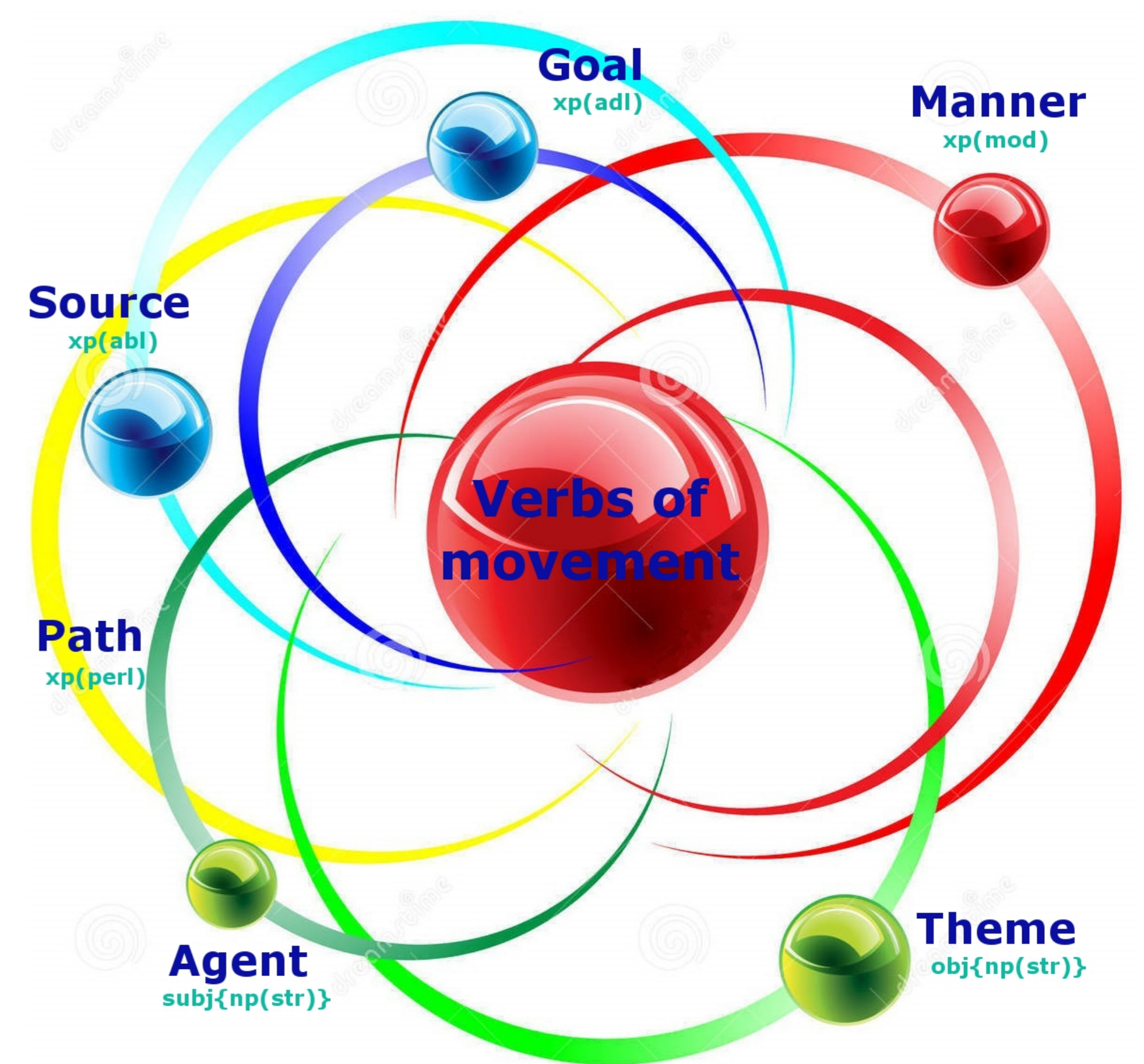


Using Polish Wordnet for Predicting Semantic Roles for the Valency Dictionary of Polish

Natalia Kotsyba
Institute of Computer Science PAS
Jana Kazimierza 5, Warsaw



Resources

Polish Valency Dictionary (*Walenty*)

Currently contains information about: grammatical aspect of verbs, arguments and their syntactic realizations, prevailing in the form of nominal or prepositional phrases; currently ca. 12K entries.

A sample entry for the verb *objechać* 'bypass, etc.':

_:perf:subj{np(str)}+obj{np(str)}+{np(inst)}

Polish Wordnet (*Słowniec*)

A network of lexical-semantic relations with two major dimensions for classifying verbs, explicated as 'artificial synsets'; version 2.1 contains over 19K verb synsets:

- lexical semantic groups (LS)
 - abstract: abstract relations, their change, or causing their change;
 - body: reaction of organism or physiological activity;
 - contact: physical contact or physical contact that influences an object;
 - loc: location or locative relations, their change, causing their change;
 - logic: situations connected with cause-effect relations or causing their change;
 - phys: physical phenomena, their change, or causing their change;
 - poss: change of possession or causing that change;
 - prod: producing something;
 - psych: emotional, mental, intellectual states, their change, or causing their change;
 - soc: situation connected with social interaction, their changes, and causing their changes;
 - temp: temporal sequence of events;
- lexical aspects (LA) and their features:

LA	dynamic	change	telic	intentional	perfective
act	+	+	-	+	+
activity	+/-	-	-	+	-
action	+	+	+	+	+/-
process	+	+	+	-	+/-
stative	-	-	-	-	-
accident	+	+	-	-	+
event	+	-	-	-	-
non-punctuals	+	+	-	+/-	+

The goal is to blend the two resources and assign semantic roles to each argument of each verb sense.

Lexical aspects are dependent on the lexical-semantic features. The table below shows the numbers of verbs belonging to a certain LA within the same LS group. The prevailing LAs are highlighted.

LA	abstract	body	contact	loc	logic	phys	poss	prod	psych	soc	temp
act	85	127	263	665	616	147	392	-	642	839	15
activity	105	159	273	734	667	3	423	-	884	991	15
action	253	142	1305	1552	1125	202	126	318	208	519	-
process	244	19	-	95	82	1075	6	110	178	146	94
stative	109	-	-	63	-	62	-	-	313	-	-
accident	156	32	118	238	371	218	32	-	218	21	213
event	134	53	81	243	304	295	27	-	64	52	89

Presently used semantic roles

	COLLECTION SCENARIO TYPE		ENTITY EVENT, STATE-OF-AFFAIRS ATTRIBUTE, LOCATION, TIME BACKGROUND, FOCUS				PART PHASE INSTANCE		
	INITIATOR		PATH		UNDERGOER		RESULT		
	AGENT	CAUSE	MEDIATOR	INSTRUMENT	PATIENT	OBJECT	RESULTEE	PRODUCT	GOAL
SOURCE *REASON (logic)	MANIPULATOR	*STIMULUS	BENEFACTOR	*ASSET(poss)	*EXPERIENCER	*THEME	*RECIPIENT	*CONTENT	
	*PROTAGONIST	EFFECTOR			*DURATION	(psych)	*MATERIAL	(pos)	
	*COMMUNICATOR				*TRACE (loc)	(soc, psych)		*BENEFICIARY	*DESTINATION
	COGNIZER (psych)				*PERCEIVER		(loc)	*PURPOSE	
					(psych)		(logic)		

Examples: ↗

Legend:

- METARoles
- ABSTRACT VERBS' ROLES
- CONCRETE ROLE GROUPS
- CONCRETE VERBS' ROLES
- LS-SPECIFIC ROLES

The proposed inventory uses several levels of abstraction for the roles. The most general ones are: (INITIATOR, UNDERGOER, PATH, RESULT(E)).

INITIATOR and UNDERGOER are further divided into their intentional/non-intentional (AGENT vs CAUSE) or animate/inanimate (PATIENT vs OBJECT) representatives. INITIATOR and RESULT, are those that do not undergo any changes. PATH refers to means by which INITIATOR operates to achieve the desired RESULT focusing on UNDERGOER, the affected group. RESULT appears if there is any change being a consequence of the process, and defines the essence of this change. Roles with asterisks are LS specific (soc, poss, psych, etc.).

Inferences

The intersection of the two dimensions singles out groups of verbs that can be characterized by common sets of roles, alternation patterns, etc., giving rise to protoframes.

Some LA-dependent features:

- number of roles: Activities and Accidents, and especially States, tend to have one participant; Actions, being telic and intentional, have most often at least two of them;
- possibility of phase modification: only some Accident verbs are inchoative (contain the initial phase).

Some LS-dependent features:

- nature of roles and the character of interaction among them: social verbs have humans or human organizations as participants, abstract ones have processes and concepts as participants; production verbs have results; possession verbs have a pair of animate subjects and two objects of exchange; the body class frequently uses the part-whole alternation, etc.

Protoframes

- prod: INITIATOR produces RESULT (with the help of INSTRUMENT from MATERIAL):
 - prod-action: INITIATOR → AGENT (intentional);
 - prod-process: INITIATOR → EFFECTOR (non-intentional);
- psych: EXPERIENCER has some feeling/is in some state (due to CAUSE); INITIATOR arouses a feeling/state in EXPERIENCER; COGNIZER interacts with CONTENT; COMMUNICATOR addresses ADDRESSEE with CONTENT;
- phys: CAUSE affects UNDERGOER; UNDERGOER gets affected;
- loc: AGENT/THEME moves from SOURCE to DESTINATION; THEME moves/is moved from SOURCE to DESTINATION; INITIATOR moves THEME (from SOURCE to DESTINATION using INSTRUMENT in some MANNER).

Such preliminary schemes and roles can be treated as working hypotheses to be manually checked by linguists; they can also be useful for further consistency checks. For some of linguistic and/or technical tasks this level of the generalization of roles and frames may be useful in its present state as well.

Semantically enriched Walenty

Additionally the number of sense is indicated, as identical syntactic structures often correspond to different semantic frames, e.g., the verb *objechać*-1 in the sense 'bypass':

[subj{np(str)}:Agent Theme]+[obj{np(str)}:Location]+[{np(inst):Instrument}]

Jan objechał jezioro rowerem. 'John bypassed the lake on a bike.'

objechać-2 'visit':

[subj{np(str)}:Agent Theme]+[obj{np(str)}:Protagonist Collection]

Jan objechał całą rodzinę nowym samochodem. 'John visited the whole family by his new car.'

objechać-3 'scold':

[subj{np(str)}:Focus Protagonist]+[obj{np(str)}:Background Protagonist]

+[{np(inst):Content}]

Jan objechał sąsiada najgorszymi słowami. 'John has scolded his neighbour with the worst words.'



Institute of Computer Science
Polish Academy of Sciences

Natalia Kotsyba
n.kocyba@ipipan.waw.pl