



ADAM MICKIEWICZ UNIVERSITY IN POZNAŃ

Zygmunt Vetulani
Department of Computer
Linguistics and Artificial
Intelligence
www.amu.edu.pl/~vetulani
vetulani@amu.edu.pl

Grażyna Vetulani
Faculty of Modern
Languages and
Litterature
Institute of Roman
Philology
gravet@amu.edu.pl

Verb-Noun Collocations in PoINet 2.0

by

Zygmunt Vetulani
Grażyna Vetulani

CCLCC 2014, Tübingen, Germany, 14.08.2014

www.amu.edu.pl



Our intention is to present the context of our work on Polish verb-noun collocations which is **PolNet**

The project (of extending PolNet with predicative synsets) is a part of a **long term program** of designing and implementing a Lexicon Grammar for Polish.



Lexicon Grammar

Lexcon Grammar is an idea of semantic grammar introduced by Maurice Gross under the influence of Zellig Harris

Maurice Gross started working on LGs in late 1960s

„...the basic unit of meaning is the elementary sentence rather than the word”
therefore

in order to describe a predicative word it is necessary to take into account
properties of the class of elementary sentences that one can built using this
word as predicate.

Thus:

- the predicative words should be described in terms of syntactic and semantic properties of elementary sentences (valency structure)

Gross implemented the LG for French in form of „syntactic tables”.



Lexicon Grammar – similar approaches

In 1970s a Polish slavist, Kazimierz Polański, started (independently of Gross) his monumental description of Polish verbs (7,000) according to syntactic-semantic criteria

Kazimierz Polański (ed.). 1992. *Słownik syntaktyczno-generatywny czasowników polskich* vol. I-IV, Ossolineum, Wrocław, 1980-1990, vol. V, Kraków./* Syntactic-generative dictionary of Polish verbs */

MAMIĆ /* to lead on*/

I. 'działać łudzaco, bałamucić, zawodzić, tumanić'

1. NP1(N) __ NP1(ACC) + (NP(I))

2. NP2(N) __ NP2(ACC)

NP1(N) → [+Hum]

NP1(ACC) → [+Hum]

NP(I) → [-Abstr, -Anim][+Abstr]

NP2(N) → [-Abstr, -Anim][+Abstr]

NP2(ACC) → [oczy][wzrok][+Hum]

/* notation simplified, examples of use omitted*/

Semantic features used by Polański

[+Abstr] – abstract	[Fl] – plant
[-Abstr] – concrete	[Inf] – information
[+Anim] – animate	[Instit] – institution
[-Anim] – non-animate	[Instr] – instrument
[+Hum] – human	[Liqu] – liquid
[-Hum] – non-human	[Mach] – machine
[Coll] – collective	[Mat] – material
[Elm] – element	[Pars] – part

Recently: FrameNet (Fillmore, ...) and VerbNet (Palmer, ...)



Lexicon Grammar – our approach

**NOW, in my lab,
we implement the Polish lexicon-grammar
as a wordnet (PoINet)**



Princeton WordNet (PWN)

Wordnet technology started in 1985 with (Princeton) WordNet. It is a lexical database for the **English** language designed in the Cognitive Science Laboratory (Princeton University, George A. Miller).

Organising concepts:

- synsets (classes of synonyms)
- hierarchical relations (hyponymy/hyperonymy)

Nowadays: over 200,000 synsets (word-senses)

In computer science WordNet (PWN) and similar resources are often used as ontologies



Princeton WordNet followers

Euro WordNet

EuroWordNet is a system of PWN-inspired wordnets for several European languages, interconnected with interlingual links stored in the ILI (Interlingual Index). (www.illc.uva.nl/EuroWordNet/) (P. Vossen)

Balknet

Has produced WordNets for six European languages (Bg, Cz, Gr, R, T and Srb). DEBVisDic client-server wordnet development tool. (D. Christodoulakis, K. Pala, D. Tufis) (www.dblab.upatras.gr/balkanet/)

.....
many other (GermaNet...)

PolNet (UAM, Poznań) and plWordnet (TU Wrocław)

PolNet - Polish wordnet developed at the UAM since 2006 (www.amu.edu.pl/~vetulani)



Inspirations for PoINet 1.0 and 2.0

Projects:

PWN (Miller, Fellbaum),
Euro WordNet (Vossen),
Balkanet EU (Pala, Vitas)
FrameNet (Fillmore),
VerbNet (Palmer),
Lexicon-Grammar (Gross).

Resources and tools:

Traditional dictionaries of Polish,
"Syntactic-Generative dictionary of Polish Verbs" (Polański),
Dictionaries of predicative nouns and collocations (Grażyna Vetulani)
DEBVisDic (Pala, Czech Republic)
IPI PAN Corpus for Polish (Przepiórkowski et al.)



„PolNet - Polish WordNet” project - PolNet 1.0

PolNet is being developed at the UAM since 2006

- merge model of development (respects original conceptualisation of Polish speakers) (from scratch, **no translation from PWN, no automatic generation of synsets**)
- first public release (PolNet 1.0) in 2011/2012 (at LTC and GWC), free download for research e.g. META SHARE (<http://metashare.elda.org/repository/search/?q=PolNet>)
- **development algorithm** was first published in 2007 at LTC (and the revised version in LNAI 5603 of 2009)
- PolNet 1.0.:
 - nouns – 11,700 synsets for 20,300 word-meaning pairs and 12,000 simple nouns,
 - verbs – 1,500 synsets for 2,900 word-meaning pairs and 900 simple verbs

We applied as development tools VisDic and DEBVisDic (Pala), well suited to support verb **valency structure** and **semantic roles**.



Verb Valency Structures and Semantic Roles in PolNet

In PolNet the **valency structures** are determined by the values of **semantic roles** associated to the argument positions and by **morphosyntactic properties** of the arguments.

Semantic roles:

- the semantic roles are functions (in mathematical sense) associated to the argument positions in the syntactic patterns.
- values of these functions are ontology concepts in form of noun synsets or top-level ontology concepts (SUMO).

See the example below.

Action, Agent, Asset,
Benef, Cause, Dest,
Exp, Giver, Instr, Loc,
Object, Partner, Patient,
Product, Proposition,
Purpose, Recipient,
Source, State, Time,
Value

(after M. Palmer, 2008)



Description of the verbal synset {*pomóc:1, pomagać:1*} /to help/ in PolNet 1.0

```
<SYNSET>
<VALENCY>
<DEF>"to participate in somebody's work in order to make it easier "</DEF>
<SYNONYM>
<FRAME>Agent(N)_Benef(D)</FRAME>
<FRAME>Agent(N)_Benef(D) Action('w'+L)</FRAME>
<FRAME>Agent(N)_Benef(D) Manner</FRAME>
<FRAME>Agent(N)_Benef(D) Action('w'+L) Manner</FRAME>
</VALENCY>
<ILR type="category_domain" link="1356">CITTA:1</ILR>
<ILR type="Agent" link="ENG20-02383992-n">człek:1, człowiek:1, istota ludzka:1, zwierzę:2, ....</ILR>
<ILR type="Benef" link="ENG20-02383992-n">człek:1, człowiek:1, istota ludzka:1, zwierzę:2, ....</ILR>
<ILR type="Action" link="PL_PK-2035015933">czynność:1</ILR>
<ILR type="Manner" link="2214">CECHA_ADVERB_JAKOŚĆ:1</ILR>
<WORD>pomóc</WORD> <WORD>pomagać</WORD>
<LITERAL Inote="U1" sense="1">pomóc</LITERAL>
<LITERAL Inote="U1" sense="1">pomagać</LITERAL>
</SYNONYM>
<ID>3441</ID>
<USAGE>Agent(N)_Benef(D); "Pomogłam jej."</USAGE>
<USAGE>Agent(N)_Benef(D) Action('w'+L); "Pomogłam jej w robieniu lekcji."</USAGE>
<USAGE>Agent(N)_Benef(D) Manner Action('w'+L); "Chętnie pomogłam jej w lekcjach."</USAGE>
<USAGE>Agent(N)_Benef(D) Manner;"Chętnie jej pomagałam."</USAGE>
<CREATED>agav 2010-11-27 18:49:47</CREATED>
<POS>v</POS>
</SYNSET>
```

Gloss

Both perfective and imperfective forms

Syntactic patterns

Sem. roles values

Use examples



Present PoINet Development (from PoINet 1.0 to PoINet 2.0) Importance of collocations in the PoINet

This development phase consists in the inclusion of **verb-noun collocations** which are very frequent (and specific) in Polish and many other languages.

They play the role of compound verbs:

- support verb (light verb) + predicative noun
- the support verb supplies formal properties while the predicative noun opens argument positions

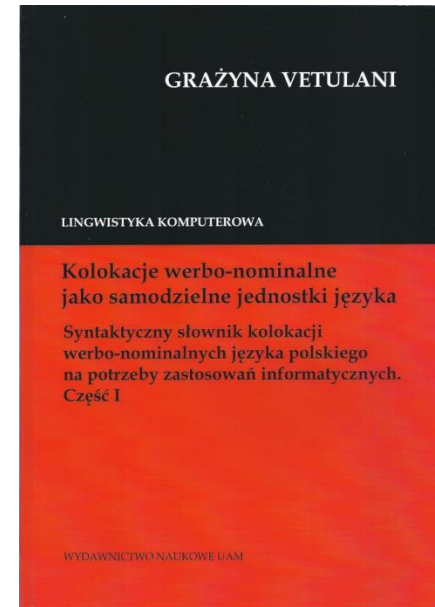
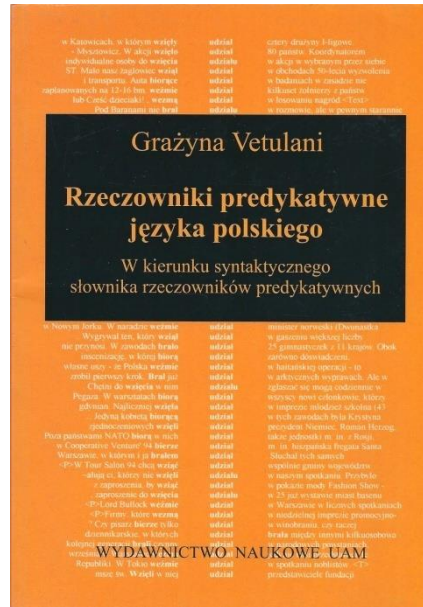
Notice that

- the support verbs in the VN collocations are hardly predictable.
- in Polish some VN collocations do not have simple verbs as equivalents (e.g. *ponieść klęskę* /≈to suffer a defeat/, *mieć nadzieję* /≈to hope/)

Example

ambicja, f/
/* ambition */

- mieć(Acc)/N1(D),
/* to have an ~ of sth1 */
- mieć(Acc,pl)/MOD,
/* to have MOD ~s */
- posiadać(Acc,pl)/MOD,
/* to own MOD ~s */
- ujawniać(Acc,pl)/MOD,
/* to show MOD ~s */
- zaspokoić(Acc)/N1(Gen),
/* to fulfill one's ~ of sth */
- zaspokoić(Acc,pl)/MOD,
/* to fulfill MOD ~s ~*/
- zaspakajać(Acc)/N1(Gen)
/* to fulfill one's ~ of sth */
- zaspakajać (Acc,pl)/MOD
/* to fulfill MOD ~s */



5 classes of predicative nouns:

- Class I (typical) - env. 2862 (env. collocations 16,000 identified in corpora)
- Class II (features) – env. 2826
- Class III (diseases) – env. 275
- Class IV (profession) – env. 1478
- Class V (circumstantials) – env. 213

To be included in PoNet (until end 2014) env. 3600 collocations of classes I and II for env 1200 predicative nouns



Example (G. Vetulani 2012) *kierunek* (direction)

dryfować w **kierunku**/ dryfować w(Ms)/N1(D);MOD,
iść w **kierunku**/ iść w(Ms)/N1(D);MOD,
kontynuować **kierunek**/ kontynuować(B)/MOD/N1(D),
mieć **kierunek**/ mieć(B)/MOD,
nadać **kierunek**/ nadać(B)/MOD/N1(C),
nadawać **kierunek**/ nadawać(B)/MOD/N1(C),
nakreślić **kierunek**/ nakreślić(B)/MOD/N1(D),
obrać **kierunek**/ obrać(B)/MOD/N1(D),
podążać w **kierunku**/ podążać w(Ms)/MOD,
posuwać się w **kierunku**/ posuwać się w(Ms)/MOD,
podtrzymać **kierunek**/ podtrzymać(B)/MOD/N1(D),
pójść w **kierunku**/ pójść w(Ms)/MOD,
przybrać **kierunek**/ przybrać(B)/MOD,
przyjąć **kierunek**/ przyjąć(B)/MOD, przyjmować
kierunek/ przyjmować(B)/MOD

realizować **kierunek**/ realizować(B)/MOD/N1(D),
trzymać się **kierunku**/ trzymać
się(D)/MOD/N1(D),
utrzymać **kierunek**/ utrzymać(B)/MOD/N1(D),
wskazać **kierunek**/ wskazać(B)/N1(C),
wskazywać **kierunek**/ wskazywać(B)/N1(C),
wytyczać **kierunek**/ wytyczać(B)/MOD/N1(D),
wytyczyć **kierunek**/ wytyczyć(B)/MOD/N1(D),
wyznaczać **kierunek**/ wyznaczać(B)/MOD/N1(D)
wyznaczyć **kierunek**/ wyznaczyć(B)/MOD/N1(D)
zachować **kierunek**/ zachować(B)/MOD/N1(D),
zdążać w **kierunku**/ zdążać w(Ms)/MOD,
zmierzać w **kierunku**/ zmierzać w(Ms)/MOD,

Syntactic patterns



Two main challenges for lexicographers (among many): granularity, attribution of semantic roles

Problem: No consensus about the concept of synonymy

Leibnitz approach: synonyms are to be substituable in all contexts

small synsets=fine granulation=huge number of synsets

In opposition to:

Miller and Fellbaum: synonyms must be substituable in at least one context

large synsets=coars granulation=low number of synsets

(Many intermediate solutions)

Why granularity problem is important?

Using a wordnet as ontology requires that **all elements of a given synset** must represent **the same concept**.

In PolNet: synonymy → common valency structure

Synonymous are solely such word/meanings for which

- the given argument position is associated with just one semantic role
- the corresponding semantic roles take the same values
- corresponding arguments have the same morphosyntactic properties

(this condition is necessary but not sufficient for verb synonymy).



Example of a problem (typical mistakes)

The valency structures for all verbs in the synset must be pairwise compatible, i.e. the same semantic role must be associated with the corresponding syntactic positions

Strict application of the definition is difficult – lexicographers make mistakes

```
<VALENCY> /*fragment*/ (szanować (kogoś)=darzyć szacunkiem (kogo))
<DEF>To respect sb or sth </DEF>.
<FRAME>Agent(N) _ Benef(Acc)</FRAME> /* Andrzej szanuje Sonię*/ /* Andrew respects Sonia */
<FRAME>Agent(N) _ Benef(Acc) Cause('za'+Acc)</FRAME> /* Andrzej szanuje Sonię za mądrość*/
-----
<FRAME>Agent(N) _ Object(Acc) Cause('za'+Acc)</FRAME> /* Andrzej szanuje prawo za ład i
porządek*/ /* Andrew respects public order */
</VALENCY> (szanować (coś)=mieć w poszanowaniu (coś)
=have a good opinion about (sth))
<STAMP>artusia 2014-03-05 23:04:35</STAMP>
```

Fig.1. Fragment of the PoInet 1.0 code for the synset {szanować:1} with the gloss <DEF>To respect sb or sth </DEF>.

This simple example illustrates at least two dilemmas:

- 1) about granularity (why not two different synsets: for „to respect somebody” or „to respect something”?)
- 2) about the choice of the semantic roles in the first two frames („Benef” or (more neutral) „Object” ?)



Example of a problem (typical mistakes)

Another meaning {szanować:3} is close to the English *to protect , preserve* (e.g. nature). It corresponds to another synset defined as follows (fragment) («chronić coś (przed zniszczeniem lub nadmiernym zużyciem)/to protect sth (against sth)»)

```
<VALENCY> /*fragment*/  
<DEF> To protect sth</DEF>  
<FRAME>Agent(N) _ Object(Acc)</FRAME> /* Wise people protect nature*/  
</VALENCY>  
<STAMP>artusia 2014-03-05 22:49:26</STAMP>
```



Transformations. A collocations' proper granularity dilemma

Piotr upoważnił **adwokata** do zakupu = Peter authorized the advocate to buy

Piotr dał pełnomocnictwo do zakupu **adwokatowi** = Peter gave advocate enablement to buy

<DEF>Nadać/nadawać komuś prawo do reprezentowania siebie celem wykonywania pewnych czynności</DEF>

The above observed granularity problem for *szanować* is formally solvable **on the basis of glosses to capture the meaning variation**. In case of collocations this may be impossible.

“Piotr upoważnił **adwokata**(Acc) do zakupu” /*upoważnić=authorize*/
<VALENCY>
<FRAME>Agent(N) _ **Patient**(Acc) Purpose('do'+G)
</FRAME>
</VALENCY>

“Piotr dał pełnomocnictwo do zakupu **adwokatowi**(D)”
<VALENCY>
<FRAME>Agent(N) _ **Patient**(D) Purpose('do'+G)
</FRAME>
</VALENCY>

SOLUTION: to consider two different synsets with the same gloss: <DEF>to entitle sb else to represent the concerned person to do sth </DEF>.

and to relate the two synsets by a relation describing the case transformation e.g. TRANS_CASE_PATIENT(Acc,D)



To be continued...



Verb-Noun Collocations in PoINet 2.0
Zygmunt Vetulani, Grażyna Vetulani
Polish National Program for Humanities
0022/FNiTP/H11/80/2011

THANK YOU



REFERENCES

- Charles J. Fillmore, Collin F. Baker, and Hiroaki Sato. 2002. The FrameNet Database and Software Tools. In: *Proceedings of the Third International Conference on Language Resources and Evaluation. Vol. IV*. LREC: Las Palmas.
- Aldo Gangemi, Roberto Navigli, and Paola Velardi. 2003. The OntoWordNet Project: Extension and Axiomatization of conceptual Relations in WordNet (PDF). *Proc. of International Conference on Ontologies, Databases and Applications of SEMantics (ODBASE 2003)* (Catania, Sicily (Italy)). pp. 820–838.
- Maurice Gross. 1994. Constructing Lexicon-Grammars. In: Beryl T. Sue Atkins, Antonio Zampolli (eds.) *Computational Approaches to the Lexicon*, Oxford University Press. Oxford, UK, pp. 213–263.
- George A. Miller, Richard Beckwith, Christiane D. Fellbaum, Derek Gross, and Katherine J. Miller. 1990. WordNet: An online lexical database. *Int. J. Lexicograph.* 3, 4, pp. 235–244.
- Karel Pala, Aleš Horák, Adam Rambousek, Zygmunt Vetulani, Paweł Konieczka, Jacek Marciniak, Tomasz Obrębski, Paweł Rzepecki, and Justyna Walkowska. 2007. DEB Platform tools for effective development of WordNets in application to PolNet, in: Zygmunt Vetulani (ed.) *Proceedings of the 3rd Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, October 5-7, 2005, Poznań, Poland*, Wyd. Pozn., Poznań, pp. 514-518.
- Martha Palmer. 2009. Semlink: Linking PropBank, VerbNet and FrameNet. In *Proceedings of the Generative Lexicon Conference*. Sept. 2009, Pisa, Italy: GenLex.
- Adam Pease. 2011. *Ontology: A Practical Guide*. Articulate Software Press, Angwin, CA. ISBN 978-1-889455-10-5.11
- Adam Przepiórkowski. 2004. *Korpus IPI PAN. Wersja wstępna / The IPI PAN CORPUS: Preliminary version*. IPI PAN, Warszawa.
- Grażyna Vetulani. 2000. *Rzeczowniki predykatywne języka polskiego. W kierunku syntaktycznego słownika rzeczowników predykatywnych*. (In Polish). Poznań: Wyd. Nauk. UAM.

Verb-Noun Collocations in PolNet 2.0

Zygmunt Vetulani, Grażyna Vetulani
Polish National Program for Humanities
0022/FNiTP/H11/80/2011

CCLCC 2014 Workshop, Tübingen, Germany, 14.08.2014



Verb-Noun Collocations in PolNet 2.0

Zygmunt Vetulani, Grażyna Vetulani
Polish National Program for Humanities
0022/FNiTP/H11/80/2011

Grażyna Vetulani. 2012. *Kolokacje werbo-nominalne jako samodzielne jednostki języka. Syntaktyczny słownik kolokacji werbo-nominalnych języka polskiego na potrzeby zastosowań informatycznych. Część I.* (In Polish). Poznań: Wyd. Nauk. UAM.

Zygmunt Vetulani, Justyna Walkowska, Tomasz Obrębski, Jacek Marciniak, Paweł Konieczka, Przemysław Rzepecki (2009): An Algorithm for Building Lexical Semantic Network and Its Application to PolNet - Polish Wordnet Project, in: Z. Vetulani and H. Uszkoreit, *Human Language Technology. Challenges of the Information Society. LTC 2007. Revised selected papers. LNAI 5603*, Springer, 369-381.

Zygmunt Vetulani. 2012. Wordnet Based Lexicon Grammar for Polish. *Proceedings of the Eith International Conference on Language Resources and Evaluation (LREC 2012)*, May 23-25, 2012. Istanbul, Turkey, (Proceedings), ELRA: Paris. ISBN 978-2-9517408-7-7, pp. 1645-1649.

Zygmunt Vetulani and Jacek Marciniak. 2011. Natural Language Based Communication between Human Users and the Emergency Center: POLINT-112-SMS. In: Zygmunt Vetulani (ed.): *Human Language Technology. Challenges for Computer Science and Linguistics. LTC 2009. Revised Selected Papers*. LNAI 6562, Springer-Verlag Berlin Heidelberg, str. 303-314.

Zygmunt Vetulani and Grażyna Vetulani. (2013): Through Wordnet to Lexicon Grammar, in: Fryni Kakoyianni Doa (Ed.). *Penser le lexique-grammaire : perspectives actuelles*, Editions Honoré Champion, Paris, France, 531-545.

Zygmunt Vetulani, Justyna Walkowska, Tomasz Obrębski, Paweł Konieczka, Paweł Rzepecki, and Jacek Marciniak. 2007. PolNet - Polish WordNet project algorithm, in: Zygmunt Vetulani (ed.) *Proceedings of the 3rd Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, October 5-7, 2005, Poznań, Poland*, Wyd.Poznańskie, Poznań, pp. 172-176.

Piek Vossen, Laura Bloksma, Horacio Rodríguez, Salvador Climent, Nicoletta Calzolari, and Wim Peters. 1998. The EuroWordNet Base Concepts and Top Ontology, Version 2, Final, January 22, 1998 (Euro WordNet project report) (<http://www.vossen.info/docs/1998/D017.pdf>; access 22/04/2011). (also: Piek Vossen (et al.)



Concordances for Vpred+*udział*

Weźmie w nim	udział	6 wozów ciężarowych, które
. W spotkaniu tym wzięli	udział	przedstawiciele: władz Tczewa,
również może wziąć w nim	udział	...<P>Prezydent Gintowt-
<P>Niemiecki gość weźmie	udział	w polsko-niemiecko-rosyjskim
- Nie zostawimy. Nie bierzemy	udziału	!<P>- O, pan kurator! Prosimy do
Łużkow.<P> Rosja nie weźmie	udziału	w III wojnie światowej tylko
startem ligi weźmiemy jeszcze	udział	w turnieju, organizowanym przez
Oprócz tego wzięli oczywiście	udział	w oficjalnym treningu, na który
w Katowicach, w którym wzięły	udział	cztery drużyny I-ligowe.
- Mysztowicz. W akcji wzięło	udział	80 państw. Koordynatorem
indywidualne osoby do wzięcia	udziału	w akcji w wybranym przez siebie
ST. Malo nasz żaglowiec wziął	udział	w obchodach 50-lecia wyzwolenia
i transportu. Auta biorące	udział	w badaniach w zasadzie nie
planowanych na 12-16 bm. weźmie	udział	kilkuset żołnierzy z państw
Pod Baranami nie brał	udziału	w rozmowie, ale w pewnym starannie