

MICHAŁ MARCIŃCZUK<sup>A</sup>, MARCIN OLEKSY<sup>B</sup>, TOMASZ BERNAŚ<sup>C</sup>,  
JAN KOCON<sup>D</sup>, & MICHAŁ WOLSKI<sup>E</sup>

G4.19 Research Group, Wrocław University of Technology, Poland

<sup>A</sup>michal.marcinczuk@pwr.edu.pl ; <sup>B</sup>marcin.oleksy@pwr.edu.pl ; <sup>C</sup>t.b@gazeta.pl ;

<sup>D</sup>jan.kocon@pwr.edu.pl ; <sup>E</sup>misiek.wolski@gmail.com

## TOWARDS AN EVENT ANNOTATED CORPUS OF POLISH

### Abstract

The paper presents a typology of events built on the basis of TimeML specification adapted to Polish language. Some changes were introduced to the definition of the event categories and a motivation for event categorization was formulated. The event annotation task is presented on two levels — ontology level (language independent) and text mentions (language dependant). The various types of event mentions in Polish text are discussed. A procedure for annotation of event mentions in Polish texts is presented and evaluated. In the evaluation a randomly selected set of documents from the Corpus of Wrocław University of Technology (called KPWr) was annotated by two linguists and the annotator agreement was calculated. The evaluation was done in two iterations. After the first evaluation we revised and improved the annotation procedure. The second evaluation showed a significant improvement of the agreement between annotators. The current work was focused on annotation and categorisation of event mentions in text. The future work will be focused on description of event with a set of attributes, arguments and relations.

**Keywords:** information extraction; event recognition; corpus annotation

### 1. Introduction

Event recognition is a subtask of information extraction task. The goal of information extraction is to understand the meaning of a text at some level on which one can catch given type of information and present it in a structured manner. Event recognition focuses on finding in text references to some situations and extracting their descriptions. Event recognition has practical applications in many tasks from the field of natural language processing, like text summarization (Maybury, 1995), discourse analysis, events aggregation and reporting (Vossen et al., 2014; van Erp, Fokkens, & Vossen, 2014; Aggeri et al., 2014). Within the Clarin-PL project<sup>1</sup> we

<sup>1</sup>Project web page: <http://clarin-pl.eu/>.

plan to develop methods and tools for event recognition for Polish. We want to identify event mentions in Polish texts, categorise them on a coarse-grained level and identify event attributes, arguments and relations in order to enable deeper text understanding. In order to create and evaluate such tools we need a practical guideline for event annotation dedicated to Polish and a corpora annotated with events. According to our best knowledge, the most popular and widely used specification for event annotation is TimeML Annotation Guidelines Version 1.2.1 (Saurí, Littman, Knippen, Gaizauskas, Setzer, & Pustejovsky, 2006) (henceforth, TimeML). The specification has been already adopted to several languages, including Spanish (Saurí, Batiukova, & Pustejovsky, n.d.), Catalan (Saurí, Batiukova, & Pustejovsky, n.d.), French (Bittar, 2010) and Italian (Caselli, Bartalesi Lenzi, Sprugnoli, Pianta, & Prodanof, 2011). Applying an existing guideline for another language requires a careful study of that language phenomena and might need some adjustments concerning language-specific issues. In the following sections we present results of our work on adaptation the TimeML specification to Polish language and evaluation of the specification on Polish texts. In Section 2 we present a definition of event concept and what we understand as an event on the ontology level. Section 3 contains a typology of event categories and motivation for event categorization. In Section 4 we define the event mentions for Polish as a text-level representation of events. In Section 5 we present results of two evaluations of the guidelines for Polish performed on the KPWr corpora (Broda, Marcińczuk, Maziarz, Radziszewski, & Wardyński, 2012). In Section 6 we present a detailed procedure for event mention annotation which was created as a result of first evaluation. Section 7 presents a summary of our current work and future plans including event description with attributes, arguments and relations.

## 2. What is an event?

*Event* is one of the primary concepts in almost any upper-level ontology. According to the Oxford Dictionary *event* “is a thing that happens or takes place”.<sup>2</sup> In the Suggested Upper Merged Ontology ontology (Pease, 2011) (henceforth, SUMO) an event is represented as a concept called *Process*,<sup>3</sup> which is defined as following:

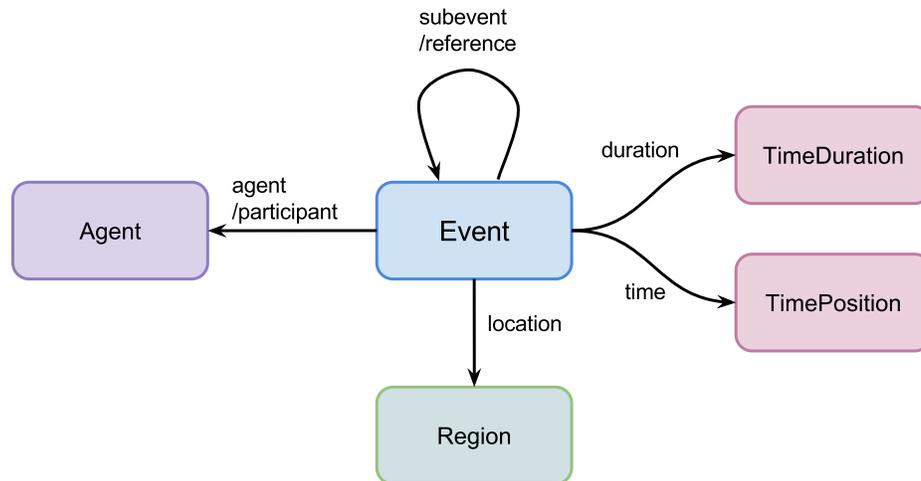
“The class of things that happen and have temporal parts or stages. Examples include extended events like a football match or a race, actions like pursuing and reading, and biological processes. The formal definition is: anything that occurs in time but is not an object. Note that a process may have participants ‘inside’ it which are objects, such as the players in a football match”

In other words *event* is anything that takes place in time (date, time and/or duration) and space (has a location), may involve agents (executor or participants), may contain or be part of other events and may produce some outcome (object). In our work we will consider as event all situations which can be mapped onto the *Process* concept or any concept which is a subclass of *Process* in the SUMO

<sup>2</sup><http://www.oxforddictionaries.com/definition/english/event>

<sup>3</sup><http://sigma-01.cim3.net:8080/sigma/Browse.jsp?lang=EnglishLanguage&flang=SUO-KIF&kb=SUMO&term=Process>

ontology. The relations of *event* with other concepts from the SUMO ontology are illustrated on the Figure 1.



**Figure 1:** Event relations with other concepts from SUMO ontology.

The states are also treated as *events* (Saurí et al., 2006) but they have a specific ontological status. It isn't simple to map the words denoting a state to *Process* in the SUMO ontology. As Vendler said suggestively, states are “that puzzling category in which the role of verb melts into that of predicate, and actions fade into qualities and relations” (Vendler, 1957, p. 109). This feature brought us to individual treatment of the mentions of the states.

### 3. Event Categories

We used seven coarse-grained categories of events, i.e. *action*, *state*, *reporting*, *perception*, *aspectual*, *intensional action* and *intensional state*. The categorisation was based on the TimeML guideline with some modifications. Instead of the *occurrence* term we used *action*. The *occurrence* category from TimeML refers only to specific temporally located events. Generics — actions which refer to some general rules (for example, a *boil* event in sentence “Water boils in 100 °C”) are not tagged. We noticed that the distinction between specific and generic events can be applied to any category of events what indicates that the event generality should be defined as an event attribute rather than its category. Taking into account Polish terminological tradition (e.g. Laskowski, 1998), we've decided to use the term *action* instead *occurrence*,<sup>4</sup> to accent its generality and to make visible the key opposition between the two core categories: *action* and *state*. In addition this change emphasizes the distinction between the *state/action* and *intensional state/action*. The remaining categories can be treated as auxiliary categories, as they refer to another events and introduce some additional information about the event.

<sup>4</sup>Although they are sometimes treated as synonyms (e.g. Mourelatos, 1978).

TimeML specification doesn't introduce the higher level classification. Still the categories of events can be divided into four groups in respect to two factors: *dynamycity* (*course in time*) and *event argument* (see Table 1). The *course in time* factor divides events into *static* and *dynamic* events. The *static* events endure or persists over some period of time and though they may provide the potential change, they do not constitute a change (Mourelatos, 1978, p. 192). The *event argument* factor indicates if the event have (or might have) an argument that is an event. For example the *start* event indicates the beginning of some other event.

**Table 1:** Groups of event categories.

	Without an event argument	With an event argument
Static	<i>state</i>	<i>intensional state</i>
Dynamic	<i>action</i>	<i>perception</i> <i>reporting</i> <i>aspectual</i> <i>intensional action</i>

In addition we've noticed that actions or states which connect with an event argument could be divided into two groups. Occurring some of the events in the text (i.e., *perception*, *reporting* and *aspectual*) signals that an event which is an argument occurred (or should have occurred) in the real world and occurring the other (i.e. *intensional state* and *intensional action*) doesn't gave such certainty. It wasn't the statement of TimeML specification authors but we treat it as the important remark for the future processing of extracted events.

Furthermore, we've decided to introduce a separate category for synsemantic verbs that occur with nominalizations (*light predicates*). Since they have specific grammatical function, they are described in the section on event mentions.

Figure 2 shows the final classification of events.

### 3.1. Action

*Action* represents a dynamic situation which occurs in time and space. The event could have some type of outcome that can be a product, achievement or change from one state to another.

Examples: build, dance, jump, running

### 3.2. State

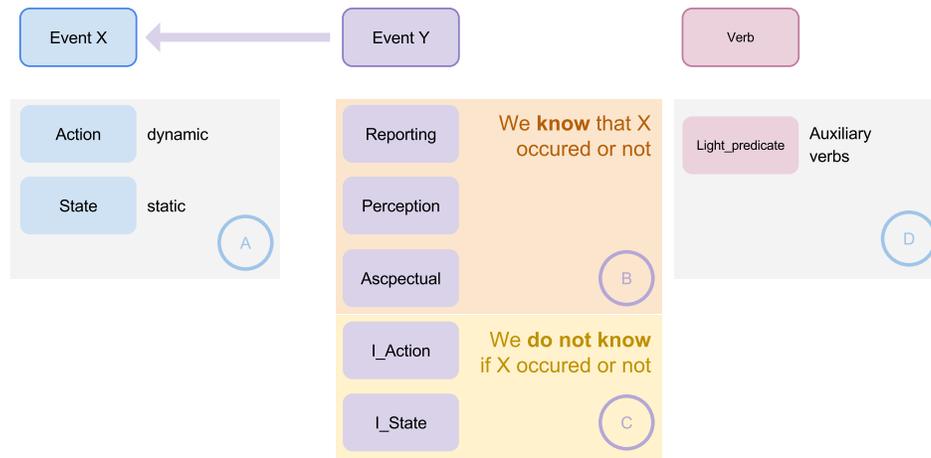
*State* represents a static situation. It refers to object attributes (Apresjan, 2000, pp. 47–48) or situations which are stable and does not change over given period of time (Laskowski, 1998, p. 153).

Examples: sleep, stand

### 3.3. Reporting

*Reporting* refers to a dynamic situation where an agent inform about an event or narrate an event. If the *reporting* refers to an *action* or a *state* then it is a strong indication that the *action* or *state* took place or was true.

Examples: report, explain, relation, telling



**Figure 2:** Categories of events.

### 3.4. Perception

*Perception* refers to a physical perception of an event by an agent. This class indicates that the agent was an observer of the event. The *perception* event is a strong indicator that the observed event took place or was true.

Examples: see, hear, observation, hearing

### 3.5. Aspectual

*Aspectual* refers to a dynamic situation which indicates a change of a phase of another event. The change can be (following TimeML):

1. Initiation — an event was started,
2. Reinitiation — an event was stopped and started again,
3. Termination — an event was stopped before it was completed,
4. Culmination — an event was completed,
5. Continuation — an event is continued.

If the *aspectual* event refers to an *action* or a *state* and it is not referred by any *intensional action* or *state*, then it is a strong indication that the action or state was true for some period of time.

Examples: start, stop, continuation, interruption

### 3.6. Intensional action

*Intensional action* is a situation where an agent declare his or her will to perform an action or give a command to another agent to perform an action. We cannot infer if the action was or will be performed in the future. The possible groups of intensional actions (following TimeML) are:

1. Attempt — the agent tried to do X but failed to accomplish it.
2. Delay — the agent postpone some action in time.

3. Avoid — the agent prevent same action which may happen.
4. Ask — the agent asks somebody to do something.
5. Promise — the agent promises to do something.
6. Propose — the agent propose to do something.

Comparing to TimeML we removed two groups of events from this category, i.e.: investigation (investigate, delve) and naming (name, nominate, appoint, etc.). Those two groups does not require any other event as an argument, thus they can be treated as an *action*.

Examples: try, delay, promise, ordering

### 3.7. Intensional state

*Intensional state* is a state which refers to some possible *actions* or *states*. It indicates, than an agent refers to some possible event, which may or may not occur in the future. Most of the intensional states are connected with mental activities, emotions and needs. The possible groups of intensional states (following TimeML) are:

1. Thinking — agent A thinks about doing X.
2. Will — agent A want to do X.
3. Expectation — agent A expect X to happen.
4. Emotions — agent A is afraid of doing X.
5. Needs — agent A need to do X.
6. Be ready — agent A jest ready to do X.
7. Ability — agent A jest able to do X.

Examples: believe, fear, wish, desire

## 4. Event Mentions in Text

Authors of TimeML state that “events may be expressed by means of tensed or untensed verbs, nominalizations, adjectives, predicative clauses, or prepositional phrases” (Saurí et al., 2006, p. 3). In our approach the list was limited to three types of mentions: tensed or untensed verbs (including participles and gerundial forms), nominalizations (in the wider lexicalistic sense) (Jędrzejko, 1993, pp. 53–56) and adjectives. The complete list of mention types is presented at Figure 3.

### 4.1. Verbs

Statement that events may be expressed by verbs seems almost axiomatic. Philosophers studying processes and their typology, e.g., Ryle (1949), Vendler (1957) or Kenny (1963) developed differentiated classificatory schemes of ‘action verbs’ or verb denotations (Seibt, 2004). Despite some difficulties,<sup>5</sup> their approach had

<sup>5</sup>There are some terminological difficulties with the “event”. As Comrie said, “it is often necessary to refer to the differences between states, events, processes, etc. (...). However, while ordinary nontechnical language provides, with a limited amount of systematisation, a metalanguage for these various subdivisions, it does not provide any general term to subsume them all.” (Comrie, 1989, p. 13). Metalanguage provided such terms as: “situation” (Lyons, 1977, p. 483), “eventuality” (e.g. Bach, 1986), and “event” (e.g. Langacker, 2010, p. 166). TimeML specification prefers the term “event” but according to Comrie suggestions (1989) and Polish terminological tradition we use in Polish specification general cover-term “sytuacja” (eng. situation).

Predicate			Annotation	
verb	personal form	in simple predicate		yes
		in complex predicate	copula	no
			modal	yes
			aspectual	yes
			generic	yes
			in periphrastic metaphorical expressions, idioms	yes
	impersonal form			yes
	predicative			yes
	infinitive			yes
	subject complement (ellipsis of copula)			no
<i>winię</i> -like verb			no	
<i>to</i>			no	
<i>żał...</i>			yes	
adverbs			no	
context			no	
accent, intonation			no	
pause			no	
gerund/nominalization			yes	
participle			yes	
adjective			yes/no	

**Figure 3:** Categories of event mentions.

a great impact on the future classifications including Polish tradition. For Laskowski (1998, p. 152–153), situation is a denotate of sentence constituted by verb.

It was important to introduce a method of annotating periphrastic predication. Ewa Jędrzejko points out several types of complex predicates (Jędrzejko, 2011, pp. 34–37):

- Standard nominal predicates [VCOP + NKONKR // Nabstr // Adj // Adv],
- The so-called modal predicates [VMOD + VINFIN] + ...,
- The phase-aspectual complex predicates [VFAZ + V//NA] + ...,

- The most common type of the VNA with basic ‘generic’ verbs [VGENER + NA//NE//Nabstr] + ...,
- Periphrastic predicates in the strict sense of the term [VMETAF// METAPRED + Nabstr//NA//NE ] + ...,<sup>6</sup>
- So-called phraseological predicates, i.e. ‘typical’ idioms functioning as verbs [Vmetafor + N +  $\diamond$  +  $\diamond$  + ] IDIOM

We’ve decided to exclude some verbs from annotation. Tagged elements should introduce enough information to classify the situation. Our assumption was that copulae and other auxiliary verbs (e.g. components of analytic future tense) are semantically (referentially) empty so they are not very useful for event extraction. Verbal part of modal predicates, phase-aspectual predicates and predicates with generic verbs may be called *light predicate* or *light verb* (Jespersen, 1965). According to Zolotova, Onipenko, and Sidorova (1999) they are modifiers (phase and modal) and compensators (accompanied by deverbal noun) There is no agreement concerning the definition and the semantics of light verbs (Kotsyba, 2014) but it was valid to include them to annotation as they carry a grammatical and very general but sufficient lexical meaning. We tag both elements of such predicates because they are relevant to different kind of event information — after that two tagged events will be linked as identical.

#### 4.2. Nominalizations

Nominalizations in which the original verb appears as a gerund or deverbative noun should be treated as equivalent of verb predications. According to Topolińska the result of nominalization is a nominal phrase correlated on syntactic level with a sentence or verbal realization of the same propositional model (Topolińska, 1984, p. 355). Still, we don’t consider nominalization as a simple transposition. Even the transformations, that aren’t structurally motivated, base on the principle of semantic transposition and suppletivism (“miłość” — “kochać”, “uroda” — “ładny”, “kłęska” — “pokonać”) and they could be secondarily verbalized by means of synsemantic verbs<sup>7</sup> (“czuć radość”, “mieć urodę”, “ponieść kłęskę” etc.) (Jędrzejko, 1993, pp. 46–47), so we decided to include them.

We have made an exception for one category. At the first stage of annotation we don’t mark the nominalizations for the states due to the specific ontological status of this situations and the features of their nominalizations (Mourelatos, 1978, pp. 204–210). It is an open question if we need to recognize them in the future.

#### 4.3. Other mentions

Taking into account the scope of our task it was important to consider all predicative expressions. Jodłowski (1976, pp. 31–33) introduced one of the first and basic classification. It includes many types of nonverbal predicates such as accent, intonation, context, pause or adverbs. Still, identification of these mentions would require context analysis or some additional data (e.g., conversational), so we decided to exclude them from annotation.

<sup>6</sup>“concrete’ verbs are used in these constructions metaphorically not only in a verbalising function” (Jędrzejko, 2011, p. 36).

<sup>7</sup>light verbs

As Sauri et al. (2006) stated events may be expressed by adjectives. Some of them are nominalizations so it is valid to annotate them. We have decided that the mentions that introduce other situations (i.e. that have an event argument) are the most important.

### 5. Annotators Agreement

The inter-annotator agreement was measured on randomly selected documents from the Corpus of Wrocław University of Technology called KPWr (Broda et al., 2012). We used the *positive specific agreement (psa)* (Hripcsak & Rothschild, 2005) as there are no negative decisions to count to measure the agreement between two linguists. The documents were annotated using the Annotator perspective from the Inforex system<sup>8</sup> (see Figure 4) (Marciniuk, Kocoń, & Broda, 2012).

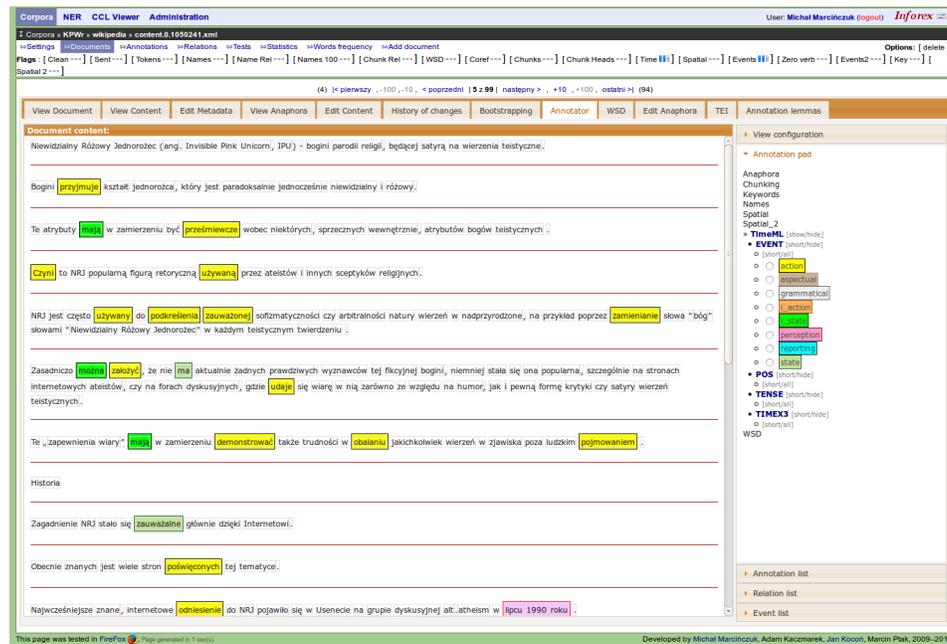


Figure 4: Document annotator perspective in Inforex.

In the first iteration we randomly selected 100 documents. The results are presented in Table 2. The agreement for event mentions without categorisation was ca. 85% and with categorisation it drops to 68%. The results show, that the most confusing categories were *state* (36.98%) and *light predicate* (39.60%). The best agreement was achieved for *aspectual* (86.79%) and *action* (77.89%). We have carefully analyzed the discrepancies between annotators and introduce some clarifications in the guideline. The major changes were:

- Categories which require another event as an argument (*aspectual*, *perception*, *reporting*, *intensional action* and *intensional state*) can be annotated regard-

<sup>8</sup>Web page: <http://www.nlp.pwr.wroc.pl/inforex/>.

**Table 2:** Agreement between two linguists (A and B) after first iteration.

Mention category	A and B	Only A	Only B	<i>psa</i>
Mentions without categorization	3184	393	664	85.76%
Mentions with categorization	2561	1016	1287	68.98%
action	2085	766	418	77.89%
state	213	92	634	36.98%
perception	20	2	37	50.63%
reporting	39	29	28	57.78%
aspectual	46	4	10	86.79%
intensional action	23	19	21	53.49%
intensional state	115	61	70	63.71%
light predicate	20	41	20	39.60%

less the argument is directly stated in the sentence or not (for example the event argument might be omitted or referenced by a pronoun).

- We have formulated a procedure for recognition state mentions. We have defined the following criteria:
  - If the mention as a lexical unit is present in the plWordNet<sup>9</sup> (Piasecki, Szpakowicz, & Broda, 2009) then the synset containing the lexical unit must be a direct or indirect hyponym of an artificial synset call “state verb”.
  - State has no dynamic.
  - State does not change in time.
  - Verb representing state has imperfect aspect.
  - Verb representing state does not have an perfective form.
- Passive construction does not indicate a state. Event category results from the semantic of the verb, not the grammatical construction. For example sentences “John was killed” and “Tom killed John” represent the same action of killing a person named John.
- We defined a procedure to determine if given mention should be annotated and with what category. The procedure consists of a set of yes-no questions. The procedure is presented in Section 6.

Next, we have performed a second evaluation to check, if the clarifications improved the agreement between annotators. In the second iteration we randomly selected another set of 50 documents. The documents were annotated by the same two linguists. The results for second evaluation are presented in Table 3. The agreement for mention annotation without categorisation raised from 85% to 93%. There were noticeable improvement in annotation of the categories which were annotated with unacceptable agreement (*state* and *light predicate*) in the first iteration (from 36.98% to 74.26% for *state*; from 39.60% to 50% for *light predicate*). The overall agreement in the second evaluation was significantly higher. However, there are

<sup>9</sup>Web page: <http://plwordnet.pwr.wroc.pl/>.

**Table 3:** Agreement between two linguists (A and B) after second iteration.

Mention category	A and B	Only A	Only B	<i>psa</i>
Mentions without categorization	2427	140	225	93.01%
Mentions with categorization	1856	346	430	82.71%
action	1531	253	198	87.16%
state	135	45	81	68.18%
perception	23	1	6	86.79%
reporting	21	18	14	56.76%
aspectual	26	2	8	83.87%
intensional action	20	8	57	38.10%
intensional state	88	13	48	74.26%
light predicate	12	6	18	50.00%

still two categories with low agreement which should be verified, i.e. *intensional action* with agreement of 38.10% and *light predicate* with agreement of 50%.

## 6. Mention Annotation Procedure

After the first iteration of measuring the annotation agreement we formulated a procedure for mention annotation and classification. The procedure consists of a series of yes-no questions.

Input: M — Mention

**Q1** Does M name an *action* (activity, accomplishment or achievement) or *state* (or their collections) in the real world?

YES ⇒ goto Q2  
NO ⇒ don't mark

**Q2** Is M a complex predicate?

YES ⇒ goto Q3  
NO ⇒ goto Q4

**Q3** Is M a metaphore or idiom?

YES ⇒ mark head only — goto Q7  
NO ⇒ goto Q4

**Q4** Is M a predicative expression?

YES ⇒ don't mark  
NO ⇒ mark separately both elements:

- verb (copula) — goto Q5;
- complement — goto Q1.

**Q5** Is M the LIGHT PREDICATE?

YES ⇒ goto Q6  
NO ⇒ goto Q7

**Q6** Is M connected with nominalization?

YES ⇒ mark as LIGHT PREDICATE

NO ⇒ don't mark

**Q7** Does M has an event argument?

YES ⇒ goto Q9

NO ⇒ goto Q8

**Q8** Could M have an event argument?

YES ⇒ goto Q9

NO ⇒ goto Q13

**Q9** Is M a *state*?

YES ⇒ mark as INTENSIONAL STATE

NO ⇒ goto Q10

**Q10** Does M inform about a phase of the situation?

YES ⇒ mark as ASPECTUAL

NO ⇒ goto Q11

**Q11** Does M describe the reporting of other situation?

YES ⇒ mark as REPORTING

NO ⇒ goto Q12

**Q12** Does M describe the perception of other situation?

YES ⇒ mark as PERCEPTION

NO ⇒ mark as INTENSIONAL ACTION

**Q13** Is S the state?

YES ⇒ mark as STATE

NO ⇒ mark as ACTION

## 7. Conclusions and future work

The evaluation showed that the annotation of event mentions is relatively simple — the agreement after second iteration was 93%. The categorisation of event mentions causes more problems and the agreement drops to 82%. This shows that the task is not trivial and if we want to obtain a good quality of data with high agreement the final annotation of whole KPWr corpus will require the “2+1” approach. This means that each document in the corpus will be annotated separately by two linguists and the differences will be evaluated by a third linguist — supervisor.

The future plan is to prepare guidelines for event description with attributes, arguments and relations. The attributes, we are considering, are:

- generality — is the event specific or general,
- polarity — is the form of the mention affirmative or negative,
- modality — is it assertoric, optative, imperative or interrogative,

- tense and aspect

Most of them (in particular generality) have to be annotated manually. Although there are tools<sup>10</sup> that could be used for automatic annotation of tense and aspect.

The events will be linked with their generic arguments, i.e. agent, time and location. In the last step we will mark the relations between the events. The categories of relations include identity and references between reporting, perception, aspectual, intensional action and intensional state and their event arguments.

### References

- Agerri, R., Agirre, E., Aldabe, I., Altuna, B., Beloki, Z., Laparra, E., de Lacalle, M. L., Rigau, G., Soroa, A., and Urizar, R. (2014). Newsreader project. In *30th Conference of the Spanish Society for Natural Language Processing (SEPLN)*.
- Apresjan, J. D. (2000). *Semantyka leksykalna: Synonimiczne środki języka*. (Z. Kozłowska, Z. & A. Markowski, Trans.). Warszawa.
- Bach, E. (1986). The algebra of events. *Linguistics and Philosophy*, 9, 5–16.
- Bittar, A. (2010). *Building a TimeBank for French: A Reference Corpus Annotated According to the ISO-TimeML Standard* (Unpublished Phd thesis). Université Paris Diderot.
- Broda, B., Marcińczuk, M., Maziarz, M., Radziszewski, A., & Wardyński, A. (2012). KPWr: Towards a free corpus of Polish. In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, & S. Piperidis (Eds.), *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul: European Language Resources Association (ELRA).
- Caselli, T., Bartalesi Lenzi, V., Sprugnoli, R., Pianta, E., & Prodanof, I. (2011). Annotating events, temporal expressions and relations in Italian: The It-TimeML Experience for the Ita-TimeBank. In *Proceedings of the 5th Linguistic Annotation Workshop, LAW V '11* (pp. 143–151). Stroudsburg, PA, USA: Association for Computational Linguistics.
- Comrie, B. (1989). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.
- Hripcsak, G. & Rothschild, A. S. (2005). Technical brief: Agreement, the f-measure, and reliability in information retrieval. *Journal of the American Medical Informatics Association*, 12(3), 296–298. <http://doi.org/10.1197/jamia.M1733>
- Jędrzejko, E. (1993). *Nominalizacje w systemie i w tekstach współczesnej polszczyzny*. Katowice: Uniwersytet Śląski. (*Prace naukowe Uniwersytetu Śląskiego w Katowicach*, 1335)
- Jędrzejko, E. (2011). The problematics of describing periphrastic predication: Between word and image. *Studies in Polish Linguistics*, 6, 27–44.
- Jespersen, O. (1965). *A modern English grammar — on historical principles* (Pt. 6: *Morphology*). London: Read Books.
- Jodłowski, S. (1976). *Podstawy polskiej składni*. Warszawa: PWN.
- Kenny, A. (1963). *Actions, Emotions and Will*. London: Routledge & Kegan Paul.
- Kotsyba, N. (2014). How light are aspectual meanings? A study of the relation between light verbs and lexical aspects in Ukrainian. In K. Robering (Ed.), *Events, arguments, and aspects: Topics in the semantics of verbs* (pp. 261–299). Amsterdam:

<sup>10</sup>Morphological taggers, i.e. WCRFT (Radziszewski, 2013).

- John Benjamins Publishing Company. (*Studies in Language Companion Series*, 152). Retrieved from <https://benjamins.com/catalog/slcs.152.07kot>
- Langacker, R. W. (2010). Control and the mind/body duality: Knowing vs. effecting. In E. Tabakowska, M. Choiński, & Ł. Wiraszka (Eds.), *Cognitive linguistics in action: From theory to application and back* (pp. 165–207). Berlin: Mouton de Gruyter. (*Applications of Cognitive Linguistics*, 14)
- Laskowski, R. (1998). Kategorie morfologiczne języka polskiego — charakterystyka funkcjonalna. In R. Grzegorzczkowska, R. Laskowski, & H. Wróbel (Eds.), *Gramatyka współczesnego języka polskiego: Morfologia*. Warszawa: PWN.
- Lyons, J. (1977). *Semantics* (Vol. 1). Cambridge: Cambridge University Press.
- Marcińczuk, M., Kocoń, J., & Broda, B. (2012). Inforex — a web-based tool for text corpus management and semantic annotation. In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, S. Piperidis (Eds.), *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul: European Language Resources Association (ELRA).
- Maybury, M. T. (1995). Generating summaries from event data. *Information Processing & Management*, 31(5), 735–751. [http://doi.org/10.1016/0306-4573\(95\)00025-C](http://doi.org/10.1016/0306-4573(95)00025-C)
- Mourelatos, A. P. D. (1978). Events, processes, and states. *Linguistics and Philosophy*, 2(3), 415–434. <http://doi.org/10.1007/BF00149015>
- Pease, A. (2011). *Ontology: A practical guide*. Angwin, CA: Articulate Software Press.
- Piasecki, M., Szpakowicz, S., & Broda, B. (2009). *A wordnet from the ground up*. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej.
- Radziszewski, A. (2013). A tiered CRF tagger for Polish. In H. Rybiński, M. Kryszkiewicz, M. Niezgodka, R. Bembenik, & Ł. Skonieczny (Eds.), *Intelligent tools for building a scientific information platform: Advanced architectures and solutions*. Berlin: Springer Verlag. Retrieved from [http://link.springer.com/10.1007/978-3-642-35647-6\\_16](http://link.springer.com/10.1007/978-3-642-35647-6_16)
- Ryle, G. (1949). *The Concept of Mind*. London: Barnes and Nobles.
- Saurí, R., Batiukova, O., & Pustejovsky, J. (n.d.). Annotating Events in Spanish TimeML Annotation Guidelines.
- Saurí, R. & Pustejovsky, J. (n.d.). Annotating Events in Catalan. TimeML Annotation Guidelines.
- Saurí, R., Littman, J., Knippen, B., Gaizauskas, R., Setzer, A., & Pustejovsky, J. (2006). TimeML Annotation Guidelines, Version 1.2.1.
- Seibt, J. (2004). *Process theories: Crossdisciplinary studies in dynamic categories. Studies in Philosophy and Religion*. Dordrecht: Springer Netherlands.
- Topolińska, Z. (1984). Składnia grupy imiennej. In Topolińska, Z. (Ed.) *Gramatyka współczesnego języka polskiego* (pp. 301–384). Warszawa.
- van Erp, M., Fokkens, A., & Vossen, P. (2014). Finding stories in 1,784,532 events: Scaling up computational models of narrative. In *Workshop on Computational Models of Narrative (CMN'14), Quebec City, Canada, July 31 – August 2*.
- Vendler, Z. (1957). Verbs and times. *Philosophical Review*, 66(2), 143–160. <http://doi.org/10.2307/2182371>
- Vossen, P., Rigau, G., Serafini, L., Stouten, P., Irving, F., Van Hage, W. (2014). News-Reader: Recording history from daily news streams. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)*, Reykjavik, Iceland, May 26–31.
- Zolotova, G. A., Onipenko, N. K., & Sidorova, M. I. (1999). *Kommunikativnaia grammatika russkogo jazyka*. Moskva: RAN.

---

**Acknowledgment**

Work financed by the Polish Ministry of Science and Higher Education, a program in support of scientific units involved in the development of a European research infrastructure for the humanities and social sciences in the scope of the consortia CLARIN ERIC and ESS-ERIC, 2015–2016.

The authors declare that they have no competing interests.

The authors' contribution was as follows: MM, MO and JK are responsible for the ideas in the research and the methodology; MO and MM created annotation guidelines and were responsible for inter-annotator agreement increasing and performed data analysis, TB, MW collected the examples; all authors participated in drafting the manuscript.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 PL License (<http://creativecommons.org/licenses/by/3.0/pl/>), which permits redistribution, commercial and non-commercial, provided that the article is properly cited.

© The Authors 2015

Publisher: Institute of Slavic Studies, PAS, University of Silesia & The Slavic Foundation