

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

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

<sup>A</sup>jan.kocong@pwr.edu.pl ; <sup>B</sup>fmichal.marcinczuk@pwr.edu.pl ;

<sup>C</sup>marcin.oleksy@pwr.edu.pl ; <sup>D</sup>t.b@gazeta.pl ; <sup>E</sup>misiek.wolski@gmail.com

## TEMPORAL EXPRESSIONS IN POLISH CORPUS KPWR

### Abstract

This article presents the result of the recent research in the interpretation of Polish expressions that refer to time. These expressions are the source of information when something happens, how often something occurs or how long something lasts. Temporal information, which can be extracted from text automatically, plays significant role in many information extraction systems, such as question answering, discourse analysis, event recognition and many more. We prepared PLIMEX — a broad description of Polish temporal expressions with annotation guidelines, based on the state-of-the-art solutions for English, mainly TimeML specification. We also adapted the solution to capture the local semantics of temporal expressions, called LTIMEX. Temporal description also supports further event identification and extends event description model, focusing at anchoring events in time, ordering events and reasoning about the persistence of events. We prepared the specification, which is designed to address these issues and we annotated all documents in Polish Corpus of Wrocław University of Technology (KPWR) using our annotation guidelines.

**Keywords:** PLIMEX; TIMEX; temporal expressions; TimeML; machine learning; natural language processing; information extraction

### 1. Introduction

The recognition of temporal expressions and events is one of the major tasks in Information Extraction (henceforth IE) and plays significant role in many natural language processing systems. It became an active area of research after the first attempts to apply the temporal and event-based reasoning in language and text to IE.

Currently we live in a dynamic world, full of information given to us mostly as unstructured data (text, sound, view). The analysis of information in that

form (e.g. natural language) is still a challenging task for researchers. Here we focus on the aspect of changes over time, which can be tracked in text written in natural language. Reasoning about how the world changes requires extracting of information about temporally grounded events.

Temporal expressions refer to the time and tell us when something happens, how long something lasts or how often something occurs. To capture the semantic meaning of temporal expression we often have to analyse the whole context of document. Human usually is conscious of its location in time and knows what is the day and month of the year or what is the part of the month (begin or end). Thus people use expressions like *15th of december*, *the next week*, *4 days before* to refer to the specific point in time. Even newspaper articles, written mostly in formal language, often contain temporal expressions, which global semantic meaning can be deduced by analyzing the whole context of the document or even document metadata, e.g. document creation/publication time. Knowing of full temporal context is essential to determine the exact date, to which the temporal expression refers. These examples do not cover the complexity of temporal expression understanding. Sometimes the part of text describes past or possible future, but it is not explicitly stated. However there are some clues in other part of text to find out what tense is given. Sometimes the purpose of using temporal expression is not the reference to real world, but just to describe some fictional events. Other kind of problems is to determine the temporal function of the expression, e.g. *three days* can be used to describe duration (how long something lasts) or point in time, e.g. *in three days*. Automatic systems should distinguish between different categories of temporal expressions to interpret their semantic meaning.

Another purpose is the normalization of temporal expressions. It is crucial to share information between different computer systems. For example a software developer from Poland could create the system, which recognize the temporal expression *dziewiątego grudnia* (*ninth of december*) and normalize it to *09.12.2014*. Another developer from USA could also create similar system, which normalizes the same temporal expression to *12/9/2014*. A serious problem may be dealing with two forms of the same date.

This article describes the overview of existing specifications for marking up text with temporal expressions and what is missing in some concepts. It shows how to classify temporal expressions and how to represent and annotate them. It presents annotation and normalization guidelines with the special focus given to Polish language. Finally it describes annotating of Polish Wrocław University of Technology Corpus, validating the result of annotation with comparison to other corpora annotated with temporal expressions and a summary of the planned work for the future.

## 2. Related Works

### 2.1. Information Extraction

Information Extraction task is a subpart of natural language engineering and it focuses at extracting of structured information from unstructured documents. In case of human language texts processing we can observe many directions of IE, with subtasks aimed at recognition of:

**Named entities** — recognition of known entity name, e.g. person names, organization names, location names. Several solutions were proposed for Polish as open source frameworks, e.g. *Liner2* (Marcinićzuk, Kocoń, & Janicki, 2013; Marcinićzuk, & Kocoń, 2013), which consists of several universal methods for sequence chunking like dictionary look-up, statistical processing or pattern matching.

**Events** — event recognition (Andersen, Hayes, Huettner, Schmandt, Nirenburg, & Weinstein, 1992; Llorens, Saquete, & Navarro-Colorado, 2010) is important subtask in more complex natural language processing tasks like question answering<sup>1</sup> or automatic summarization (Daniel, Radev, & Allison, 2003).

**Temporal expressions** — many texts, mainly newspapers, describe the world by talking about people, events and states they participate in. The exact temporal locations of events is rarely explicit, many temporal expressions are vague. Automatic extraction of temporal expressions identifies when something occurred by the recognition and normalization of expressions which refer to time. Often used in automatic question answering systems (Pustejovsky, Knippen, Littman, & Saurí, 2005).

The above list is not exhaustive and still the number of IE subtasks increases, because it is becoming an interesting topic for many researchers. IE allows the information from the text to be expressed in a structured way and allows to join the information from multiple sources of documents.

## 2.2. Temporal Information

Temporal expressions in text refer to the time. According to (Mazur, 2012) these expressions are divided into two main categories: instants and intervals. These are atoms of time, which can be used to represent and reason about time. In the literature we can find many terms to describe instants, e.g. *a time point*, *a point*, *a point in time*, *a moment*. Also interval sometimes is called period like in (Benthem, 1983), where Benthem uses interval as something which is between boundaries. On the other side in (Allen, 1995) we find interval temporal expressions in Benthem's meaning denoted by the term *duration*. The main difference between instants and intervals is that instants have no duration (treated as a feature of a period).

One of the most widely used specification for English to describe temporal information in natural language corpora is TimeML (Saurí, Littman, Knippen, Gaizauskas, Setzer, & Pustejovsky, 2006). It was developed in the context of a workshop TERQAS,<sup>2</sup> as a part of the ARDA-funded program AQUAINT<sup>3</sup> in a multi-project effort to improve the performance of question answering systems over documents written in natural language (Pustejovsky, Ingria, Saurí, Castano, Littman, Gaizauskas, Setzer, Katz, & Mani, 2005). The aim of this research was to improve accessing of information from text through content rather than keywords. The main problem was the recognition of events and their temporal anchorings.

<sup>1</sup><http://www.timeml.org/site/terqas/index.html>

<sup>2</sup>Time and Event Recognition for Question Answering Systems. An Advanced Research and Development Activity Workshop on Advanced Question Answering Technology

<sup>3</sup><http://www.informedia.cs.cmu.edu/aquaint/index.html>

Events are naturally anchored in time and these temporally grounded events are the source of information from which we can reason about the world changes. Also other entities and their properties change over time, therefore a database of assertions about entities will be incomplete if it does not capture how these entities are temporally updated (Pustejovsky, Ingria, et al., 2005).

Previous works, focused at temporal and event-based reasoning task, were *ACL Workshop on Spatial and Temporal Reasoning (2001)* and *LREC Workshop “Annotation Standards for Temporal Information in Natural Language (2002)*. Several papers from these workshops were about the time representation and identification (Saurí, Littman, Gaizauskas, Knippen, Setzer, & Pustejovsky, 2006). One of the results created on the basis of workshops’ conclusions is TimeML — Markup Language for Temporal and Event Expressions. It is designed to address these issues which remain unresolved during workshops. In article (Pustejovsky, Ingria, et al., 2005). four major problems were introduced:

- Time stamping of events
- Ordering events with respect to another
- Reasoning with contextually underspecified temporal expressions (such as *last week, two weeks before*)
- Reasoning about the persistence of events (how long does an event last)

We see that capturing of the temporal information is important in the process of reasoning about the changes described in text documents. This document describes the first stage of the temporal information capturing and focuses at annotating of the temporal expressions in text.

### 2.3. PLIMEX

PLIMEX is the temporal annotation language, suitable to describe temporal expressions in Polish text documents. It is based on TIDES Instruction Manual for the Annotation of Temporal Expressions (Ferro, 2001), which describe TIMEX2 annotation format. TIDES manual is also the core of TIMEX3 annotation format, used in TimeML specification (Saurí et al., 2006). Both documents present how to use the special Standard Generalized Markup Language tags to annotate temporal expressions, by inserting them directly into the text. In section (Annotation Process) we describe our solution of annotating temporal information. We adapted types of temporal expressions from TIMEX3: DATE, TIME, DURATION and SET. These specific types are described in section (Types of Temporal Expressions in PLIMEX).

TimeML was successfully adapted to many languages and one of the most widely used rule-based system (HeidelTime (Strötgen, & Gertz, 2013; Strötgen, Zell, & Gertz, 2013)), which uses TIMEX3 annotation standard, currently supports 11 languages: English, German, Dutch, Vietnamese, Arabic, Spanish, Italian, French, Chinese, Russian, and Croatian. Our research gives the opportunity to create the cross-domain temporal tagger which supports Polish.

### 3. Types of Temporal Expression in PLIMEX

In this section we define the TimeML types of temporal expressions, adapted to Polish. All English translations of Polish examples are given in parentheses. The extent of the annotation in text (if needed) is marked with square brackets.

#### 3.1. DATE

DATE is a type of temporal expression which denotes a point on a timeline, i.e., a unit of time greater than or equal to the day. The key question is *when*.

In general, different categories of DATE can be mentioned. The first one consider dates, like *czerwiec 2004* (*June 2004*) or *15.08.2014*, as contextually independent. At this point, the researchers of this issue are in agreement with each other, although the terminology that is used differs. The most commonly used is fully-specified (Schilder & Habel, 2001; Han, Gates, & Levin, 2006) and then explicit (Saquete, Muñoz, & Martínez-Barco, 2003; Schilder, 2004), absolute (Negri & Marseglia, 2005; Vicente-Díez, Samy, & Martínez, 2008) and other names.

Other temporal expressions are named and categorized variously. It is possible to mention the types of temporal expressions such as:

- (1) *wczoraj* (*yesterday*), *jutro* (*tomorrow*)
- (2) *[5 dni] temu* (*5 days ago*), *[5 dni] później* (*5 days later*)
- (3) *28 lipca* (*July 28*)

Some researchers use only one name for all of these types, like relative (Negri & Marseglia, 2005), implicit (Saquete et al., 2003), indexical (Schilder, 2004) and others. In other literature, two (e.g. Han et al. uses deictic for (1) group and relative for (2,3) group) or three names (e.g. Smith mentions deictic for (1), dependent for (2) and flexible anchoring for (3), while Busemann et al. uses deictic, anaphoric and underspecified for (1, 2, 3)) are in use. Mazur provides a full list of the terminology used in the literature.

We need also to mention point-temporal expressions with vague boundaries: *zima* (*winter*), *weekend* (*weekend*). These units can be used like any type of point expression:

**fully-specified** — *zima 2007* (*winter of 2007*)

**deictic** — *ostatnia zima* (*last winter*)

**anaphoric** — *następująca zima* (*the following winter*)

**underspecified** — *zima* (*winter*)

Other examples of DATE:

- (4) *[poniedziałek, 16 marca 1985 roku]* (*on [Monday, 16th March 1985]*)
- (5) *to wydarzyło się [drugiego listopada]* (*it happened on [the second of November]*)
- (6) *w [październiku 1963 roku]* (*in [October 1963]*)
- (7) *to będzie we [wtorek osiemnastego]* (*it will be on [Tuesday, the eighteenth]*)
- (8) *byłem nad jeziorem [latem tamtego roku]* (*I was at the lake in [the summer of that year]*)

### 3.2. TIME

It is type of point expression that describes temporal expressions which refer to the time of day, even if it is not clearly defined. The key question is also *when*. For example *Smith wrócił* (*Smith returned*):

- (9) [*za dziesięć trzecia*] (at [*ten to three*])
- (10) [*dwadzieścia po dwunastej*] (at [*twenty past twelve*])
- (11) at [*eight in the morning*] (o [*ósmej rano*])
- (12) at [*9 am on Friday, October 1st, 1999*] (o [*9.00 w piątek 1 października 1999 roku*])
- (13) [*yesterday late at night*] ([*wczoraj późno w nocy*])
- (14) [*last night*] ([*wczoraj w nocy*])

### 3.3. DURATION

DURATION, in contrast to DATE, has two points on a timeline associated with it — a start and an end point. A different name used in literature is period (Saquete et al., 2003). The key question is *how long*.

Sometimes range expressions are also included to this group (Mizobuchi, Sumitomo, Fuketa, & Aoe, 1998), but these expressions can be treated as separate points in time (Mani & Wilson, 2000). For example *Smith był tutaj* (*Smith stayed there*):

- (15) [*dwa miesiące*] (for [*two months*])
- (16) [*48 godzin*] (for [*48 hours*])
- (17) [*trzy tygodnie*] (for [*three weeks*])
- (18) [*całą ostatnią noc*] ([*all last night*])
- (19) [*20 dni*] w lipcu ([*20 days*] in July)
- (20) przez [*trzy godziny*] w zeszły poniedziałek (for [*three hours*] last Monday)

If a specific piece of information, which relates to the calendar, occurs in the temporal expression, then DATE is the right type of annotation. This is true even if the context suggests that this type of temporal expression indicates the duration of an event, e.g. [*Cały 1985*] przebywał na emigracji ([*The entire 1985*] he lived in exile).

#### 3.3.1. DURATION or DATE/TIME

The location in time has two forms. The first is a temporal expression which directly indicates a point in a timeline (e.g. *July 3, 20:30, Sunday*). The signal of a temporal expression, which appears in this case, indicates the relation of the temporal expression to the described situation (a situation occurs at this time), for example *at 20.30 on Sunday*.

The second form of the localization is associated with the determination of a period which separates the moment of the situation from another point in time. These temporal terms are prepositional phrases. The signal of a temporal expression indicates a different relation than the one included. For example, in the expression *za [sześć godzin] będę w domu* (in [*six hours*] I'll be home), the preposition indicates

that it took *six hours* to the point in time associated with the beginning of the given situation. The temporal expression itself in this situation is the DURATION type. This problem can be solved as follows:

- If a temporal term answers the question *when* or *how long* and the temporal expression answers the question *how long, how many* (e.g. *how many days, months*), then we have to deal with DURATION;
- If a temporal term answers the question *how long* or *when* and temporal expression answers the question *when, what* or another question unrelated to the quantification, then we have to deal with DATE or TIME.

Examples:

- (21) *jak długo?* — *przez 4 dni* (*how long?* — *for 4 days*)  
*przez ile?* — *4 dni* (*for how many?* — *4 days*) = DURATION
- (22) *kiedy?* — *za pięć dni* (*when?* — *in five days*), but:  
*za ile?* — *pięć dni* (*in how many (days)?* — *five days*) = DURATION
- (23) *kiedy?* — *o ósmej* (*when?* — *at eight*)  
*o której?* — *ósmej* (*what (time)?* — *eight*) = TIME
- (24) *jak długo?* — *od wczoraj* (*how long?* — *from yesterday*), but:  
*od kiedy?* — *wczoraj* (*since when?* — *yesterday*) = DATE

### 3.4. SET

The SET expression is a type of temporal expression, relating to more than one instance of a time unit — either a point or a period. The key question is *how often*.

This type of an expression is named variously, but the most popular name is SET (Mani & Wilson, 2000). Han et al. also distinguished (inside of this type):

**recurrence** — e.g. *co poniedziałek* (*every Monday*)

**rate** — e.g. *trzy razy na miesiąc* (*three times per month*)

Other subtypes are given by Niemi and Koskeniemi (2007):

**parity expressions** — e.g. *nieparzyste poniedziałki miesiąca* (*odd Mondays of the month*)

**ordinal expressions** — e.g. *co czwarty rok stulecia* (*every fourth year of the century*)

**containment expressions** — e.g. *lata z 29 lutego* (*the years with a 29th of February*)

Examples — *Jan wraca pijany* (*John comes back drunk*):

- (25) *dwa razy w tygodniu* (*twice a week*)  
(26) *co dwa dni* (*every two days*)  
(27) *każdej niedzieli* (*every Sunday*)

#### 4. Annotation Guidelines

The aim of that guideline is to describe which temporal expressions are annotated, how to determine the semantic meaning of temporal expression and how to identify correctly boundaries and the class of temporal expressions, which occur in text.

##### 4.1. What to Annotate

Our purpose is to annotate expressions which refer to time. We define three types of temporal expressions:

- obvious and precise
- obvious and not precise
- not obvious

##### 4.1.1. Obvious and Precise Temporal Expressions

We call temporal expression obvious and precise if it is possible to determine the exact meaning of the expression, for example:

- (28) *W dniu [15.11.1999 r.] doszło do wypadku (On [15.11.1999] there was an accident)*
- (29) *Przez [dwa dni] nie mógł mówić (He could not speak for [two days])*

These temporal expressions can be assigned to one of the following classes: DATE, TIME, DURATION or SET.

##### 4.1.2. Obvious and Not Precise Temporal Expressions

We call temporal expression obvious and not precise if it is not possible to determine the exact meaning of the expression (*fuzzy* reference to time) with the obvious type of time quantification (e.g. *year, month, day, etc.*), as it was underlined in the following example:

- (30) *Przez [kilka lat] pozostawał w ukryciu (He remained in hiding for [several years])*
- (31) *Miał czekać za drzwiami przez [następnych parę minut] (He had to wait behind the door for the [next few minutes])*

We annotate temporal expressions which are part of the event name without determining if it refers to named entity or to common expression, for example:

- (32) *Kampania [wrześniowa] była wielką klęską (The [September] Campaign was a huge failure)*

##### 4.1.3. Not Obvious Temporal Expressions

We call temporal expression not obvious if we do not know the type of time quantification, for example:

- (33) *W przeszłym okresie byli do tego zdolni (In the past period they were capable of this)*

These are mainly expressions like *bieżący, przeszły, przyszły* (*current, former, future*). For these expressions the new categories were introduced: PRESENT\_REF, PAST\_REF, FUTURE\_REF and DONT\_KNOW. The last one is used if it is difficult to determine to which category the expression should be assigned. Later we decided not to annotate temporal expressions which are not obvious and we focused only at the following types: DATE, TIME, DURATION and SET.

#### 4.2. What Not to Annotate

We do not annotate temporal expressions referring to holidays, because they are taken under the consideration at another annotation level (the name of an event), for example:

(34) *Dzień Niepodległości* (*Independence Day*)

(35) *Wielkanoc* (*Easter*)

We decided also not to annotate some temporal expressions which are relational (describe temporal location with respect to event or another point in time (Laskowski, 2003, 2005), mainly:

- expressions which determine the order of events, e.g. *poprzednio, wcześniej, później, następnie* (*previously, earlier, later, next*)
- prepositional phrases referring directly to other events, e.g. *po wojnie, przed świtem* (*after the war, before the dawn*)
- relative anaphoric pronouns, e.g. *tego dnia, w tym czasie, w tamtym roku* (*that day, at this time, in that year*)

#### 4.3. Exceptions

Relational expression can be part of the temporal expression as the modifier of temporal expression. The modifier is underlined in the following example:

(36) *Spotkali się [następnego dnia]* (*They met the [next day]*)

In another example we do not include underlined expression, because it is not the modifier of temporal expression:

(37) *Spotkali się [dwie godziny] później* (*They met [two hours] later*)

#### 4.4. Control Questions

1. Are you able to determine the location in time or duration of an event without referring to other event?
  - NO — do not annotate
  - YES — go to 2.
2. Does the value of temporal expressions can be precisely determined?
  - NO — go to 3.
  - YES — we annotate as one of the following: DATE, TIME, DURATION, SET

3. Does the expression contain obvious type of time quantification (time unit)?

NO — do not annotate

YES — annotate

#### 4.5. Annotation Extent

This section describes how to identify the extent of the annotation. The possible grammar categories of temporal expression’s constituents are defined in Table 1.

**Table 1:** Possible grammatical categories of temporal expression’s constituents.

| Category         | Example (Polish) | Example (English) |
|------------------|------------------|-------------------|
| noun             | latami           | years             |
| noun phrase      | tego lata        | this summer       |
| adjective        | bieżący          | current           |
| adjective phrase | pierwszy piątek  | first friday      |
| adverb           | ostatnio         | recently          |
| adverb phrase    | bardzo późno     | very late         |

##### 4.5.1. Temporal Expressions as Prepositional Phrases

If a temporal designation<sup>4</sup> is a prepositional phrase, then (if it is interpreted as in (Laskowski, 2003, 2005)) a preposition is treated as a signal and the remaining part of a phrase as a temporal expression (interpreted as in (Saurí et al., 2006)). The extent of the annotation cannot contain a preposition. The Table 2 with examples is not exhaustive, but shows the most common occurrences of temporal expressions as constituents of prepositional phrases. There are two exceptions where

**Table 2:** The most common occurrences of temporal expressions as constituents of prepositional phrases

| Text                  |                          | Temporal expressions |                |
|-----------------------|--------------------------|----------------------|----------------|
| (PL)                  | (EN)                     | (PL)                 | (EN)           |
| przed czwartkiem      | before Thursday          | czwartkiem           | Thursday       |
| w południe            | at noon                  | południe             | noon           |
| w czwartek            | on Thursday              | czwartek             | Thursday       |
| przez ostatnie 10 lat | during the last 10 years | ostatnie 10 lat      | last ten years |

the preposition can be the part of the temporal expression:

- The preposition (underlined> is the integral (and often internal) part of the temporal expression:

<sup>4</sup>In Polish literature terms *wyrażenie temporalne* (*temporal expression*) and *określenie temporalne* (*temporal designation*) are often used interchangeably (Laskowski, 2003, 2005). We use the temporal designation to describe the expression containing the full syntactic component of the expression which refers to the time. The temporal expression in TimeML specification (Laskowski, 2003, 2005) is only the part of the temporal designation, which denotes time (without temporal relation signals, mainly prepositions).

(38) za pięć minut dwunasta (five minutes to twelve).

- The preposition in the following expressions: co rok, co miesiąc, co czwartek (each year, each month, each Thursday). It is currently the only option to indicate the temporal expression which refers to recurring event (set of times). The class of these temporal expressions is SET.

#### 4.5.2. Modifiers of Temporal Expressions

Temporal expressions contain modifiers at every position in the expression. The Table 3 contains examples of temporal expressions with modifiers as their constituents.

**Table 3:** Examples of temporal expressions with modifiers as their constituents

| Temporal expression (PL) | Temporal expression (EN) |
|--------------------------|--------------------------|
| tego zimnego dnia        | this cold day            |
| następnego dnia          | next day                 |
| wiele niedziel           | many sundays             |
| więcej niż miesiąc       | more than one month      |
| nie mniej niż 60 dni     | not less than 60 days    |

If a temporal expression is the modifier of a phrase (underlined) which is not a temporal expression, then we annotate only the temporal expression, for example:

(39) Czy przyjdiesz na [sobotni] piknik w parku? (Do you come for a [Saturday] picnic in the park?)

#### 4.5.3. Event-anchored Temporal Expressions

Event-anchored temporal expression refer to an event (underlined), e.g. [dzień] po naszym spotkaniu (the [day] after our meeting). The extent of the annotation covers only the temporal expression. Events are annotated separately with the tag EVENT (event description is not the part of this work).

#### 4.5.4. Ranges

In case of temporal expressions which describe the range, we annotate only the starting and the ending point (according to TimeML specification (Saurí et al., 2006)) as DATE or TIME. The extent of the annotation cannot contain prepositions or their replacement characters. In Example 40 we have two separate temporal expressions:

(40) Wizyta premiera będzie miała miejsce [6]-[8 sierpnia] (Prime Minister's visit will take place on [6]-[8 August])

We propose the new type of the temporal expression, called RANGE. If the expression includes two temporal expressions denoting the start and the end point, we also annotate the whole extent as RANGE (underlined):

(41) W latach [1945]-[1954] w budynku mieścił się sąd (In the years [1945]-[1954] the building housed the court)

- (42) *Bank jest czynny w godzinach [8.00]–[15.00] (The bank is open [8.00]–[15.00] hours)*
- (43) *Pracował tam od [2002] do [2005] roku (He worked there from [2002] to [2005])*

#### 4.5.5. Joined Temporal Expressions

In the case of temporal expressions which contain conjunctions, we annotate them in a similar way as constituents of RANGE expressions, defined in TimeML (Saurí et al., 2006). In Example 44 we annotate two separate temporal expressions:

- (44) *Wielka Brytania pozostanie poza unią walutową przynajmniej [rok] lub [dwa] (Great Britain will remain outside the monetary union at least a [year] or [two])*

#### 4.5.6. Using One Tag in The Complex Temporal Expressions

We treat temporal expression as indivisible in the following cases:

- Expressions denote the same point on a timeline:

(45) *dwunasta w południe (twelve at midday)*

- Expressions denote a structured hierarchy of the description of a single point in a timeline:

(46) *[piątek wieczorem]; [listopad 1943] ([friday evening]; [November 1943])*

- Even if there are commas in the complex temporal expression, we annotate it if a structured hierarchy of the description is preserved and it refers to a single point in a timeline:

(47) *[24 września, godzina 19:00] ([September 24, 19:00])*

## 5. Normalization Guidelines

### 5.1. The Interpretation of Temporal Expressions

The purpose of the interpretation of temporal expressions is to achieve the contextual meaning of those expressions. From the information processing point of view it is a case of normalization of these temporal expressions. They need to be given in an unambiguous format that can be understood by a computer. A solution that takes both the local and the global semantics of an expression into account needs to be adopted.

The local semantics is the result of an analysis of a temporal expression without taking the context of the document into account. Only tokens building the temporal expression and tokens forming a sentence in which the expression occurs are taken into account. One of the proposed specifications which takes the semantics of this type into account is the LTIMEX (Mazur, 2012).

In the case of a determination of the global semantics, the context of the document and, in some cases, other sources of knowledge (e.g. dates of holidays and

well known events) are taken into account. This level of semantic meaning is defined in the TimeML specification for both TIMEX2 and TIMEX3.

We need to adopt the normalization process of temporal expressions which takes into account the interpretation at both the local and the global levels. First, the local semantics is determined by an analysis of the structure of a temporal expression and of a local context. Then the expression is analyzed in the context of the entire document, taking into account other sources of knowledge (geolocation, calendar etc.).

### 5.2. The Local and The Global Semantics

The semantic meaning of a temporal expression is determined independently from the information concerning the context of the expression. The local semantics reflect the meaning of the individual tokens forming the expression. Sometimes the local and the global semantics are the same, especially for unambiguous points in time:

- (48) *Pierwsza wojna światowa wybuchła [28 lipca 1914 roku] (The First World War broke out on [July 28, 1914])*
- (49) *Przysłał nam swoje zdjęcia z [lat dziewięćdziesiątych dwudziestego wieku] (He sent us his pictures from [the 1990's])*

Expressions which are context-dependent do not contain enough information in order to obtain the global semantics. Only the contextual analysis of these expressions makes the definition of an exact point in time:

- (50) *Otwarcie gospodarcze stało się faktem na początku [lat dziewięćdziesiątych] (The opening up of the economy became a reality at the beginning of [the 90's])*
- (51) *Byłem [wczoraj] w kinie na ciekawym filmie ([Yesterday] I was at the cinema for an interesting film)*

The local semantics are always the same, regardless of the broader context. For example *lipiec* (*July*) always refers to the seventh month of the year and *wczoraj* (*yesterday*) always refers to the day before the current date. The global semantics depend on the context and, for that reason, the interpretation of these examples may differ in different contexts. The value of such an expression will be different, depending on the context. In the case of vague expressions, the interpretation process assumes the determination of a maximally precise meaning of the temporal expression. The global semantics must take into account the local information and all of the information that can be deduced from an analysis of the document. For example, at the local semantics level, expression *7 rano w listopadowy poranek* (*7 am on a morning in November*) specifies a month and an hour. To determine the global semantics, other information (different than the hour and the month, e.g. the day and the year) must occur in a broader context and refers to the analyzed expression. Sometimes that information does not exist, and in such cases the local and the global semantics are the same. The components of the reference date (date of a creation of the document or the date in the text) do not always specify a date directly. Consider the following sentence:

(52) *Wyjeżdżamy do Nowego Jorku w [lutym] (We go to New York in [February])*

If this sentence is derived from a document created in April 2014, the temporal expression in this sentence refers to February 2015.

Finding the right reference date is important in the normalization process of temporal expressions, for example:

- Reference to the date of the utterance, depending on the document's source. It can be the document creation time, the date of the publication or the date of the last modification (especially in the case of newspaper articles), but also the date of sending an e-mail or a date of the communicator message, etc. In general, all these dates are mentioned in TimeML specification (TIMEX3) as DCT (document creation time).
- Reference to another temporal expression in the text (temporal expressions that directly relate to other temporal expressions used in the text), for example:

(53) *Zbigniew urodził się 13 grudnia 1942 roku. Trzy dni wcześniej zmarł jego ojciec. (Zbigniew was born on December 13, 1942. His father had died three days earlier.)*

It is necessary to determine the correct meaning of the temporal expression A in the first sentence and to refer the local semantics of the temporal expression B in the second sentence to the date set by A, in order to determine the global semantics B.

- Reference to the event described in the text (temporal expressions that refer to elements described in the TimeML specification as descriptions of events), for example:

(54) *[Trzy dni] po wypadku Zosia odzyskała przytomność (Zosia regained consciousness [three days] after the accident)*

In this example, the term *wypadek* (*accident*) is an event description and refers to the change of state in reality. As long as it is not possible to directly determine when the accident occurred, it is not possible to determine the global semantics in the temporal expression *trzy dni* (*three days*). However, these types of expressions are not studied in this article.

### 5.3. Representation of The Global Semantics

In this article the global semantics are adopted, as e.g. in the TimeML specification, in order to describe temporal expressions (TIMEX2 and TIMEX3, both of which include the same normalization format). The attribute VAL (Value) is important in the normalization context. It specifies the normalized value of a temporal expression. VAL is a textual representation of a temporal expression, which is assigned using guidelines described in the ISO-8601 standard. In order to describe it, the first letters of the English (but not only English) names that specify time (the size of the letters is important) are used (e.g. **Y**ear, **M**onth, **D**ay, **h**our, **m**inute, **B**C

— before Christ, **AD** — Anno Domini, **T**ime, **M**orning, **H**alf, **Q**uarter, **S**ummer, **P**eriod).

Each temporal expression can be determined with respect to the proper type, using the coding proposed in the ISO-8601 standard, for example:

- calendar date: YYYY-MM-DD,
- week of the year: YYYY-Wxx,
- hour: hh:mm:ss,
- date and hour: YYYY-MM-DDThh:mm:ss,
- duration (the number of weeks): PxW.

The table below shows example values of the VAL attribute and the semantic meaning (Mazur, 2012):

**Table 4:** Examples of VAL attributes and the semantic meaning of temporal expressions.

| VAL                 | Meaning (EN)            | Meaning (PL)                 |
|---------------------|-------------------------|------------------------------|
| 1                   | the second millenium AD | drugie tysiąclecie n.e.      |
| 19                  | the 20th century AD     | XX wiek n.e.                 |
| 199                 | the 1990's              | lata 90. XX wieku n.e.       |
| 1992                | year 1992               | rok 1992                     |
| 1992-06-27          | 27 June 1992            | 27 czerwca 1992 roku         |
| 1992-06-27T18:04    | 1992-06-27 18:04        | 27 czerwca 1992 roku g 18:04 |
| 1992-06-27T18:04:56 | 1992-06-27 18:04:56     | 1992-06-27 18:04:56          |
| 1992-06-27TMO       | morning of 27 June 1992 | ranek 27 czerwca 1992 roku   |
| 1992-W04            | the fourth week of 1992 | czwarty tydzień 1992 roku    |
| 1992-SU             | summer of 1992          | lato 1992 roku               |
| 1992-H1             | 1st half of 1992        | pierwsza połowa 1992 roku    |
| 1992-Q3             | 3rd quarter of 1992     | trzeci kwartał 1992 roku     |
| BC0346              | year 346 BC             | 346 rok p.n.e.               |
| MA6                 | 6 million years ago     | 6 milionów lat temu          |
| P2Y                 | 2 years                 | 2 lata                       |
| P2Y3M               | 2 years and 3 months    | 2 lata i 3 miesiące          |
| P3W                 | 3 weeks                 | 3 tygodnie                   |
| P9D                 | 9 days                  | 9 dni                        |
| P2DT6H              | 2 days and 6 hours      | 2 dni i 6 godzin             |
| PT8H                | 8 hours                 | 8 godzin                     |
| PT12M               | 12 minutes              | 12 minut                     |
| PT8H12M             | 8 hours and 12 minutes  | 8 godzin i 12 minut          |
| PT3.5H              | 3.5 hours               | 3 i pół godziny              |
| P2DE                | 2 decades               | dwie dekady                  |
| P3CE                | 3 centuries             | trzy wieki                   |

Both standards (TIMEX2 and TIMEX3) expand the ISO-8601 standard, such that they have the ability to save a normalized form of vague expressions, using

an X sign to identify those elements of notation which are unknown. For example, a sunny day in June can be written as XXXX-06-XX. This standard was also expanded for new time tokens, such as the *dekada* (*DEcade*), *wiek* (*CEntury*) and *tyściecie* (*ML — millennium*). Normalization in the TimeML standard involves the determination of the global semantics for a temporal expression. There is no indirect form of notation of the local semantics. The introduction of the intermediate stage of determining the global semantics is reasonable from the normalization point of view. It can be seen in systems that recognize temporal expressions in the English language (e.g. HeidelTime,<sup>5</sup> (Strötgen & Gertz, 2013)), which often use their own intermediate standard of normalization. In this article, the LTIMEX standard proposed by Mazur (2012) is adapted. It can be used to determine the local semantics of temporal expressions.

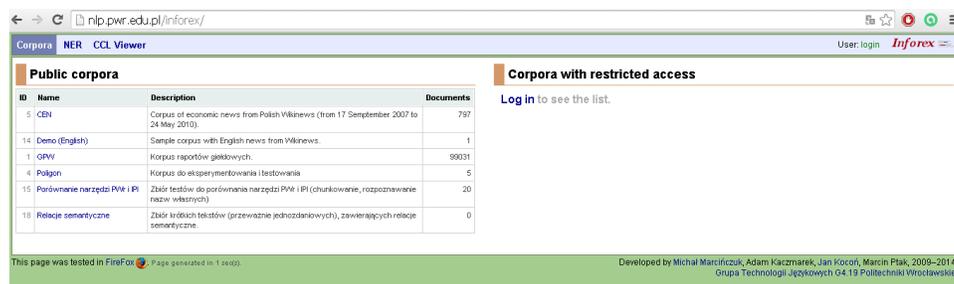
## 6. Annotation Process

### 6.1. Inforex and KPWr

Inforex (Marcińczuk, Kocoń, & Broda, 2012) is a web-based system designed for managing and annotating text corpora at the semantic level including annotation of Named Entities (NE), anaphora, Word Sense Disambiguation (WSD) and relations between named entities. The system also supports manual text clean-up and automatic text pre-processing including text segmentation, morphosyntactic analysis and word selection for WSD annotation.

KPWr — *Korpus Języka Polskiego Politechniki Wrocławskiej (Polish Corpus of Wrocław University of Technology)* (Broda, Marcińczuk, Maziarz, Radziszewski, & Wardyński, 2012; Radziszewski, Maziarz, & Wiczorek, 2012) is a corpus of written and spoken documents available on the Creative Commons license. Documents are divided into 14 categories and annotated at the level of shallow syntactic chunks, selected predicate-argument relations, named entities, relations between named entities, anaphora relations, word senses, temporal expressions and events.

### 6.2. Annotating Using Inforex



**Figure 1:** Open Inforex site — <http://nlp.pwr.edu.pl/inforex/> — and log in to the system.

<sup>5</sup><https://code.google.com/p/heideltime/>



Figure 2: Select the KPWr corpus.

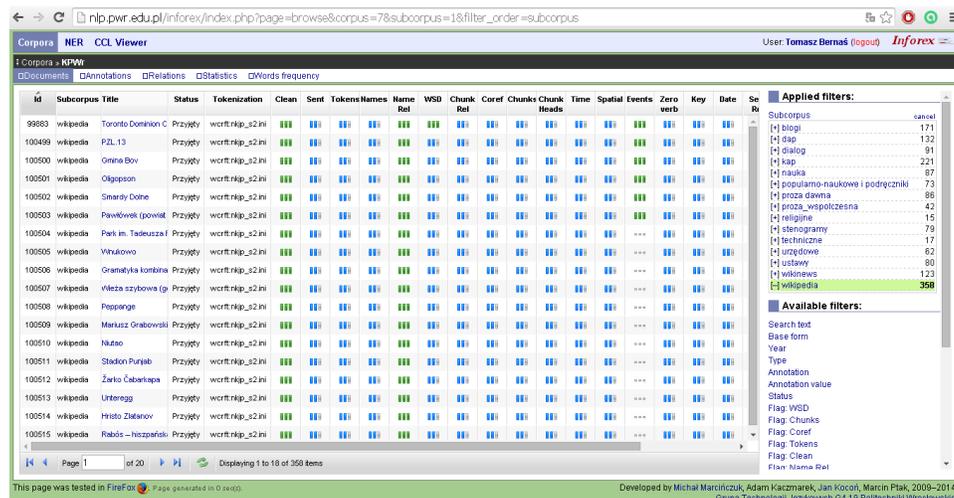
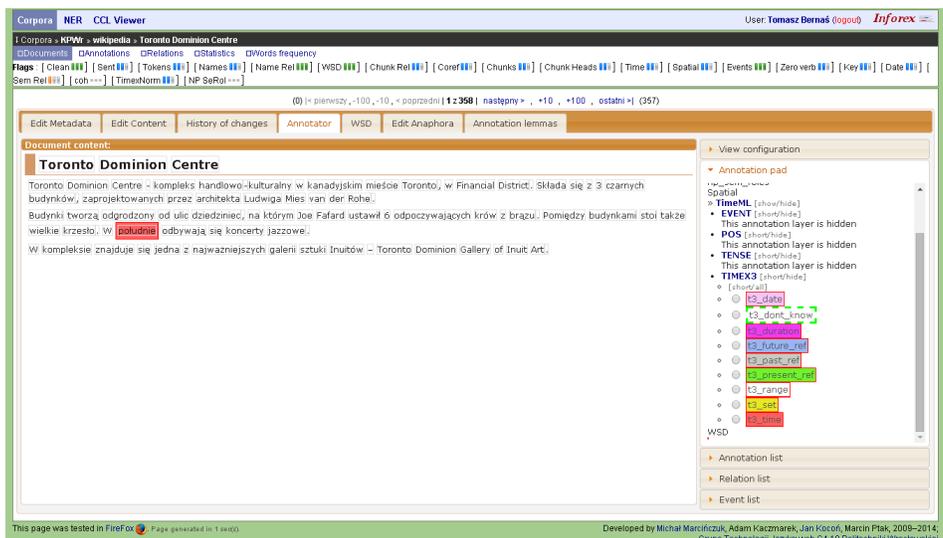


Figure 3: Select *Document* bookmark and then choose the desired subcorpus (e.g. Available filters -> Subcorpus -> wikipedia).



**Figure 4:** Go to the *Annotator* bookmark and select one of the documents and in the left menu select and set the display of TIMEX3 tokens (View configuration -> click on drop-down menu TimeML -> select *right* for TIMEX3 -> press the Apply button at the bottom of the menu).



**Figure 5:** Highlight the desired string using the cursor and select proper type of the temporal expression (*t3\_date*, *t3\_time*, *t3\_duration*, *t3\_set*) in the *Annotation pad* (in the right menu).

The screenshot shows the Inforex CCL Viewer interface. The main content area displays the text of a Wikipedia article about the Toronto Dominion Centre. The right sidebar, titled 'View configuration', shows a list of annotation layers. The layers 't3\_date', 't3\_time', 't3\_duration', and 't3\_set' are checked, indicating they are active for the current view. The 'Apply' button at the bottom of the sidebar is highlighted.

**Figure 6:** Go to the *Annotation lemmas* bookmark and select *t3\_date*, *t3\_time*, *t3\_duration* and *t3\_set* in the right menu and then press the *Apply* button at the bottom of the menu.

The screenshot shows the Inforex CCL Viewer interface. The main content area displays the text of a Wikipedia article about the Toronto Dominion Centre. A search box in the document content area shows the normalized temporal expression 'południe'. The right sidebar, titled 'View configuration', shows a list of annotation layers. The layers 't3\_date', 't3\_time', 't3\_duration', and 't3\_set' are checked, indicating they are active for the current view. The 'Apply' button at the bottom of the sidebar is highlighted.

**Figure 7:** Enter, in the proper spaces, normalized temporal expressions.



**Figure 8:** Use flags at the top of the page to determine the status of the process of the document annotation.

### 6.3. Annotation Statistics

Currently all documents from KPWr (1637) were annotated with temporal expressions. We compared statistics after the annotation process with other publicly available corpora annotated using TIMEX3 annotation format. The evaluation data was prepared to Semantic Evaluation 2013 Workshop<sup>6</sup> evaluation tasks. One of the tasks proposed was TempEval-3.<sup>7</sup> The aim of that task was to advance research on temporal information processing which could increase the quality of applications like question answering, textual entailment, text summarization, etc. It was a three-part task covering event, temporal expression and temporal relation extraction, using the complete TimeML specification for English and Spanish. The result of the evaluation is presented in (UzZaman, Llorens, Allen, Derczynski, Verhagen, & Pustejovsky, 2012). The Table 5 shows the result of analysis of the dataset provided by SemEval 2013 organizers, compared to KPWr data. The full description of these corpora is available in article (UzZaman et al., 2012).

The results of corpora analysis shows that the distribution of temporal expressions in corpora for both English and Spanish is similar to KPWr. It is worth to mention, that T3Silver corpus is a machine-annotated and automatically merged dataset based on outputs of multiple systems (UzZaman et al., 2012), whereas each document in KPWr was annotated manually with temporal expressions.

### 6.4. Agreement

The inter-annotator agreement was measured on randomly selected 100 documents from the Corpus of Wrocław University of Technology called KPWr. We used the positive specific agreement (Hripcsak & Rothschild, 2005) as it was measured for T3Platinum corpus (UzZaman et al., 2012) and two domain experts to annotate the subset of 100 documents from KPWr. We calculate the value of positive specific agreement (PSA) for each category. The results are presented in Table 6.

According to (UzZaman et al., 2012) the best quality of data was achieved for TempEval-3 platinum corpus (T3Platinum) and it was annotated and reviewed by the organizers. Every file was annotated independently by at least two expert annotators. The result of overall T3Platinum inter-annotator positive specific

<sup>6</sup><http://www.cs.york.ac.uk/semeval-2013/>

<sup>7</sup><http://www.cs.york.ac.uk/semeval-2013/task1/>

**Table 5:** The result of analysis of the dataset provided by SemEval 2013 organizers, compared to KPWr data. ALL is the sum of DATEs, TIMEs, DURATIONs and SETs in each corpus. SUM is the number of DATEs, TIMEs, DURATIONs and SETs in all corpora except KPWr. % means what part of all temporal expressions covers particular type. Languages: **EN**glish, **SP**anish, **PL** — Polish.

|                  | Aquaint | TimeBank | T3Platinum | T3Silver | T3Spanish | SUM   | KPWr  |
|------------------|---------|----------|------------|----------|-----------|-------|-------|
| <b>Language</b>  | EN      | EN       | EN         | EN       | SP        | EN&SP | PL    |
| <b>Documents</b> | 73      | 183      | 21         | 2452     | 178       | 2907  | 1637  |
| <b>DATE</b>      | 502     | 1177     | 116        | 11133    | 924       | 13852 | 4391  |
| <b>TIME</b>      | 68      | 44       | 4          | 2644     | 57        | 2817  | 928   |
| <b>DURATION</b>  | 68      | 189      | 34         | 1346     | 251       | 1888  | 653   |
| <b>SET</b>       | 14      | 16       | 4          | 68       | 37        | 139   | 144   |
| <b>ALL</b>       | 652     | 1426     | 158        | 15191    | 1269      | 18696 | 6116  |
| <b>DATE[%]</b>   | 76,99   | 82,54    | 73,42      | 73,29    | 72,81     | 74,09 | 71,80 |
| <b>TIME[%]</b>   | 10,43   | 3,09     | 2,53       | 17,41    | 4,49      | 15,07 | 15,17 |
| <b>DUR.[%]</b>   | 10,43   | 13,25    | 21,52      | 8,86     | 19,78     | 10,10 | 10,68 |
| <b>SET[%]</b>    | 2,15    | 1,12     | 2,53       | 0,45     | 2,92      | 0,74  | 2,35  |

agreement (PSA) at the level of annotating of temporal expressions with types was 0.88. In our case for 100 randomly selected documents the PSA value achieved was 86.25 (annotating using PLIMEX 1.0 specification). We manage to achieve positive specific agreement at the level of at least 0.9 in PLIMEX 2.0.

## 7. Conclusions and Further Work

In this paper we described the PLIMEX 1.0 specification to annotate Polish text documents with the temporal expressions. We performed several comparison methods to other corpora annotated using similar temporal information description methods. We evaluate the quality of the specification by annotating of the whole KPWr corpus (1637 documents) and we calculated the positive specific agreement on the randomly selected subset of 100 documents annotated by two domain experts, achieving very promising result. The next step is to construct the machine learning system which will use KPWr as the evaluation data (a resource of both training and test dataset). At the same time we plan to construct a rule-based system to normalize the recognized temporal expressions and evaluate the result using normalized values added by domain experts to KPWr.

**Table 6:** The value of positive specific agreement (PSA) calculated on the subset of 100 documents from KPWr, annotated independently by two domain experts using PLIMEX 1.0 guidelines. *1 and 2* means all annotations in which annotators 1 and 2 agreed. *Only 1* is the number of annotations made only by annotator 1 and *only 2* — the number of annotations made only by annotator 2.

| Type     | 1 and 2 | only 1 | only 2 | PSA [%] |
|----------|---------|--------|--------|---------|
| DATE     | 182     | 12     | 22     | 91,46   |
| TIME     | 28      | 13     | 8      | 72,73   |
| DURATION | 13      | 3      | 4      | 78,79   |
| SET      | 6       | 2      | 9      | 52,17   |
| ALL      | 229     | 30     | 43     | 86,25   |

## References

- Allen, J. (1995). *Natural language understanding* (2nd ed.). Redwood City, CA: Benjamin Cummings Publishing Co., Inc.
- Andersen, P. M., Hayes, P. J., Huettner, A. K., Schmandt, L. M., Nirenburg, I. B., & Weinstein, S. P. (1992). Automatic extraction of facts from press releases to generate news stories. In *Processing of the Third Conference on Applied Natural Language Processing* (pp. 170–177). Stroudsburg, PA: Association for Computational Linguistics. <http://doi.org/10.3115/974499.974531>
- Van Benthem, J. F. A. K. (1983). *The logic of time: A model-theoretic investigation into the varieties of temporal ontology and temporal discourse*. Dordrecht: Springer Netherlands. <http://link.springer.com/10.1007/978-94-010-9868-7>
- Broda, B., Marcińczuk, M., Maziarz, M., Radziszewski, A., & Wardyński, A. (2012, May). 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, Turkey: European Language Resources Association (ELRA).
- Busemann, S., Declerck, T., Diagne, A. K., Dini, L., Klein, J., & Schmeier, S. (1997). Natural language dialogue service for appointment scheduling agents. In *Proceedings of the Fifth Conference on Applied Natural Language Processing* (pp. 25–32). Stroudsburg, PA: Association for Computational Linguistics. <http://doi.org/10.3115/974557.974563>
- Daniel, N., Radev, D., & Allison, T. (2003). Sub-event based multi-document summarization. In *Proceedings of the HLT-NAACL 03 on text summarization workshop* (Vol. 5, pp. 9–16). Stroudsburg, PA: Association for Computational Linguistics. <http://doi.org/10.3115/1119467.1119469>
- Ferro, L. (2001). Instruction manual for the annotation of temporal expressions.
- Filatova, E. & Hovy, E. (2001). Assigning time-stamps to event-clauses. In *Proceedings of the workshop on temporal and spatial information processing* (Vol. 13, pp. 13:1–13:8). Stroudsburg, PA: Association for Computational Linguistics. <http://doi.org/10.3115/1118238.1118250>
- Han, B., Gates, D., & Levin, L. (2006). Understanding temporal expressions in emails. In *Proceedings of the Main Conference on Human Language Technology Conference of the*

- North American Chapter of The Association of Computational Linguistics (pp. 136–143). Stroudsburg, PA: Association for Computational Linguistics. <http://doi.org/10.3115/1220835.1220853>
- Hripcsak, G. & Rothschild, A. S. (2005). 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>
- Laskowski, R. (2003). Wyrażenia przyimkowe o funkcji temporalnej w języku polskim. *Studia Slavica Oldenburgensia*, 11, 193–226.
- Laskowski, R. (2005). Temporalne frazy przyimkowe o funkcji prospektywnej i retrospektywnej. In M. Grochowski (Ed.), *Przystówki i przyimki: Studia ze składni i semantyki języka polskiego* (pp. 209–225). Toruń: Wydawnictwo Uniwersytetu Mikołaja Kopernika.
- Llorens, H., Saquete, E., & Navarro-Colorado, B. (2010). TimeML events recognition and classification: Learning CRF models with semantic roles. In *Proceedings of the 23rd International Conference on Computational Linguistics* (pp. 725–733). Stroudsburg, PA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=1873781.1873863>
- Mani, I. & Wilson, G. (2000). Robust temporal processing of news. In *Proceedings of the 38th annual meeting on Association for Computational Linguistics* (pp. 69–76). Stroudsburg, PA: Association for Computational Linguistics. <http://doi.org/10.3115/1075218.1075228>
- Marcińczuk, M. & Kocoń, J. (2013, August). Recognition of named entities boundaries in Polish texts. In *Proceedings of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing* (pp. 94–99). Sofia: Association for Computational Linguistics. <http://www.aclweb.org/anthology/W13-2414>
- Marcińczuk, M., Kocoń, J., & Broda, B. (2012, May). 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).
- Marcińczuk, M., Kocoń, J., & Janicki, M. (2013). Liner2 — a customizable framework for proper names recognition for Polish. In R. Bembeni, L. Skonieczny, H. Rybiński, M. Kryszkiewicz, & M. Niezgodka (Eds.), *Intelligent tools for building a scientific information platform* (pp. 231–253). Berlin: Springer. (*Studies in Computational Intelligence*, 467). [http://link.springer.com/10.1007/978-3-642-35647-6\\_17](http://link.springer.com/10.1007/978-3-642-35647-6_17)
- Mazur, P. (2012). Broad-coverage rule-based processing of temporal expressions (Unpublished doctoral dissertation). Politechnika Wroclawska, Wrocław.
- Mizobuchi, S., Sumitomo, T., Fuketa, M., & Aoe, J.-I. (1998, October). A method for understanding time expressions. In *Systems, man, and cybernetics, 1998. 1998 IEEE International Conference on* (Vol. 2, pp. 1151–1155). <http://doi.org/10.1109/ICSMC.1998.727858>
- Negri, M. & Marseglia, L. (2005). *Recognition and normalization of time expressions: ITC-irst at TERN 2004*.
- Niemi, J. & Koskenniemi, K. (2007). Representing calendar expressions with finite-state transducers that bracket periods of time on a hierarchical timeline. In J. Nivre, H.-J. Kaalep, K. Muischnek, & M. Koit (Eds.), *Proceedings of the 16th Nordic Conference of Computational Linguistics NODALIDA-2007* (pp. 355–362). Tartu, Estonia.
- Pustejovsky, J., Ingria, B., Saurí, R., Castano, J., Littman, J., Gaizauskas, R., Setzer, A., Katz, G., & Mani, I. (2005). The specification language TimeML. In I. Mani (Ed.), *The language of time: A reader* (pp. 545–557). Oxford: Oxford University Press.

- Pustejovsky, J., Knippen, R., Littman, J., & Saurí, R. (2005). Temporal and event information in natural language text. *Language Resources and Evaluation*, 39(2–3), 123–164. <http://doi.org/10.1007/s10579-005-7882-7>
- Radziszewski, A., Maziarz, M., & Wieczorek, J. (2012). Shallow syntactic annotation in the Corpus of Wrocław University of Technology. *Cognitive Studies | Études cognitives*, 12, 129–147.
- Saquete, E., Muñoz, R., & Martínez-Barco, P. (2003). TERSEO: Temporal Expression Resolution System Applied to Event Ordering. In V. Matoušek & P. Mautner (Eds.), *Text, speech and dialogue* (pp. 220–228). Berlin: Springer. (Lecture Notes in Computer Science, 2807). [http://dx.doi.org/10.1007/978-3-540-39398-6\\_31](http://dx.doi.org/10.1007/978-3-540-39398-6_31)
- Saurí, R., Littman, J., Knippen, B., Gaizauskas, R., Setzer, A., & Pustejovsky, J. (2006). *TimeML annotation guidelines, version 1.2.1*.
- Schilder, F. (2004). Extracting meaning from temporal nouns and temporal prepositions. *ACM Transactions on Asian Language Information Processing (TALIP)*, 3(1), 33–50. <http://doi.org/10.1145/1017068.1017071>
- Schilder, F. & Habel, C. (2001). From temporal expressions to temporal information: Semantic tagging of news messages. In *Proceedings of the ACL-2001 Workshop on Temporal and Spatial Information Processing, ACL-2001. Toulouse* (Vol. 13, pp. 65–72). Stroudsburg, PA: Association for Computational Linguistics. <http://doi.org/10.3115/1118238.1118247>
- Smith, C. (2009). Temporal structures in discourse. In R. P. Meier, H. Aristar-Dry, & E. Destruel (Eds.), *Text, time, and context* (pp. 285–302). Dordrecht: Springer Netherlands. (*Studies in Linguistics and Philosophy*, 87). [http://link.springer.com/10.1007/978-90-481-2617-0\\_12](http://link.springer.com/10.1007/978-90-481-2617-0_12)
- Strötgen, J., Zell, J., & Gertz, M. (2013, June). HeidelTime: Tuning English and developing Spanish resources for TempEval-3. In *Second Joint Conference on Lexical and Computational Semantics (SEM)* (Vol. 2: *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pp. 15–19). Atlanta, Georgia: Association for Computational Linguistics. <http://www.aclweb.org/anthology/S13-2003>
- Strötgen, J. & Gertz, M. (2013). Multilingual and cross-domain temporal tagging. *Language Resources and Evaluation*, 47(2), 269–298. <http://doi.org/10.1007/s10579-012-9179-y>
- UzZaman, N., Llorens, H., Allen, J. F., Derczynski, L., Verhagen, M., & Pustejovsky, J. (2012). TempEval-3: Evaluating events, time expressions, and temporal relations. *CoRR*, *abs/1206.5333*. <http://arxiv.org/abs/1206.5333>
- Vicente-Díez, M. T., Samy, D., & Martínez, P. (2008). An empirical approach to a preliminary successful identification and resolution of temporal expressions in Spanish news corpora. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC)* (p. 2153–2158).

---

### 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: JK and MM are responsible for the ideas in the research: JK. TB and MO created annotation guidelines and were responsible for inter-annotator agreement increasing. JK performed data analyses: 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