 1)
- Developed into a lexico-semantic database
  - a network of **lexicalised concepts** (represented as **synsets** – synonym sets)
- Open licence and unbelievable carrier
  - “Princeton WordNet” > 7.5 mln. hits and “wordnet” > 8 mln. hits in Google
  - thousands of applications and citations

Princeton WordNet

- S: (n) **car#1**, auto#1, automobile#1, machine#6, motorcar#1 (a motor vehicle with four wheels; usually propelled by an internal combustion engine) *"he needs a car to get to work"*
  - ***direct hyponym*** / *full hyponym*
    - S: (n) ambulance#1 (a vehicle that takes people to and from hospitals)
    - S: (n) beach wagon#1, station wagon#1, wagon#5, estate car#1, beach waggon#1, station waggon#1, waggon#2 (a car that has a long body and rear door with space behind rear seat)
    - …
  - ***part meronym***
    - S: (n) accelerator#1, accelerator pedal#1, gas pedal#1, gas#5, throttle#2, gun#6 (a pedal that controls the throttle valve) *"he stepped on the gas"*
    - S: (n) air bag#1 (a safety restraint in an automobile; the bag inflates on collision and prevents the driver or passenger from being thrown forward)

Princeton WordNet



Relational semantic dictionary

- Underspecifications in the WordNet model
- Problematic notion of synonymy
  - difficulties with operationalised definition
- Synset – a lexicalised concept vs language data
  - use examples – illustrate `words' (or word senses)
  - contexts and collocations – related to `words'
  - occurrences of semantic relations – `words'
- Definitions
  - refer typically to `words' and `words' related to them
- Abstract concepts informally and implicitly defined

Relational semantic dictionary

- Lexical meanings
  - nodes in a graph of lexical-semantic relations
  - differentiated by properties of local graph structures
- Uniform format of meaning descriptions
  - subgraphs of lexico-semantic relations
- Formalised description (to some extent)
  - a graph structure
- Partial definition
  - limited by the types of lexico-semantic relations
  - constraints on the interpretation and use of word meanings
  - such meaning distinctions are not obvious for human users

plWordNet model

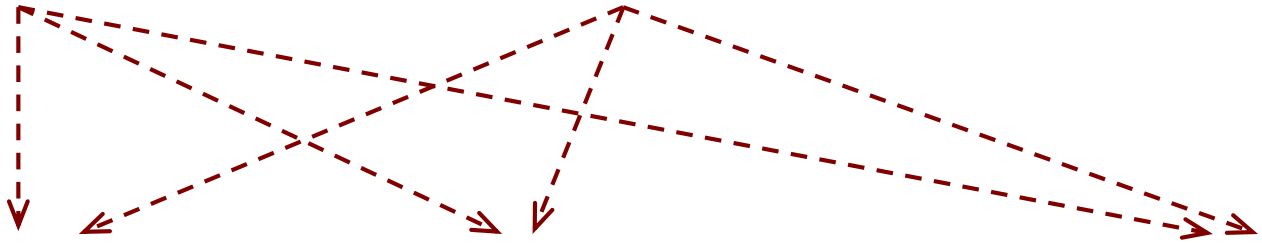
- Corpora contain words, senses discernible by context, not sets of synonyms
- Lexical unit (LU)
  - a triple: <part of speech, lemma, sense id>
  - the basic building block in plWordNet, belongs to one synset
- Synset – a group of lexical units which share
  - lexico-semantic *constitutive relations,* e.g. hyper/hyponymy, mero/holonymy
  - and *constitutive features*: stylistic register, aspect, and semantic classes for adjectives and verbs
- A relation between two synsets is a shorthand for sharing relations between lexical units

Synset

- Example

  {*wzór 1 `paragon'* ,*wzorzec 2 `*pattern*'*,...}

  —**hypernym**→



  {*idol 1 `idol'*, *bożyszcze 1 `~idol'* , *gwiazdor 1 ~`star'*}

- Synset as a notational convention
  - for a group of lexical units sharing certain constitutive relations
- What are wordnet constitutive relations?
- Are relations enough to define synsets?

Constitutive relations

- Required properties
  - well-established in linguistics
    - good understanding (e.g. paradigmatic relations)
    - existing descriptions
  - definable with sufficient specificity
  - and useful in generalisation
    - relatively frequent
    - should describe sets of lexical units systematically – a sharing factor
- Level of generalisation of a wordnet vs selection of the constitutive relations

Constitutive features

- Wordnet structure as a basis for acceptable conclusions
  - lack of formal definitions
  - some conclusions based on properties of relations, e.g. transitivity of hypernymy
- Additional constraints on the relation definitions
  - meta-conditions
  - obligatory and built into the relation definitions
- plWordNet: **stylistic registers**, **semantic verb classes** and **aspect**

plWordNet model

- ***Minimal Commitment Principle***
  - any particular theory of meaning is not assumed
  - minimal set of principles
  - all referring to facts that can be checked in the language data to as large extent as possible
  - a wordnet open to potentially many types of interpretations

Non-relational elements

- Types
  - constitutive features (attributes)
  - glosses (short definitions)
  - use examples
- Glosses
  - a genus term (dictionary) – a hypernym
  - set of differentiae, related words (dictionary) – partially by targets of other relations
  - Negatives: potential redundancy
  - Positives:
    - improved understanding by the editing team members and users
    - additional information for Word Sense Disambiguation
  - links to Wikipedia as an extension

Non-relational elements

- Use examples
  - verification of and support for the identified meaning
  - contexts as an extended description
  - very valuable for WSD
- Potential extensions
  - large number of examples
  - contextualised examples
  - sense disambiguated examples

plWordNet (Słowosieć): goal

To build a wordnet which provides a faithful and comprehensive description of the system of Polish lexical semantics

- its structure should represent accurately the lexico-semantic relations between lexical meanings in Polish
- and be motivated only by observations derived from Polish language data
- any form of translation from wordnets for other languages was excluded
- a resource with good coverage with respect to lemmas, word senses and instances of lexico-semantic relations
- in close correspondence to language data collected from very large corpora

Corpus-based wordnet development process

- A large text corpus is primary data
  - Lemmas (starting with the most frequent)
  - Examples of use and senses
- Language tools and systems support corpus exploration
  - simple, e.g. concordances
  - advanced – extraction of semantic similarity, relations, sense clusters
  - combined – semi-automated wordnet expansion (Paintball algorithm, RANLP 2013)
- Process
  - systematic extraction of lemmas, acquisition of lexico-semantic knowledge, generation of suggestions, decisions of editors
  - supported by: dictionaries, encyclopaedias, intuition, team

Corpus-based Wordnet Development

- Limited resources at the starting point
  - translation ruled out & no electronic monolingual dictionaries to leverage
- Schema
  1. A large corpus built from available sources
  2. Extraction of lemma frequency list
  3. Selection of new lemmas
  4. Building a Measure of Semantic Relatedness
  5. MSR-based clustering new lemma into packages
  6. Extraction of knowledge sources
  7. Wordnet editing supported by tools
     - Semi-automated wordnet expansions
     - Semantic exploration of corpora
     - Consulting traditional linguistic resources
  8. Linguistic work management

Corpus-based Wordnet Development

1. plWordNet Merged Corpus
   - available Polish corpora:
     - Corpus IPI PAN
     - Rzeczypospolita Corpus
     - Wikipedia (2015)
     - Texts on open licence
   - text collected from Internet
     - larger texts
     - Max. 20% tokens not recognised by Morfeusz analyser
   - The version 7.0: ~ 2 billion tokens
   - The version 10.0: **>4 billion tokens (for plWordNet 4.0)**
2.&3. Extraction of lemma frequency list and Selection
   - from the morpho-syntactically tagged and lemmatised corpus
   - necessary manual filtering
   - 7 000-9 000 new lemmas of a PoS in focus per iteration

Corpus-based Wordnet Development

4. Extraction of knowledge sources
   - Measure of Semantic Relatedness
   - relation instances (hypernymy) extracted by manually constructed lexico-syntactic patterns
   - relation instances extracted by more generic patterns developed in a remotely controlled process
   - ML-based classifier for relation instances
5. Semi-automated wordnet expansions
   - generation of suggestions for the placement of new lemmas in the wordnet structure
   - presented for final editing decisions by linguists
   - WordnetWeaver - an extension to WordnetLoom

Corpus-based Wordnet Development

4. Measure of Semantic Relatedness: coincidence matrix, tested and tuned
5. MSR-based clusters of lemmas (up to 200) -> assignment of task for linguists

wieczór

podobieństwo	jednostka leksykalna
0.206	popołudnie
0.192	noc
0.189	przedpołudnie
0.187	poranek
0.170	ranek
0.147	koncert
0.140	dzień
0.109	weekend
0.107	kolacja
0.107	gala
0.106	spotkanie
0.106	impreza
0.102	południe
0.101	niedziela
0.098	spektakl
0.096	uroczystość
0.094	chwila
0.092	obiad
0.092	sobota
0.091	biesiada

mężczyzna

podobieństwo	jednostka leksykalna
0.436	kobieta
0.365	człowiek
0.357	dziewczyna
0.332	chłopiec
0.314	młodzieniec
0.299	chłopak
0.278	facet
0.276	starzec
0.260	dziewczynka
0.248	osobnik
0.245	osoba
0.239	żołnierz
0.238	dziecko
0.217	strażnik
0.214	staruszek
0.211	policjant
0.203	człowieczek
0.201	staruszka
0.199	niewiasta
0.199	wojownik

nietoperz

podobieństwo	jednostka leksykalna
0.203	ptak
0.182	mewa
0.171	szczur
0.171	owad
0.169	sowa
0.160	jaszczurka
0.154	ćma
0.152	sęp
0.144	mysz
0.143	ropucha
0.138	gryzoń
0.136	wąż
0.133	gołąb
0.132	pszczoła
0.132	drapieżnik
0.130	komar
0.129	pająk
0.128	gad
0.127	małpa
0.126	żółw

antonym
hypernym
hyponym
co-hyponym
closely related
holonym

Corpus-based Wordnet Development

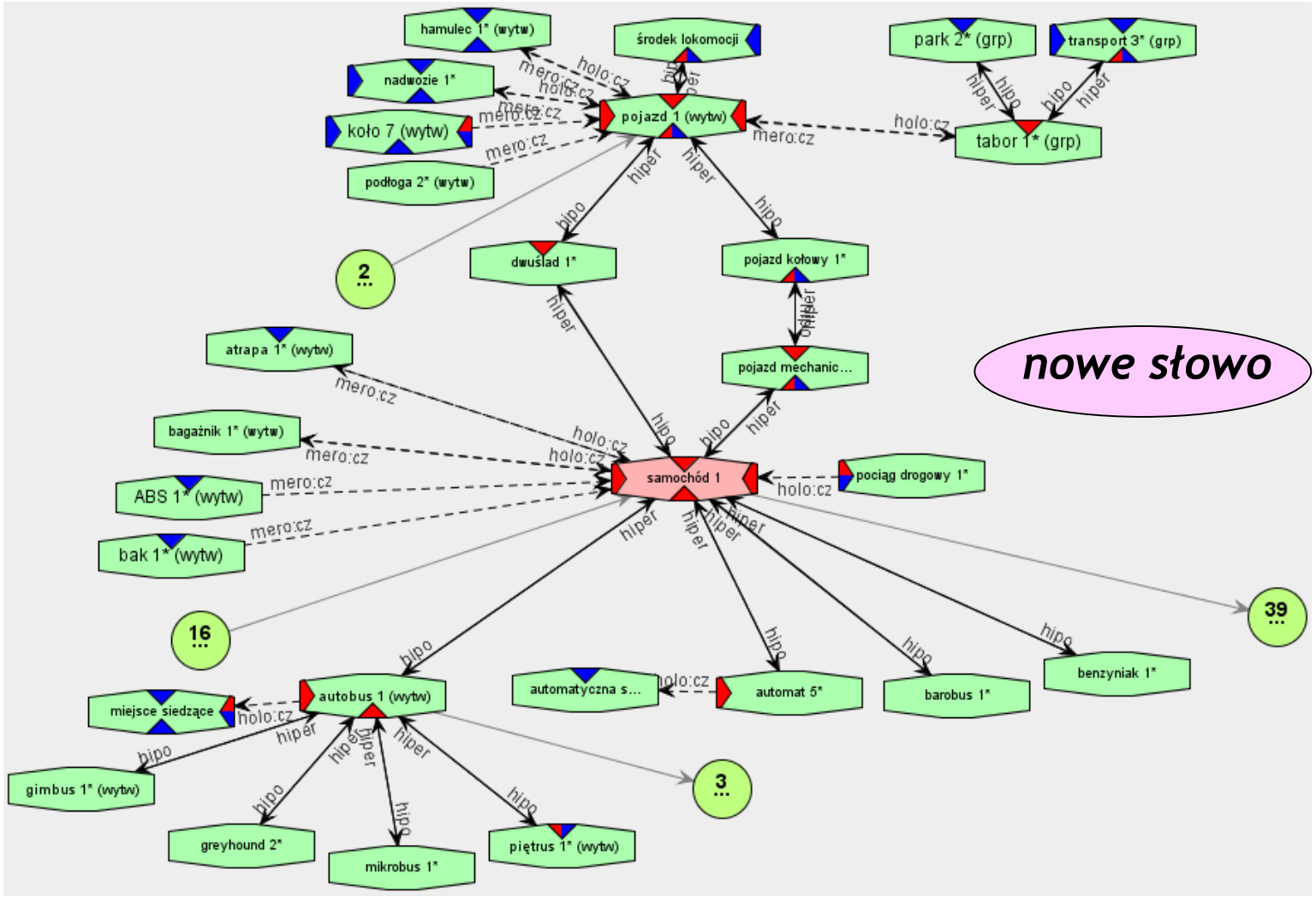
4. Measure of Semantic Relatedness: fastText on plWN Corpus 10.0, tuned
5. MSR-based clusters of lemmas (up to 200) -> assignment of task for linguists

pigment *pigment*
pigmentowy *pigmentary* ←
pigmen (*a typo*)
pigmentowo (Adv)~*pigmentary* ←
pigmentowany (Adj. Part.)
covered by pigment ←
pigmentowanie (Ger.) *covering by pig.* ←
pigmentcreme (*foreing neologism*)
aurypigment *(mineral)*
barwnik *dye* ←
pigmentacja *pigmentation* ←

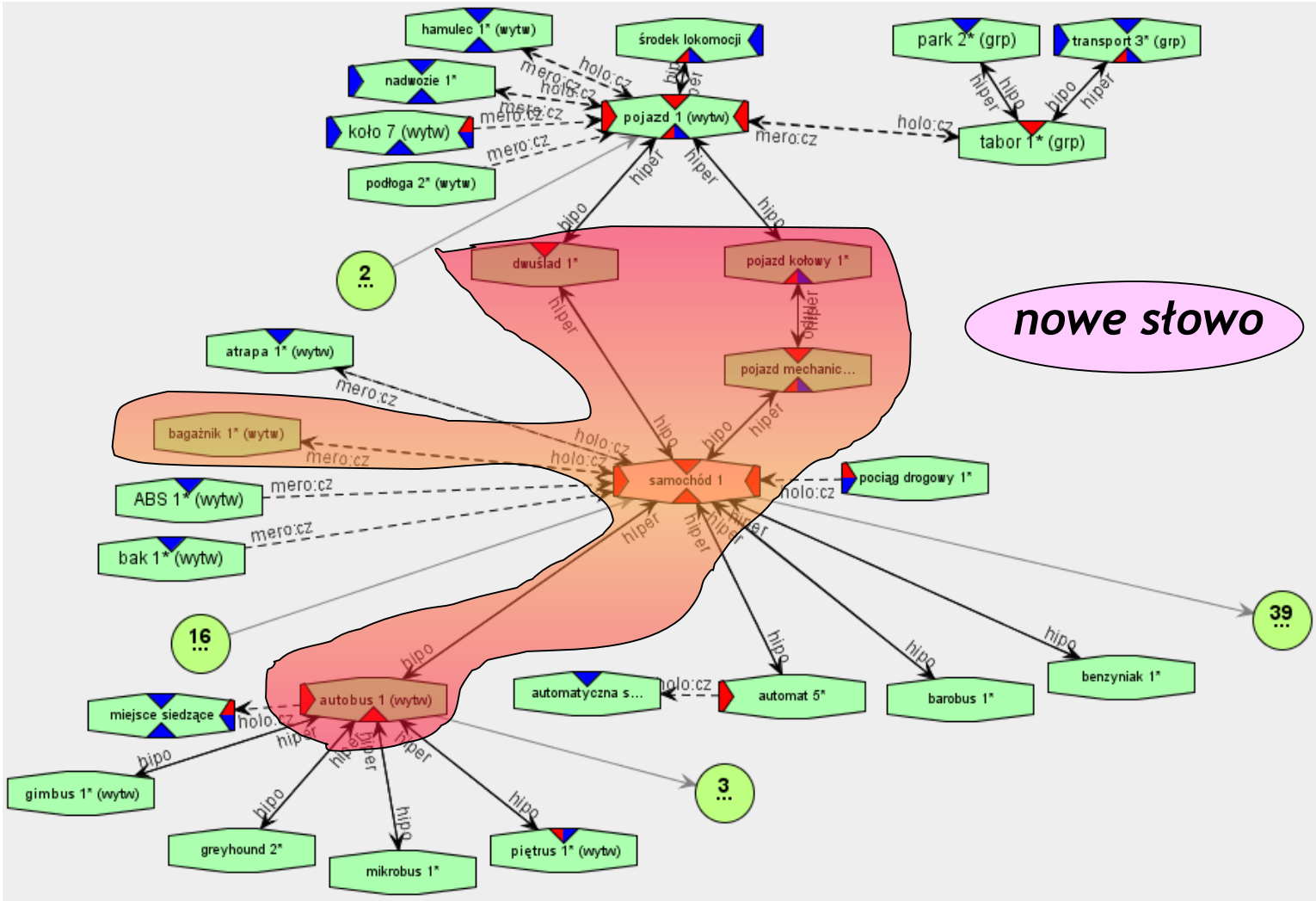
pies *dog*
piesek *pooch (deminutive)* ←
pieski *canine/wretched* ←
czworonóg *tetrapod* ←
szczeniak *puppy* ←
psiak *puppy/pooch (dem.)* ←
czworonogi *tetrapods* ←
doberman *doberman* ←
kundel *mongrel* ←
psisko *pooch* (*augmentative)* ←

← antonym ← hypernym ← hyponym ← co-hyponym
← closely related ← holonym

Paintball algorithm: initial state



Paintball algorithm: spreading activation



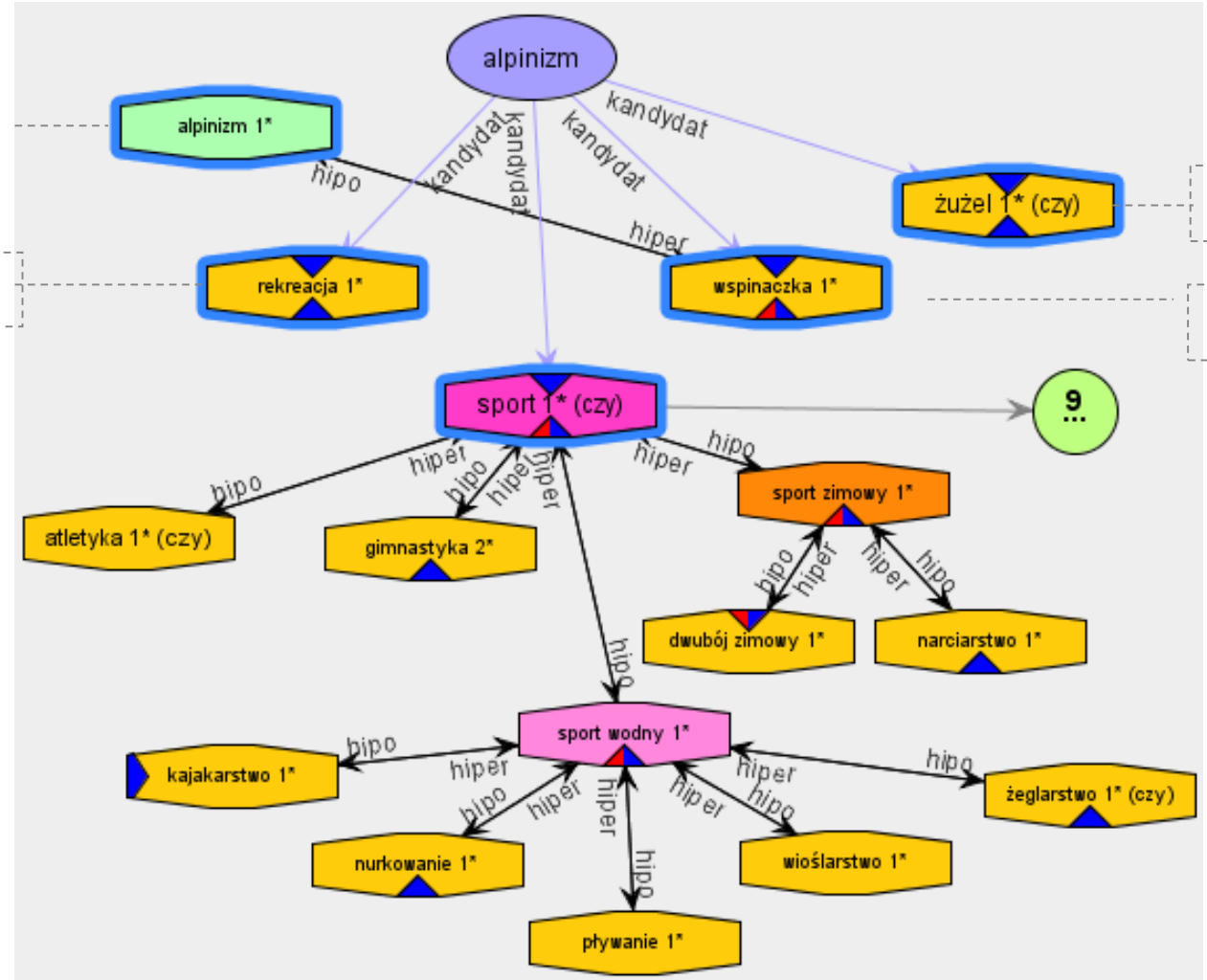
WordnetWeaver - Semi-automated Wordnet Expansion

alpinisim

recreation

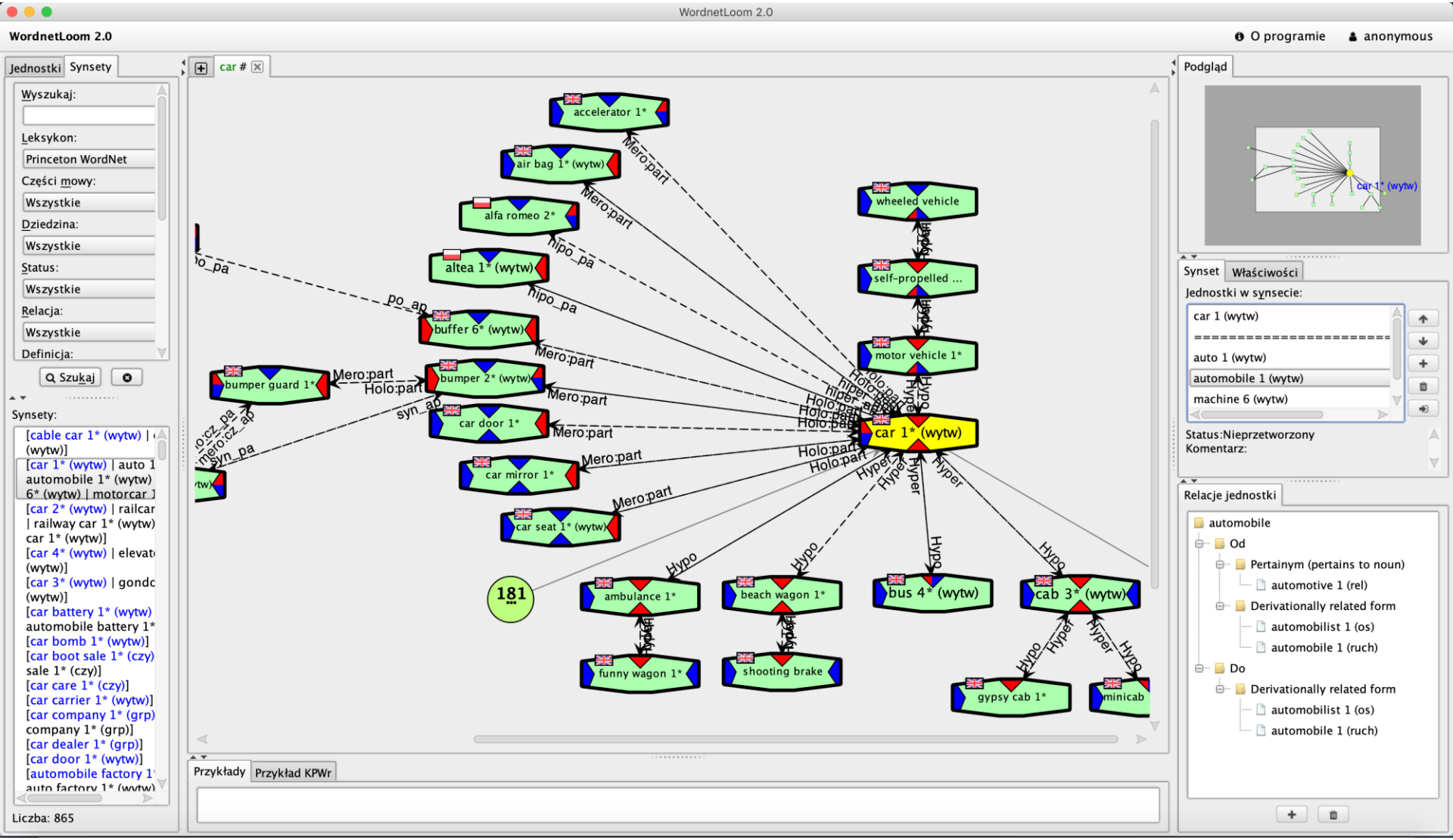
speedway

climbing



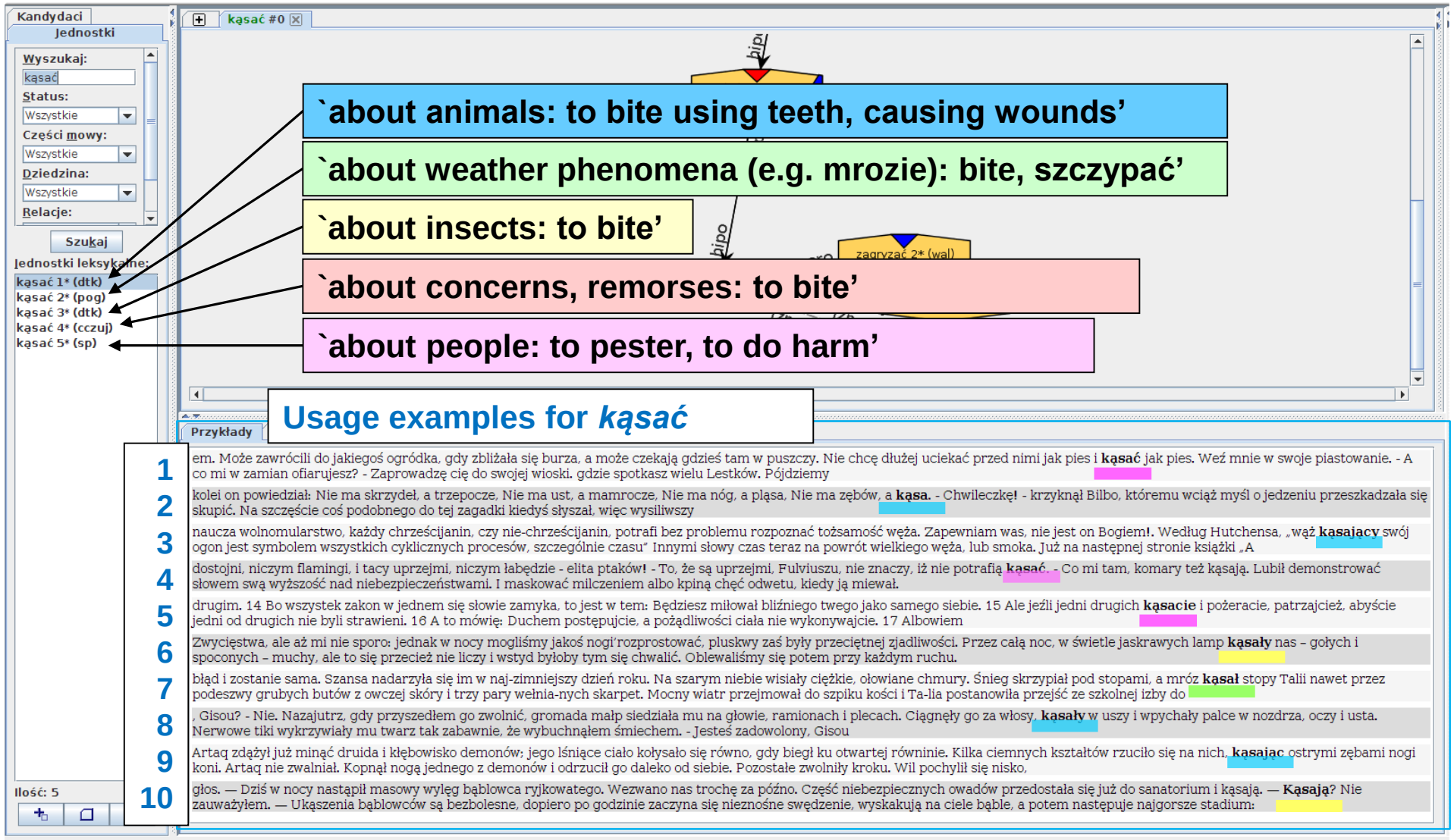
Suggestions generated by Paintball algorithm

WordnetLoom 2.0



Corpus-based Wordnet Development: Use Examples

eLex 2019
Sintra
2019-10-03
CLARIN-PL



Corpus-based Wordnet Development

6. Wordnet editing supported by tools
   - semantic exploration of corpora
     - Corpus concordancers
     - *LexCSD (character word embeddings) - usage examples – primary source for adjectives and verbs*
     - *Measure of Semantic Relatedness*
     - *WordnetWeaver*
   - consulting traditional linguistic resources
     - dictionaries, encyclopaedias (including Wikipedia), lexicons…
     - linguists' intuition, guidelines and consulting within plWordNet team
7. Linguistic work management
   - System for group work (*Redmine* system): tasks assignment, team communication etc.
   - plWordNet `Big Brother' – a web-based system for monitoring and verifying work
   - Verification and coordination: linguists plus coordinators

WordNet Tracker

eLex 2019
Sintra
2019-10-03

CLARIN-PL

WordNet Tracker

Admin (Administrator)

Search...

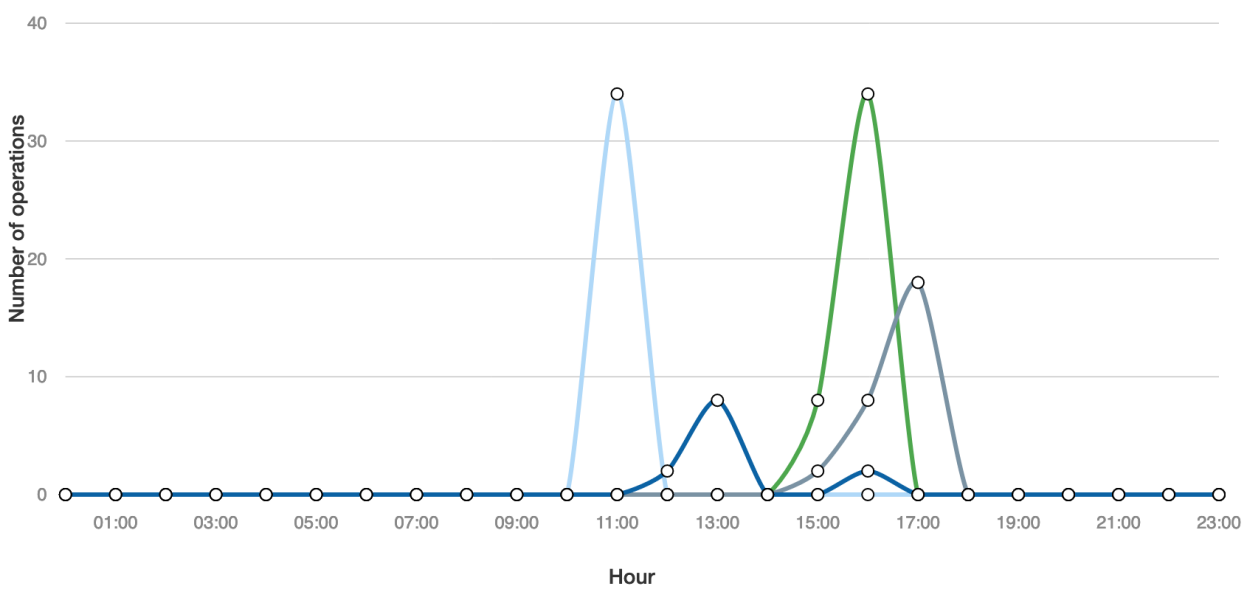
- Admin Panel
- Dashboard
- Synsets
- Senses
- Emotions
- Users
- Diagnostics
- Statistics

Dashboard

4	48	3	14
New Synsets	New Synset Relations	New Senses	New Sense Relations
View Details	View Details	View Details	View Details

Today's Users Activity

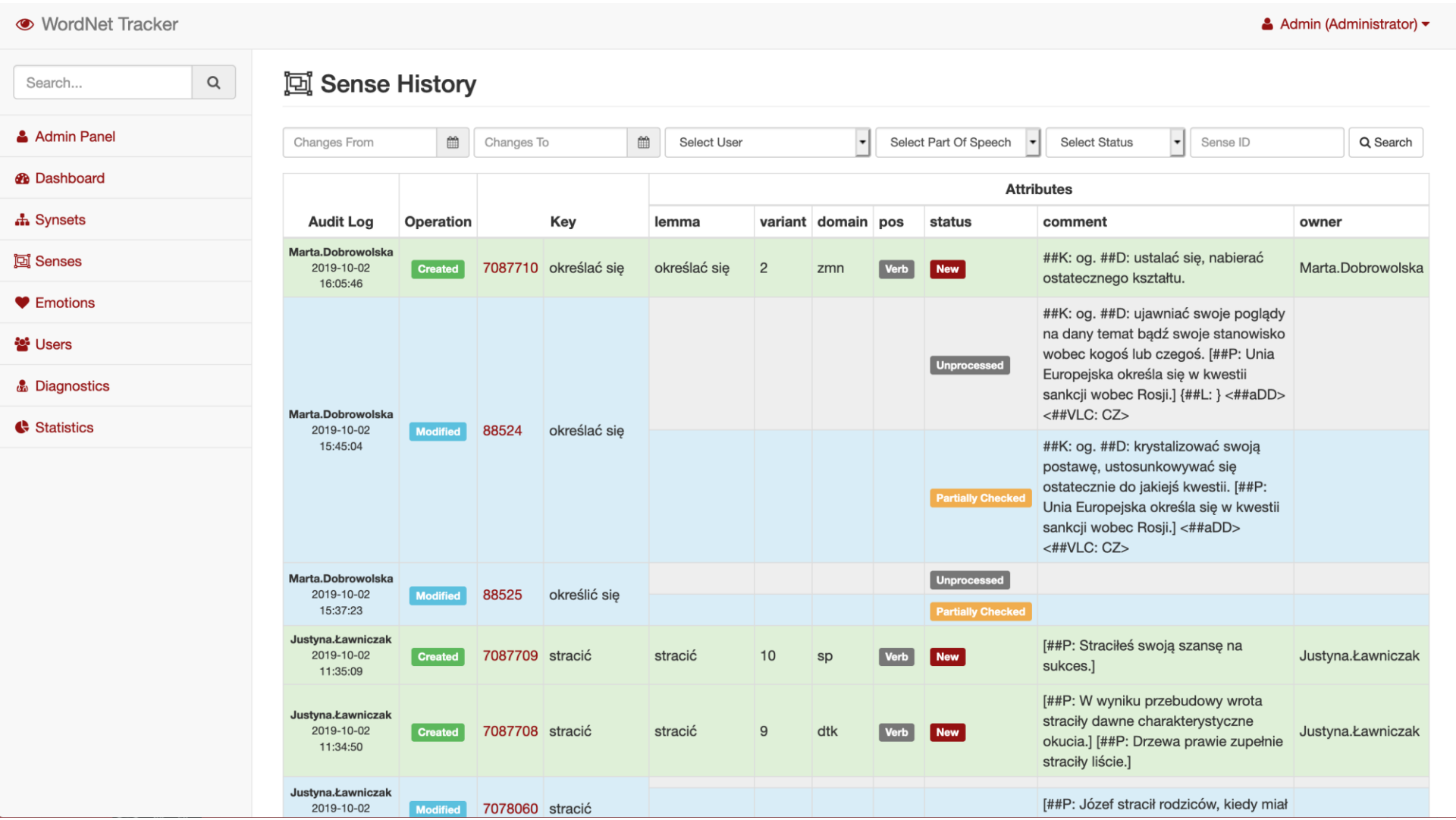
Actions



WordNet Tracker

eLex 2019
Sintra
2019-10-03

CLARIN-PL



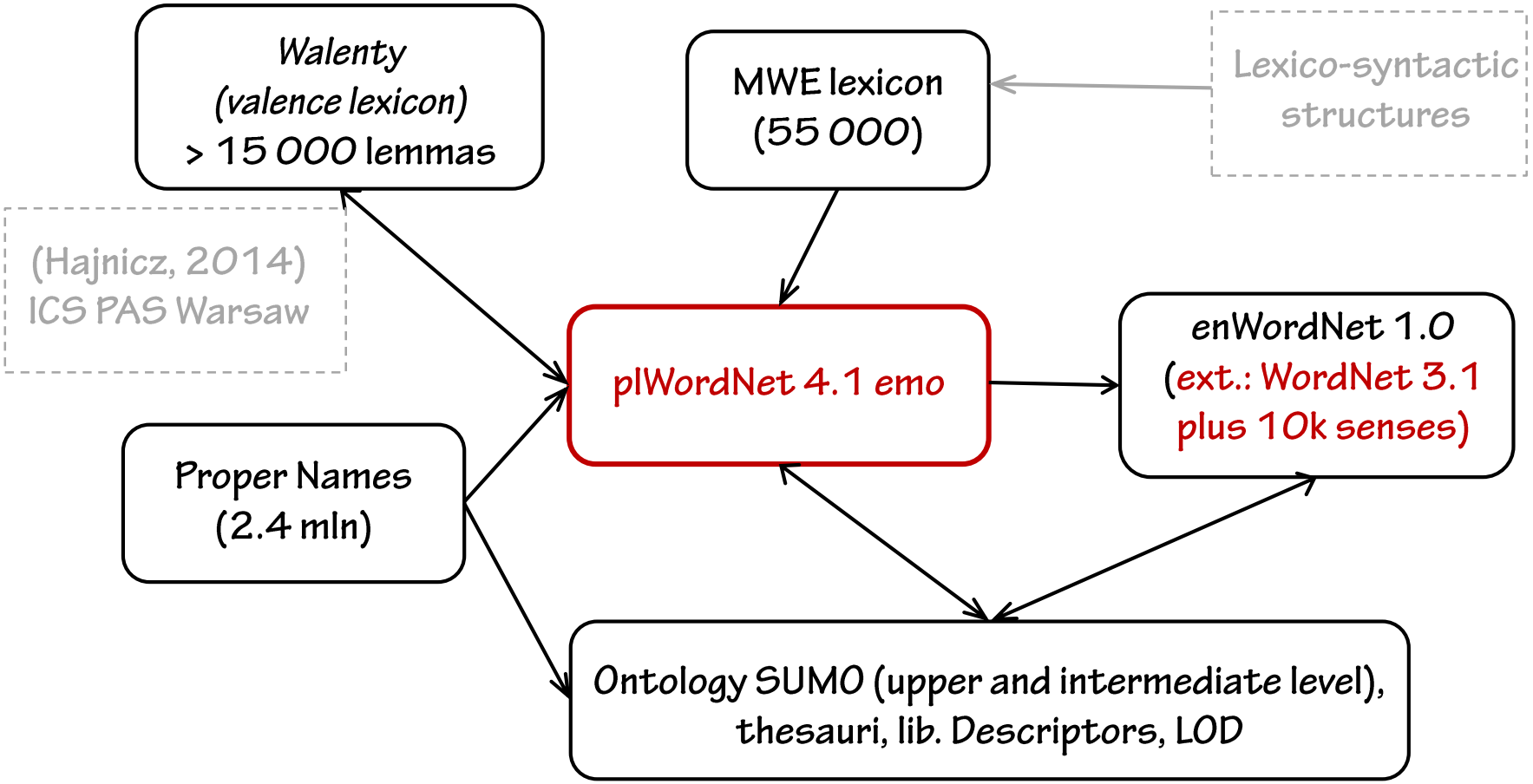
Audit Log	Operation	Key		Attributes						
				lemma	variant	domain	pos	status	comment	owner
Marta.Dobrowolska 2019-10-02 16:05:46	Created	7087710	określać się	określać się	2	zmn	Verb	New	##K: og. ##D: ustalać się, nabierać ostatecznego kształtu.	Marta.Dobrowolska
Marta.Dobrowolska 2019-10-02 15:45:04	Modified	88524	określać się					Unprocessed	##K: og. ##D: ujawniać swoje poglądy na dany temat bądź swoje stanowisko wobec kogoś lub czegoś. [##P: Unia Europejska określa się w kwestii sankcji wobec Rosji.] {##L: } <##aDD> <##VLC: CZ>	
								Partially Checked	##K: og. ##D: krystalizować swoją postawę, ustosunkowywać się ostatecznie do jakiejś kwestii. [##P: Unia Europejska określa się w kwestii sankcji wobec Rosji.] <##aDD> <##VLC: CZ>	
Marta.Dobrowolska 2019-10-02 15:37:23	Modified	88525	określić się					Unprocessed		
								Partially Checked		
Justyna.Ławniczak 2019-10-02 11:35:09	Created	7087709	stracić	stracić	10	sp	Verb	New	[##P: Straciłeś swoją szansę na sukces.]	Justyna.Ławniczak
Justyna.Ławniczak 2019-10-02 11:34:50	Created	7087708	stracić	stracić	9	dtk	Verb	New	[##P: W wyniku przebudowy wrota straciły dawne charakterystyczne okucia.] [##P: Drzewa prawie zupełnie straciły liście.]	Justyna.Ławniczak
Justyna.Ławniczak 2019-10-02	Modified	7078060	stracić						[##P: Józef stracił rodziców, kiedy miał	

A Wordnet is not Enough

- Morphological dictionary (Woliński, 2014)
- Lexicon of Multiword Expressions (structurally described) (Kurc et al., 2012)
- Lexico-semantic resources – lexical meanings
  - plWordNet 4.1 emo (Słowosieć)
  - Syntactic-semantic valency lexicon – Walenty (Przepiórkowski et al., 2014) IPI PAN
  - enWordNet 1.0 – a significant expansion of WordNet 3.1
  - Mapping of the whole plWordNet 4.1 onto WN+enWordNet
- Knowledge resources
  - NELexicon 2.0 – a large lexicon of Proper Names
  - Mapping of plWordNet onto SUMO Ontology
  - Mapping of plWordNet onto Wikipedia articles (partial), …

plWordNet as a pivot element

- A complex system of lexico-semantic resources (Maziarz et al. 2016)



More than 15 mln semi-automatically added links (Maziarz et al., GWC 2016)

plWordNet synset relations

- Hypernymy/hyponymy
  - defined for all parts of speech
  - also for verbs
  - adjectives and adverbs – limited but surprisingly numerous
- Meronymy / holonymy (with subtypes)
- Instance / type (for Proper Names and Nouns)
- Inter-register synonymy
  - nouns, verbs, adjectives and adverbs

  ≈ synonymy across different stylistic registers
  - links stylistically marked lexical units with their unmarked counterparts
  - e.g. *samochód 1* a car' – *fura 1* 'a car (slang)'

Substitution tests

- Definition of relation for wordnet editors
  - textual definition
  - substitution test
  - examples, further explanations, discussion

Condition:

Stylistic register of *Y* must be not lower in the register hierarchy than register of *X*.

Testing expressions:

If she/it is *X*, then she/it must be *Y*

If she/it is *Y*, then she/it need not be *X*

If she/it is not *Y*, then she/it cannot be *X*

Substitution tests

Condition:

Both: *ocean* ‘ocean’ and *zbiornik wodny* ‘water basin’ are of the general stylistic register.

Testing expressions:

If she/it is *oceanem* ‘ocean’, then she/it must be *zbiornikiem wodnym* ‘water basin’

If she/it is *zbiornikiem wodnym* ‘water basin’, then she/it need not be *oceanem* ‘ocean’

If she/it is not *zbiornikiem wodnym* ‘water basin’, then she/it cannot be *oceanem* ‘ocean’

Noun Lexical Relations

- Contrast
  - Complementary antonymy
  - Proper antonymy
  - Converseness
- Cross-PoS Synonymy (N-V, N-Adj)
- Feminity
- Markedness
  - Young being
  - Deminutive
  - Augmentativeness
- Feature bearer
- Role
  - Agens
  - Instrument
  - Result
  - Place
  - Patient
  - Time
  - Result with unexpressed predicate
  - Place with unexpressed predicate
- Derivation

Verb Synset Relations

- *Hypernymy – hyponymy*
- Backward relations
  - *presuposition* (V-V,N,A,Adv) - close to logical presupposition
    - *żywy* 8 ‘alive’ ←pres.- *umrzeć* 1 ‘to die’
  - *preceding* (V-V,N,A,Adv) - represents a possibility that one situation happens before the other one
    - *siedzieć* 1 ‘to sit’, *stać* 3 ‘to stand’ ←prec- *położyć się* 1 ‘to have laid down’
- Co-occurrence of two situations
  - *meronymy* (V-V_{imp}) and holonymy (V-V_{imp}) (not automatically reverse) - a situation is an element of a larger, more general situation, necessary simultaneous co-occurrence of two situations
    - meronymy: *przełykać* ‘to swallow’ is an integral part of situation *jeść* ‘to eat’
    - holonymy: *jeść* ‘to eat’ is a typical situation including *przełykać* ‘to swallow’

Verb Synset Relations

- Beginning of a situation
  - *inchoativity* (V-V_{imp},N), where the first verb represents an initial phase of a situation represented by the second element
    - *zakochać się* 1 ‘to fall in love’ → *kochać* 1 ‘to love’
- Resulting in a situation
  - *processuality* (V-N,A,Adv) – ‘to become or to be becoming
    - *zmieniać się* 1 `to be changing itself/yourself’ = to be becoming ‘*inny* 1 `different’
  - *causality* (V-V,N,A,Adv) – ‘to cause’
    - *zablokować 2* ‘to lock’ → *blokada* 4 ‘lock’
- State (V-N,Adj,Adv) – being in some state
  - *jaśnieć* 1 ‘to $shine_{imp}$’ means ‘to be bright [*jasny* 8]’ or ‘to be brightly [jasno 8]’

Synset Relations

- Multiplicativity
  - *Iterativity* (V_{imp}-V) – repetition of some state or activity
    - *grywać* '~to play a little from time to time' → *grać* 'to play'
  - *Distributivity* (V_{perf}-V_{perf}) – performing an activity by many subjects or on many objects
    - *nakupić* 'to buy many things' → *kupić* 'to buy_{perf}'
- Inter-register synonymy V-V
  - LUs are close in meaning but have incompatible stylistic registers
  - *pieprzyć* [vulgar] '~to speak (nonsense)' → *mówić* 'to speak'

plWordNet content

	synsets	lemmas	LUs	avs
GermaNet	101,371	119,231	131,814	–
Princeton WordNet 3.1	117,659	155,593	206,978	1.74
enWordNet 1.0	+7841 125,500	+10119 165,712	+11633 218,611	1.74
plWordNet 4.1 emo	**224,179**	**192,495**	**290,366**	**1.32**

- LUs – lexical units (= senses)
- avs – average synset size

PLWORDNET
SŁOWOSIEĆ

plWordNet content

- 53 different relation types (107 when counting subtypes)
  - including many relations linking lexical units of different PoS
- Semantic domains (*lexicographer files* from WordNet)
- Semantic verb classes – constitutive features, supporting defining the relation structure
- Stylistic labels (11 in total)

Description layer	Instances
lexico-semantic relations	~716K
glosses	>205K
usage examples	>78K (+ >70K emotive)
links to Wikipedia	~60K
LUs with sentiment annotation	>87K

plWordNet comparison

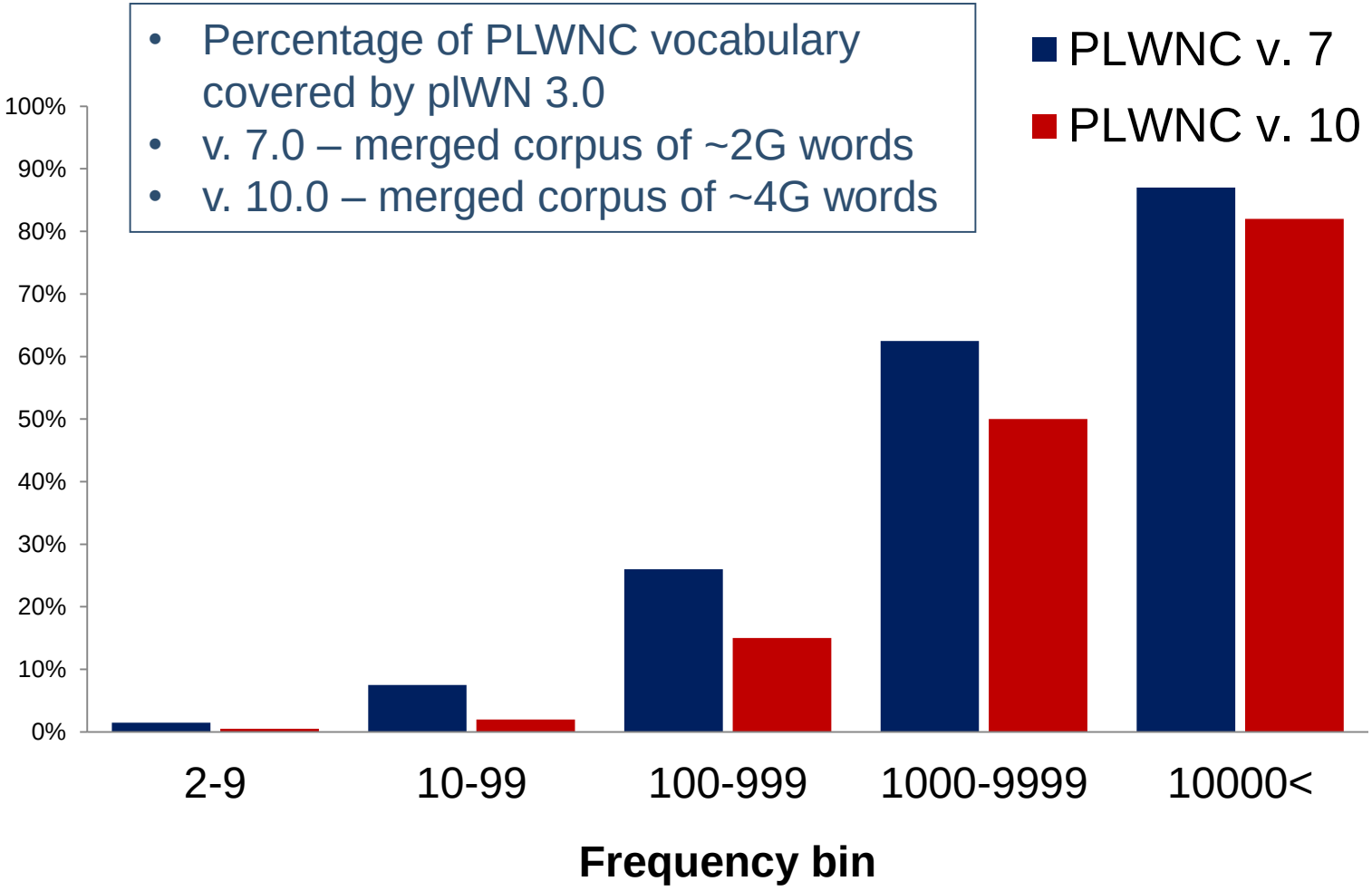
Network volume and density

WordNet 3.1	**verbs**		**nouns**		**adverbs**		**adjectives**		**all**	
	N	ρ	N	ρ	N	ρ	N	ρ	N	ρ
LU relations	24,840	0.99	44,185	0.28	720	0.13	21,636	0.72	91,381	0.42
synset relations	16,827	1.22	145,338	1.62	109	0.03	23,491	1.29	185,765	1.48
all relation types	80,280	3.20	492,457	3.12	1,015	0.18	86,221	2.87	659,973	3.02
plWordNet 3.0	**verbs**		**nouns**		**adverbs**		**adjectives**		**all**	
	N	ρ	N	ρ	N	ρ	N	ρ	N	ρ
LU relations	48,744	1.50	98,376	0.58	12,542	1.14	48,894	1.02	208,556	0.80
synset relations	36,616	1.66	219,266	1.75	19,716	2.18	48,258	1.17	323,856	1.64
all relation types	127,065	3.92	494,893	2.94	43,551	3.94	118,574	2.47	784,083	3.02

ρ is the relation density measured either for LUs, or synsets, or for all relation types

plWordNet coverage

- Percentage of PLWNC vocabulary covered by plWN 3.0
- v. 7.0 – merged corpus of ~2G words
- v. 10.0 – merged corpus of ~4G words



plWordNet 4.1 emo applications

- Wide coverage inspires a lot of applications
- plWordNet is a pivotal element of a system of language and knowledge resources
- An anchor to Linked Open Data via mapping to WordNet
- Monolingual and bilingual dictionary
  - Web-based: http://plwordnet.pwr.edu.pl
  - Android application
  - WordnetLoom for visual, graph-based browsing
  - included in a very large and popular Polish multilingual Web dictionary Ling
- WordTies (Pedersen et al., 2012), Open Multilingual WordNet (Bond and Foster, 2013)

plWordNet 4.1 emo applications

- Numerous research applications, for instance
  - Classification of gestures based on the verb categorisation in plWordNet (Lis and Navarretta, 2014)
  - Referred to in the resource for textual entailment (Przepiórkowski, 2015)
  - Language correction
  - Relation extraction
  - Text classification
  - Open Domain Question Answering
  - A quasi-ontology in document structure recognition
- An exceptional case is the practical use of plWordNet during the medical treatment of aphasia

plWordNet 4.1 emo applications

- A large number of declared applications:
- Education (at different levels) including Polish language teaching,
- Building dictionaries, extraction of synonyms and semantically related words, detection of loanwords,
- Cross-linguistic study on phonesthemes, classification of metaphorical expressions,
- Corpus studies, grammar development, comparative and contrastive studies,
- Language recognition, parsing disambiguation, semantic analysis of text, document similarity measures, semantic indexing of documents, semantic information retrieval,
- Recommendation systems, construction of chatbots and dialogue systems,
- Plagiarism detection,
- Translation evaluation, data visualisation, research on complex networks and ontologies, …

Conclusions

- Corpus-based wordnet development methods allows for good coverage of language data and close relation to the language use
- Minimal Commitment Principle wordnet models results in a relational semantic dictionary
- plWordNet is an example of a wordnet which is a relational semantic dictionary and *vice versa*

Thank you very much for your attention!

www.clarin-pl.eu

CLARIN-PL
Common Language Resources and Technology Infrastructure

Supported by the Polish Ministry of Science and Higher Education [CLARIN-PL]

Bibliography

- B. Broda M. Marcińczuk, Maziarz Radziszewski Wardyński, Adam (2012). KPWr: Towards a Free Corpus of Polish. In Khalid Choukri et al. Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), European Language Resources Association (ELRA), Istanbul, Turkey, 2012, ISBN: 978-2-9517408-7-7.
- Rudnicka Ewa, Maziarz Marek M, Piasecki Maciej, Szpakowicz Stanisław (2012) A stategy of mapping Polish WordNet onto Pinceton WordNet. In 24th International Conference on Computational Linguistics : proceedings of COLING 2012, 8-15 December 2012, Mumbai, India : posters. Vol. 3. Mumbai : The COLING 2012 Organizing Committee, 2012. s. 1039-1048.
- Kurc, R.; Piasecki, M. & Broda, B. (2012) Constraint Based Description of Polish Multiword Expressions. In Calzolari, N.; Choukri, K.; Declerck, T.; Dogan, M. U.; Maegaard, B.; Mariani, J.; Odijk, J. & Piperidis, S. (Eds.) Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), European Language Resources Association (ELRA), 2408-2413.
- Woliński, Marcin. (2014) Morfeusz Reloaded. In Nicoletta Calzolari (Conference Chair) Khalid Choukri, Thierry Declerck Hrafn Loftsson Bente Maegaard Joseph Mariani Asuncion Moreno Jan Odijk Stelios Piperidis (Ed.): Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pp. 1106–1111, European Language Resources Association (ELRA), Reykjavik, Iceland, 2014, ISBN: 978-2-9517408-8-4.
  http://nlp.ipipan.waw.pl/Bib/wol:14.pdf

Bibliography

- Przepiórkowski, Adam; Hajnicz, Elżbieta; Patejuk, Agnieszka; Woliński, Marcin; Skwarski, Filip; Świdziński, Marek Walenty (2014) Towards a comprehensive valence dictionary of Polish. In Choukri, Khalid et al. (Ed.): Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pp. 2785–2792, European Language Resources Association (ELRA), Reykjavik, Iceland, 2014, ISBN: 978-2-9517408-8-4.
- Maziarz, M.; Piasecki, M. & Szpakowicz, S. (2013)The chicken-and-egg problem in wordnet design: synonymy, synsets and constitutive relations Langauge Resources and Evaluation, 2013, 47, 769-796, http://link.springer.com/article/10.1007/s10579-012-9209-9 *(open access)*
- Maziarz, Marek and Piasecki, Maciej and Rudnicka, Ewa and Szpakowicz, Stan and Kędzia, Paweł (2016) plWordNet 3.0 -- a Comprehensive Lexical-Semantic Resource. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. The COLING 2016 Organizing Committee, pp. 2259—2268, Osaka, Japan http://aclweb.org/anthology/C16-1213