# **Event extraction with use of Frames**

Dutkiewicz Jakub, Falkowski Maciej, Nowak Maciej, Jędrzejek Czesław



Event extraction is the induction of n-ary relations from natural language texts. Some event extraction systems focus on the text coverage – the more events extracted the better. There exist ACE and ERE annotation standards specified by Entities, Relations and Events. Many different strategies for finding event word triggers and slot filling have been applied. All of the approaches accept events which are semantically nonsensical. The famous Chomsky's problem of "the colorless green ideas which sleep furiously" remains untouched. Here, we present a specific event extraction methodology which partially solves this problem. We use sets of acceptable semantic classes for event arguments, which are stored within frames. A frame is defined as a pair (R, S), where R is a semantic type of frame, and S is a set of its slots. Every slot in S is defined as a pair (Ri, T), where Ri is the relation, and T is a list of allowed semantic types for this Slot. Frame languages are fundamental for annotation event extraction and querying.

### Ontology

Extraction vocabulary is stored within the ontology. Each class in the taxonomy has assigned its lexicalizations. If word corresponds to certain class, base form of this word should appear within lexicallizations of that class. Ontology used for experimets contains 222 clasess and over 1000 lexicalizations.

#### EnergyResource PhysicalObject OtherPhysicalOb Number AstronomicalObj Hiding BreakingIn Singing Robbing Destroying Punishina Rescheduling Purchasing Event Searching Delivering Traveling Recommendating

### **Data structures** for the following sentence:

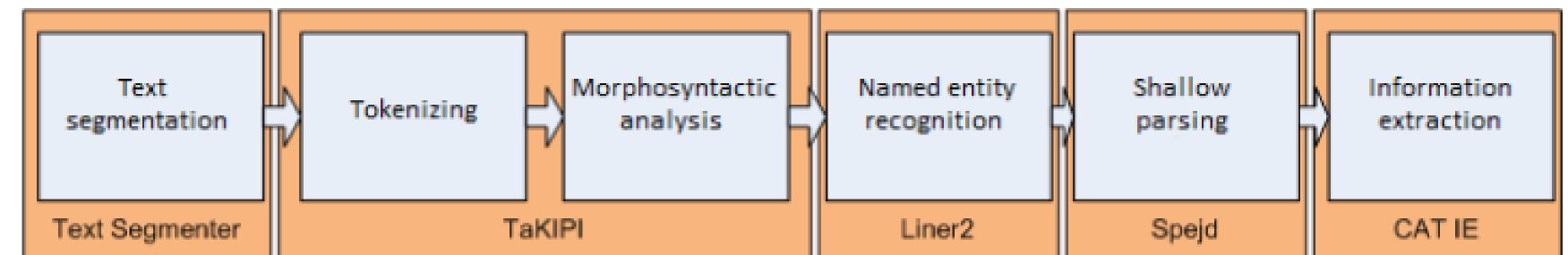
Last year Jan kowalski bought a new Fiat for a song.

W zeszłym roku Jan Kowalski kupił nowego Fiata za bezcen.

Frame		Template		Instance	
Type: Purchase		Anchor: Active verb		bought	
Slots:		Slots:		Slots:	
Relation (attribute)	Allowed type	Case	prepositions	Value	Type
Buyer	#Person,	nominative	_	Jan Kowalski	#Man
	#Organization				
Seller	#Person,	dative	_		_
	#Organization				
Bought object	#Thing	accusative	_	new Fiat	#Car
Price	#Money	accusative	"za" ("for")	for a song	#Money
Place	#Place	locative	"w" ("in")		_
Date	#Date	locative	"w" ("in")	last year	#Time

### Processing pipeline

The Extractor processes data in several steps. The extraction itself is the final step and is performed on the structures representing a parsed sentence. Figure describes consecutive stages of the analysis (inner boxes) and the names of the tools that perform those stages for the Polish language (outer boxes).



## **Extraction algorithm**

- 1. For each sentence:
- i) If verb corresponds to event type specified within the frame try to extract an event from this sentence.
- ii)For each noun phrase in the sentence:
- I. If head of the noun phrase does not correspond grammatically to any of the slots within the frame, skip this phrase continue to next noun phrase.
- II.If head of noun phrase appears in the ontology and corresponds to slot, which is gramatically acceptable, attach phrase to extracted event.
- III.If head of noun phrase appears in the ontology and does not correspond to slot, which is gramatically acceptable, terminate extraction for this verb.

IV.If head of noun phrase does not appear in the ontology and corresponds to slot, which is gramatically acceptable, attach phrase to extracted event.

### **Evaluation**

To evaluate the accuracy of described method we have downloaded data from the National Corpus of Polish. We have randomly chosen 1000 sentences with the word "kill" and 1000 sentences with the word "purchase". We have manually annotated each of the 2000 sentences. Exactly the same set of sentences was used to perform the extraction with the IE system.

Number of errors	Event type	Precision	Recall	Precision Precisioner
0	#Kill			51%
<2	#Kill	28%	84%	85%
<3	#Kill	33%	91%	100%
	#Purchase			33%
	#Purchase			69%
<3	#Purchase	47%	90%	90%

TP = number of correctly extracted events

T = number of annotated events

Precision = TP / (TP + FP)

FP = number of incorrectly extracted events

Number of errors = number of incorrectly assigned thematic roles Recall = TP / T

FP<sub>er</sub>= number of incorrectly extracted events, if type of theevent was recognized correctly