

# Evolution of Event Designation in Media: Preliminary Study

Xavier Tannier, Véronique Moriceau, Béatrice Arnulphy and Ruixin He  
LIMSI-CNRS, Univ. Paris-Sud, Orsay, France  
firstname.surname@limsi.fr



## Context: Evolution of Event Designation

### Events are often important in information extraction

But they are much less studied than named entities  
or by the mean of relation extraction

### Event designation change over time

- "September-11" or "Arab Spring" have not been so-called from the first day
- This evolution is important in IR and IE  
Querying "Arab Spring" will not lead to news articles describing the actual events

## Objectives

### Testing the following hypotheses:

- 1 Evolution from verbal to nominal designation  
When an event first occurs, media relate it in a very descriptive way (using verbal designations) whereas after some time, they use nominal designations instead.
  - 2 The number of different designations decreases over time
  - 3 Designations get shorter over time
- All these are intuitive phenomena that need to be confirmed by the observation

## Corpus



### Initial Corpus

- 1 million newswire texts in French from Agence France Presse (AFP)
- Period: 2004-2011
- Covers all areas of the news.
- Stemmed and indexed by search engine Lucene

### 7 Sub-Corpora for our Study

- 7 topics lead to 7 sub-corpora by querying Lucene index:
- Greek economic crisis
  - Tunisian and Egyptian revolutions
  - Wikileaks affair (public release of many diplomatic and secret documents)
  - Leatitia's case (the murder of a young woman which evolved into a strong judicial polemic in France)
  - H1N1 influenza (responsible for a pandemic alert in the world)
  - Icelandic ash cloud (that forced European authorities to cancel flights)

## Corpus Annotation

- In each sub-corpus, topic-related passages have been annotated  
Using the Callisto annotation tool (Day et al., 2004)
- Verbal and nominal designations have been marked up
- 2064 annotations  
that can be distributed to interested people

### Examples:

- Topic "Greek crisis":

European leaders are preparing to meet this Thursday for an emergency summit to redefine the rules in the euro zone after the **<EVT\_NOUN>Greek crisis</EVT\_NOUN>**.

**<EVT\_VERB>Drowning in debt, Greece needed help from its partners of the euro area</EVT\_VERB>**.

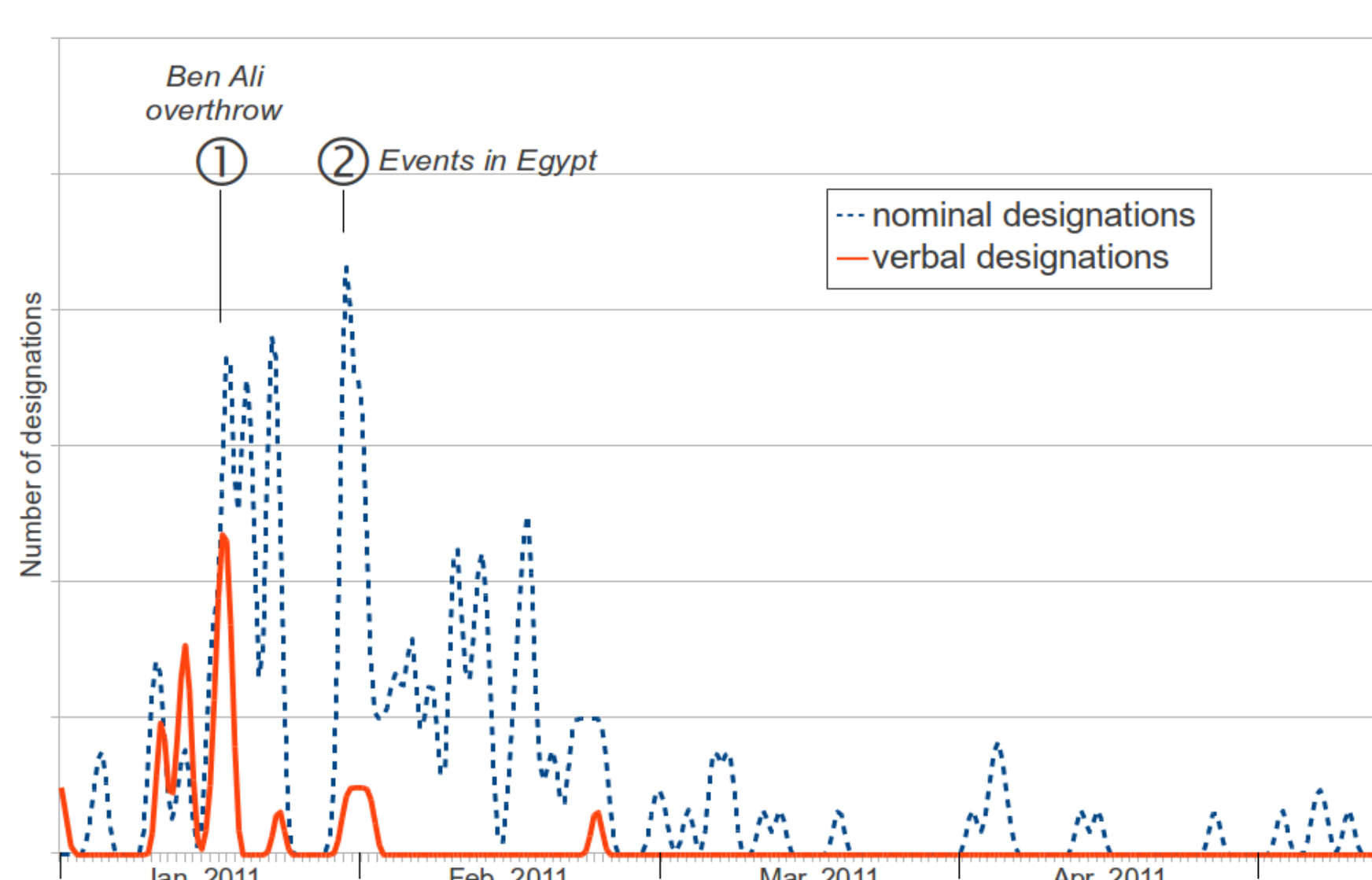
- Topic "Tunisian revolution":

**<EVT\_VERB>Ben Ali left Tunis after a month-long popular protest which Tunisians called the <EVT\_NOUN>Jasmine Revolution</EVT\_NOUN> </EVT\_VERB>**

## Observations

### From Verbal to Nominal Designations

- When an event occurs, the first designations in relation to what happened take the form of a sentence headed by a verb.  
The event must be described and contextualized (location, time, aspect).
- When the event is of a relative importance, it is no longer necessary to use the verbal forms.  
We (media and locutors) use a name for the event: the verbal designations are nominalized.  
This process can happen in one or few days.
- Then the event is no more expressed as the main information, but rather as a contextual information.  
It is already incorporated into the context of other related events: responses to this first event, declarations, consequences, etc.

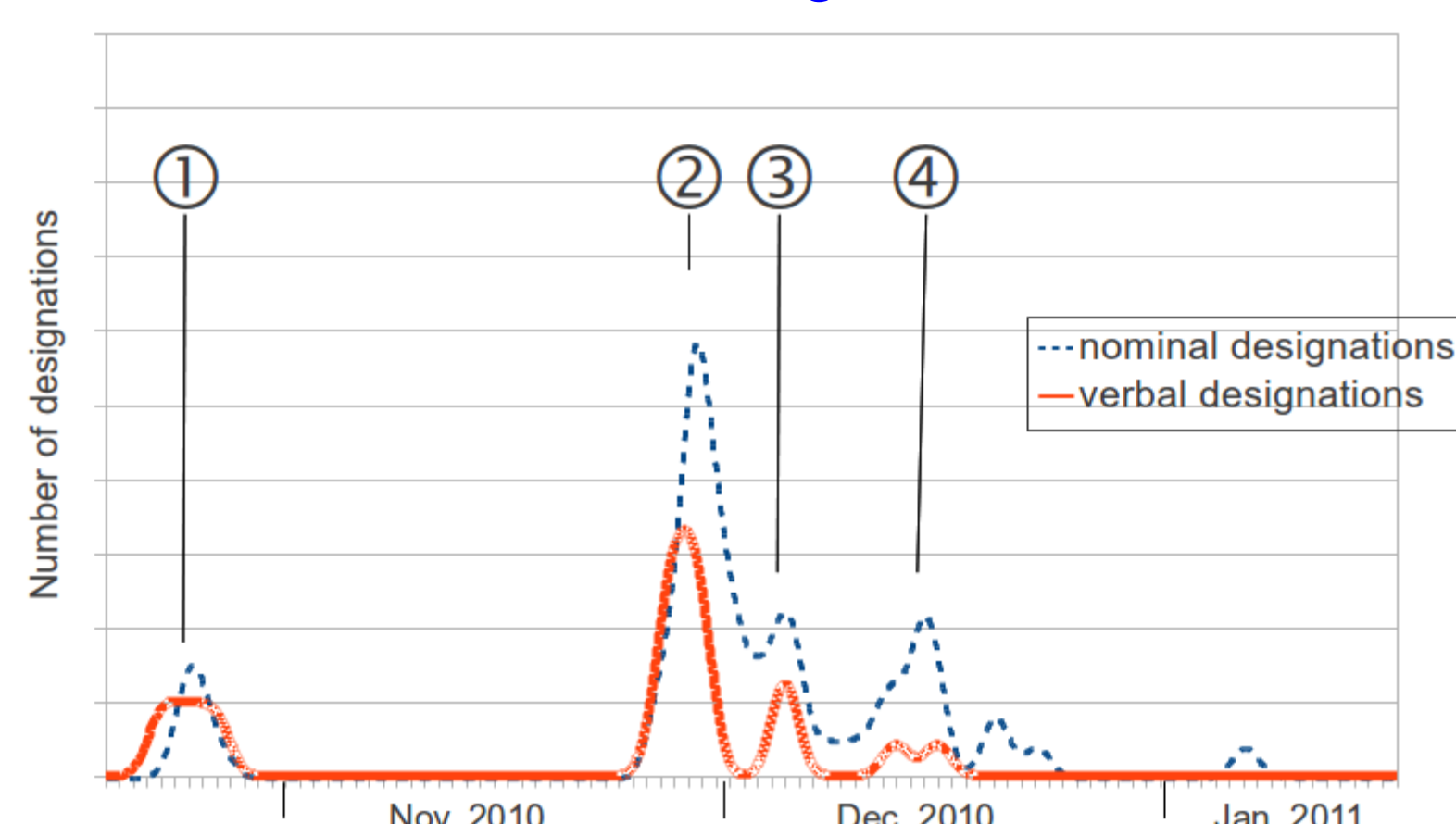


### Arab Spring

- ① President Ben Ali overthrow (Jan. 14th, 2011). This date is preceded by a progression of designations illustrating the worsening of the situation.  
  
Many verbal designations on January 14th were relayed by mostly nominal designations the following days.
- ② revolution in Egypt, where a lot of articles referred to the origins of the "Arab Spring".

### Wikileaks affair

Verbal and nominal designations are simultaneous.



Events were announced and expected before they occurred (Wikileaks announced the leaks a few days in advance).

Designations were already nominalized when the facts happened.

### The Avarice Principle

Nominal description of an event becomes shorter and shorter in time.

### Evolution of designations:

"disappearance of a young woman between Tuesday and Wednesday in Pornic"  
"the disappearance of Laetitia in Pornic"  
"the disappearance of Laetitia"  
"the Laetitia's case", mentioned 11 days after the fact.

### Freezing of denominations:

The ratio between the number of different designations of an event and its total number of designations grows up over time.

Contextual elements and verbal designations come back after a certain period of time if the event is still referred to after several months.

