



Creation, Visualization and Edition of Timelines for Journalistic Use

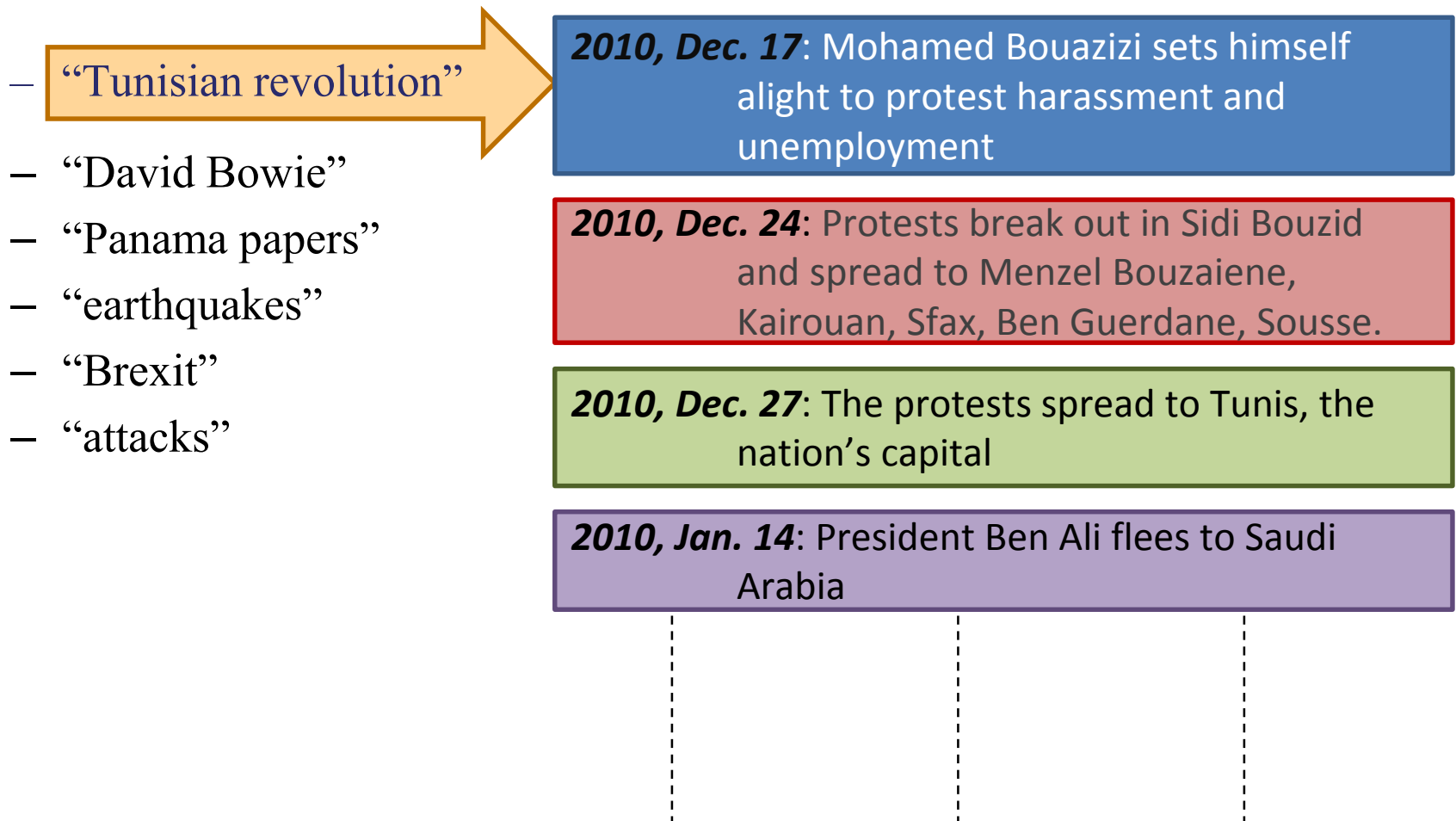
Xavier Tannier, Frédéric Vernier

LIMSI, CNRS, Univ. Paris-Sud, Université Paris-Saclay



Objectives

1. Build automatic timelines from a query





1. Timeline building

Overview

- The final goal: find **important events** from a **query**
- The **intermediate goal**: use temporal information to find **important dates**
 (Our assumption: important dates will lead to important events)
- Our system:
 - Extracts a maximum of **temporal information** from texts
 - Uses this information to extract **salient dates** for a **given query**
 - Associates a **textual description** to each date

2011, Mar. 22

2011, Jun. 12

2010, Dec. 04

Context

- Systems aiming at building timelines mainly see the problem as a traditional (multi-)document **summarization**
 - Use of textual information
 - Use only **little temporal information**
 - Document creation time (DCT)
 - Rarely, absolute dates (day + month + year)

*See Retrospective Event Detection,
New Event Detection (TDT) and others*

Context

- **Absolute dates**

At least 129 people died after a series of violent attacks around Paris, France, on **13 November 2015**.

Only 7% of all dates
(AFP corpus)

- **Relative dates**

The attacks in Paris on the night of **Friday 13 November** left 130 people dead and hundreds wounded.

At least 128 people were killed in shootings and explosions in Paris **late Friday**.

Attacks such as the one in Paris **three days ago** cannot obliterate our desire to understand.

Context: Heideitime

- **Multilingual, domain-sensitive temporal tagger** (Heidelberg University, Strötgen et al.)
- Adapted to French language at LIMSI (Moriceau et Tannier, 2014)

Document Creation Time: 2016-07-10

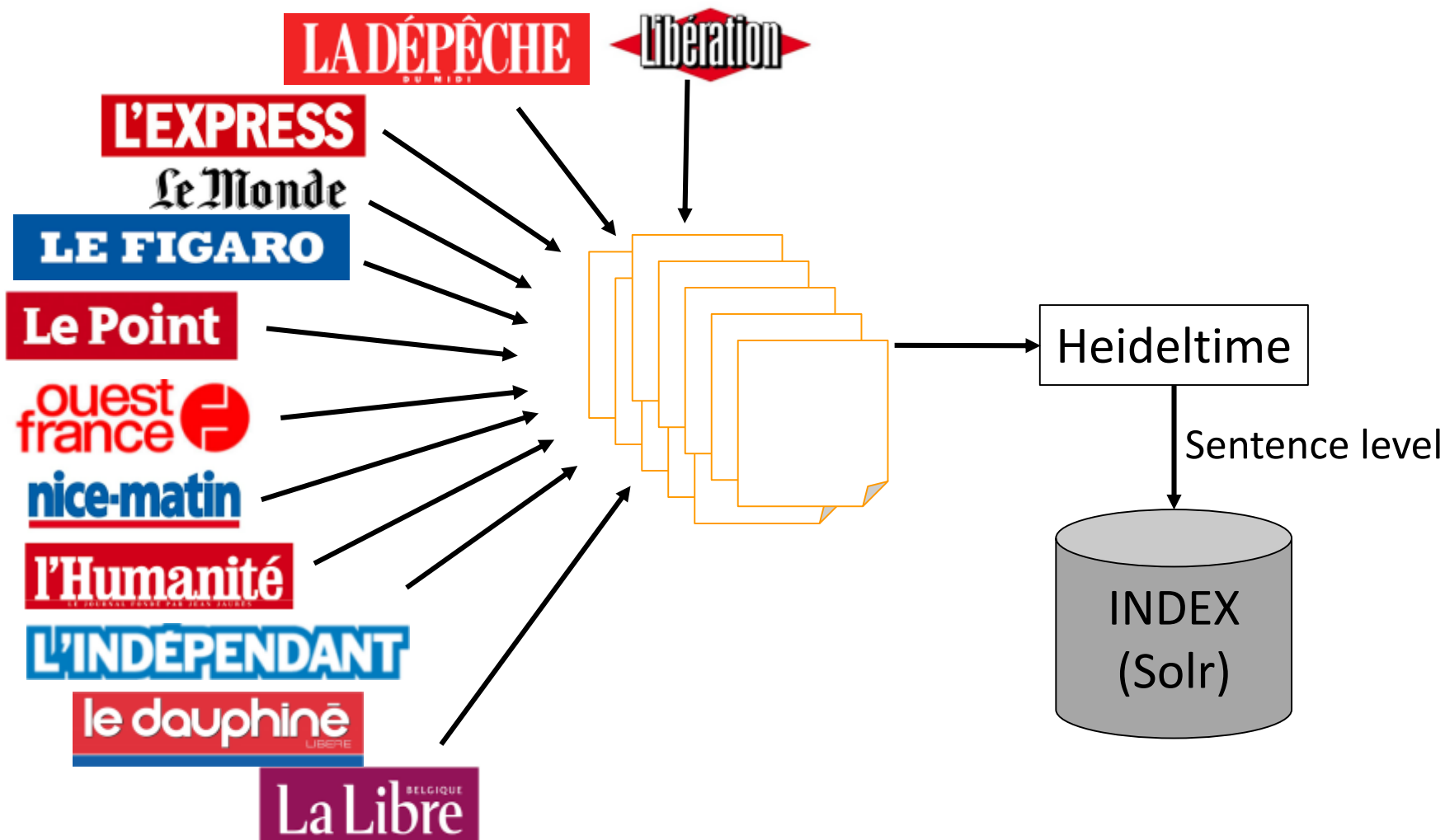
North Korea fired a ballistic missile from a submarine on [Saturday](#) but it appears to have failed soon after launch, South Korea's military said.

type: DATE

value: 2016-07-09

end of [a week](#) of sharply rising tensions only [a day](#) after the U.S. and South Korea pledged to deploy an advanced anti-missile system to counter threats from Pyongyang, and [two days](#) after North Korea warned it was planning its toughest response to what it deemed a "declaration of war" by the United States.

Approach (offline)



Approach (online)

INDEX

Query

No limit for document number
(Solr)

Le président égyptien Hosni Mubarak, qui a démissionné vendredi, et le président Zine El Abidine Ben Ali, qui a quitté le pouvoir le **14 janvier**, ont fait face à des protestations populaires sans précédent.

Zine el-Abidine Ben Ali a annoncé **jeudi soir** qu'il ne se représentera pas en 2014 au poste qu'il occupe

Ben Ali a signé sa démission **vendredi** après une vague de protestations déclenchée par le suicide d'un étudiant de 26 ans que la police avait empêché de vendre des fruits et des légumes pour vivre.

Des manifestants ont également blessés **vendredi**, y compris Chawki Belhoussine El Hadri

Ben Ali a signé sa démission **vendredi** après une vague de protestations déclenchée par le suicide d'un étudiant de 26 ans que la police avait empêché de vendre des fruits et des légumes pour vivre.

La révolte s'est propagée **le 24 décembre** dans le centre-du pays, notamment à Menzel Bouzaiane, où Mohamed Ammari est tué par balle dans la poitrine par la police.

Approach (online)

INDEX

Query

Temporal Clusters

Grouping by normalized dates
(*Heideltime*)
Ordering by date frequency

More important

2011, Jan. 14

Ben Ali a signé sa démission **vendredi** après une vague de protestations déclenchée par le suicide d'un étudiant de 26 ans que la police avait empêché de vendre des fruits et des légumes pour vivre.

Ces révélations ont lieu après la révolte tunisienne qui a mit fin à 23 ans de règne de Ben Ali, qui s'est enfui de Tunisie pour l'Arabie Saoudite **vendredi**.

2010, Dec. 24

La révolte s'est propagée **le 24 décembre** dans le centre-du pays, notamment à Menzel Bouzaiane, où Mohamed Ammari est tué par balle dans la poitrine par la police.

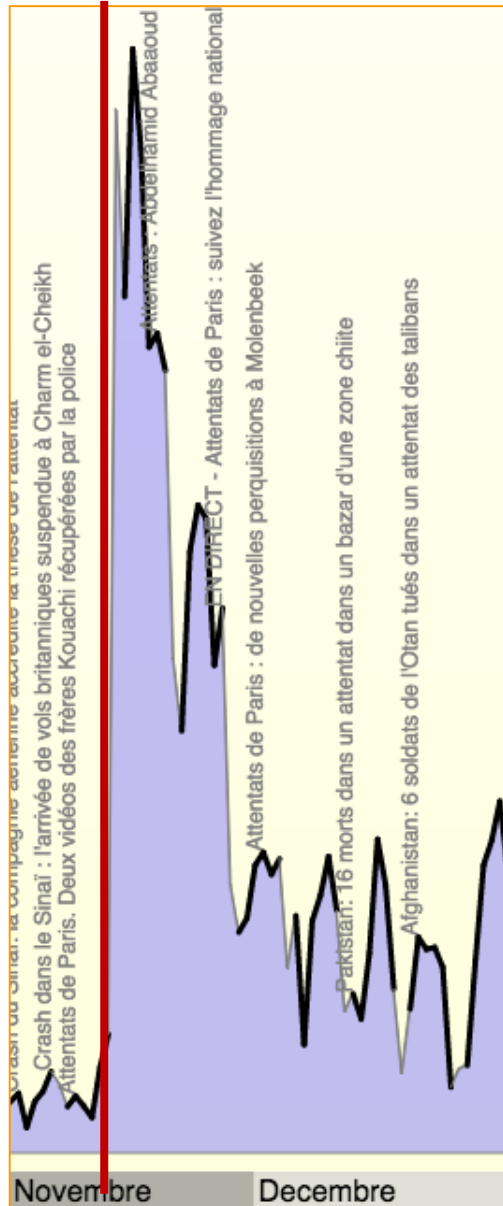
Des manifestants ont également blessés **vendredi**, y compris Chawki Belhoussine El Hadri

2011, Jan. 13

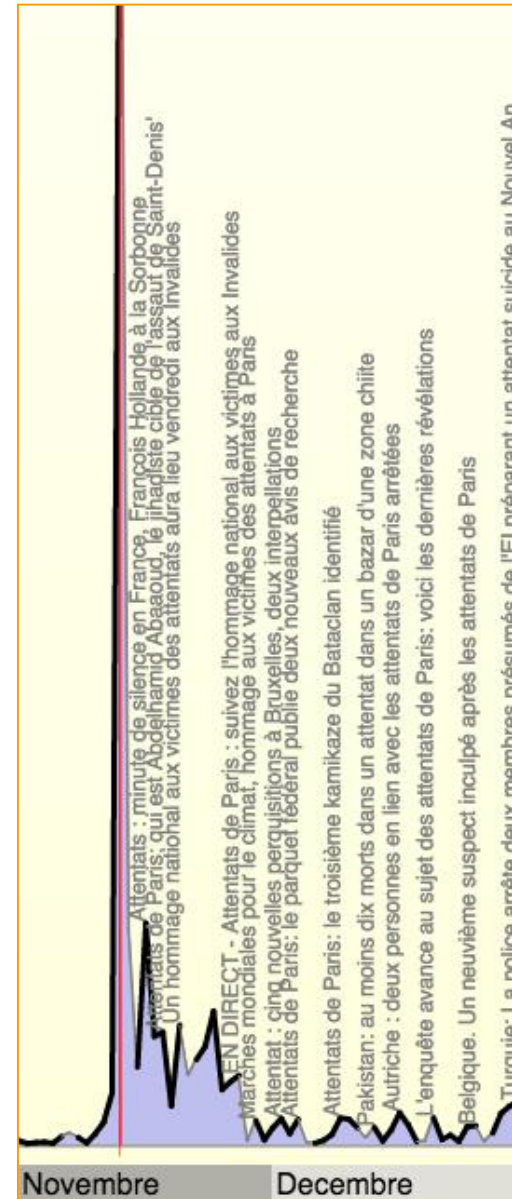
Zine el-Abidine Ben Ali a annoncé **jeudi soir** qu'il ne se représentera pas en 2014 au poste qu'il occupe

Ben Ali a donné **jeudi** l'ordre à la police de ne plus tirer sur les manifestants

Without date normalization
 (document creation times)
 → Media coverage salience



With date normalization
 (Heideltime)
 → Event salience



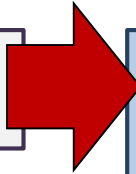
Approach (online)



INDEX

Query

Temporal Clusters



Selection of cluster specific terms
Frequent in the cluster
Not frequent in the other clusters

2011, Jan. 14

Ben Ali a signé sa démission **vendredi** après une vague de protestations déclenchée par le suicide d'un étudiant de 26 ans que la police avait empêché de vendre des fruits et des légumes pour vivre.

Ces révélations ont lieu après la révolte tunisienne qui a mit fin à 23 ans de règne de Ben Ali, qui s'est enfui de Tunisie pour l'Arabie Saoudite **vendredi**.

tf.idf weights for Jan. 14, 2011

2.45 resign
1.89 flight
0.96 Saudi Arabia
...
...

2011, Jan. 13

Zine el-Abidine Ben Ali a annoncé **jeudi soir** qu'il ne se représentera pas en 2014 au poste qu'il occupe

Ben Ali a donné **jeudi** l'ordre à la police de ne plus tirer sur les manifestants

Approach (online)



INDEX

Query

Temporal Clusters

Selection of cluster specific terms

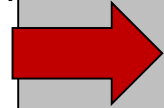
Selection of a representative article

document level

**tf.idf weights for
Jan. 14, 2011**

- 2.45 resign
- 1.89 flight
- 0.96 Saudi Arabia
- ...
- ...

Top 20 terms



Tunisia: Ex-President Ben Ali flees to Saudi Arabia

15 January 2011 | Africa

Tunisian President Zine al-Abidine Ben Ali has fled with his family to Saudi Arabia, after being forced from office after 23 years in power.

The 74-year-old leader conceded power after protests over economic issues escalated into rallies against him.

Prime Minister Mohammed



President Ben Ali resigned after weeks of unrest over corruption and unemployment

Approach (online)



INDEX

Query

Temporal Clusters

Selection of cluster specific terms

Selection of a representative article

More important ←

<p>2011, Jan. 11</p> <p>Ben Ali démissionne et s'ensuit</p>	<p>2010, Dec. 24</p> <p>Un manifestant tunisien tué par balle par la police.</p>	<p>2011, Jan. 13</p> <p>Zine el-Abidine Ben Ali renonce à se représenter</p>
--------------------------------------------------------------------	-----------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------

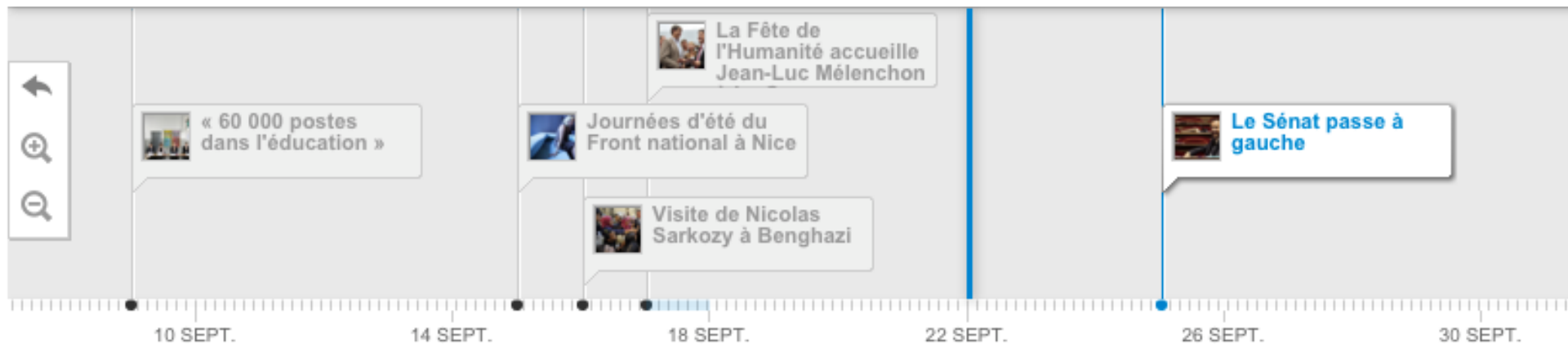
2. Visualization

Objective

- Combine information about:

- Chronology of the events
- Description of the events
- **Importance of the events**

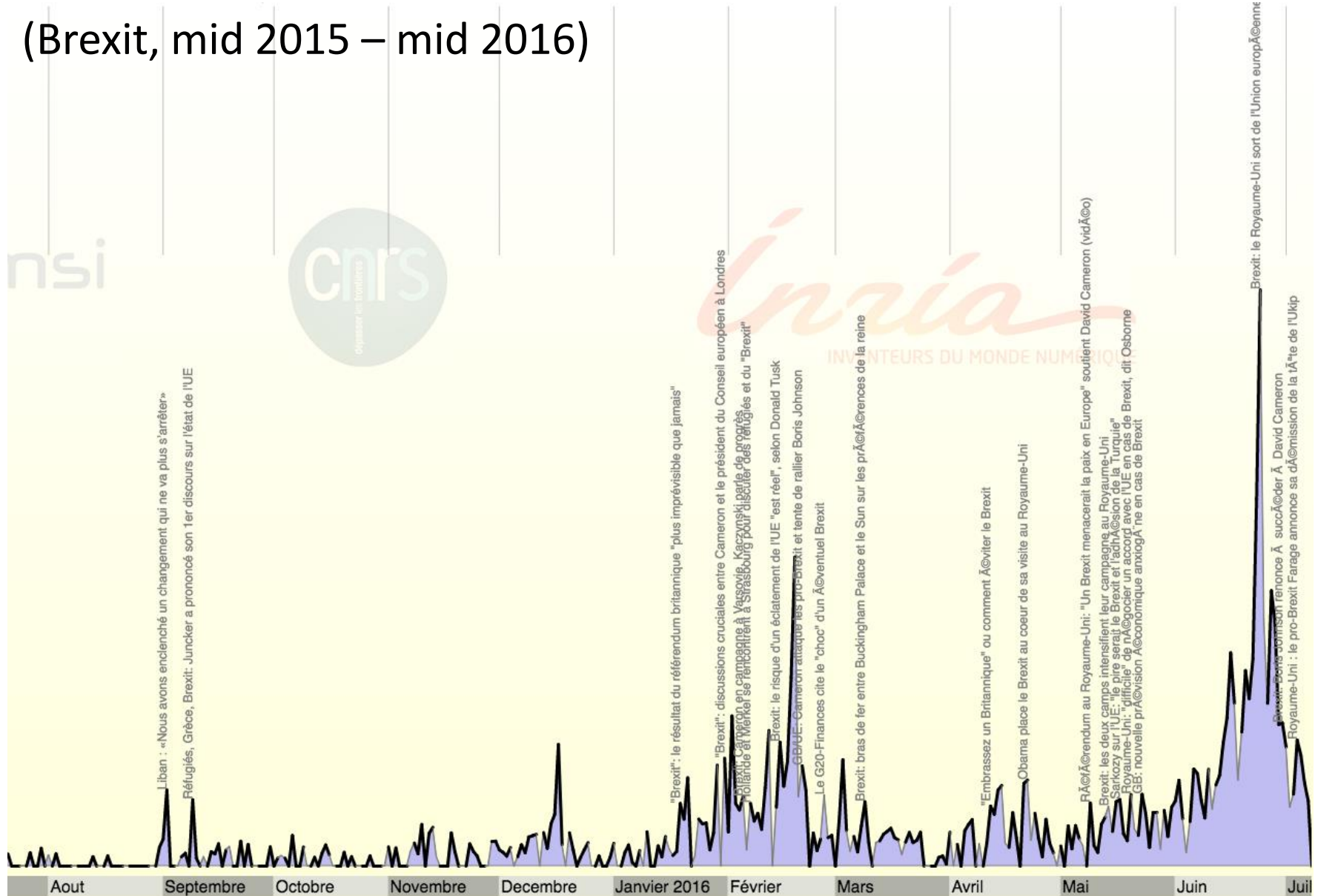
- (Current approaches:



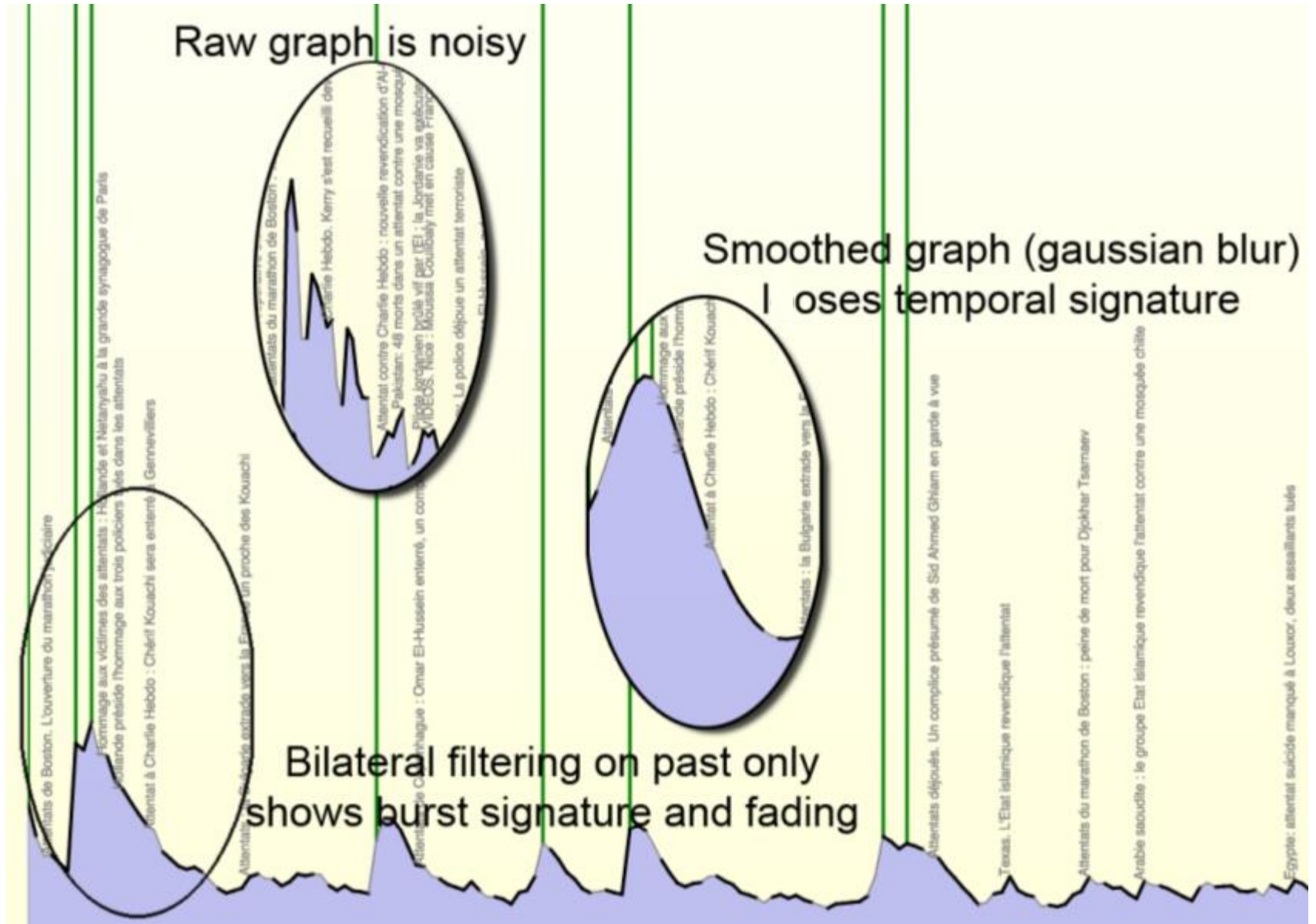
)

Raw data

(Brexit, mid 2015 – mid 2016)

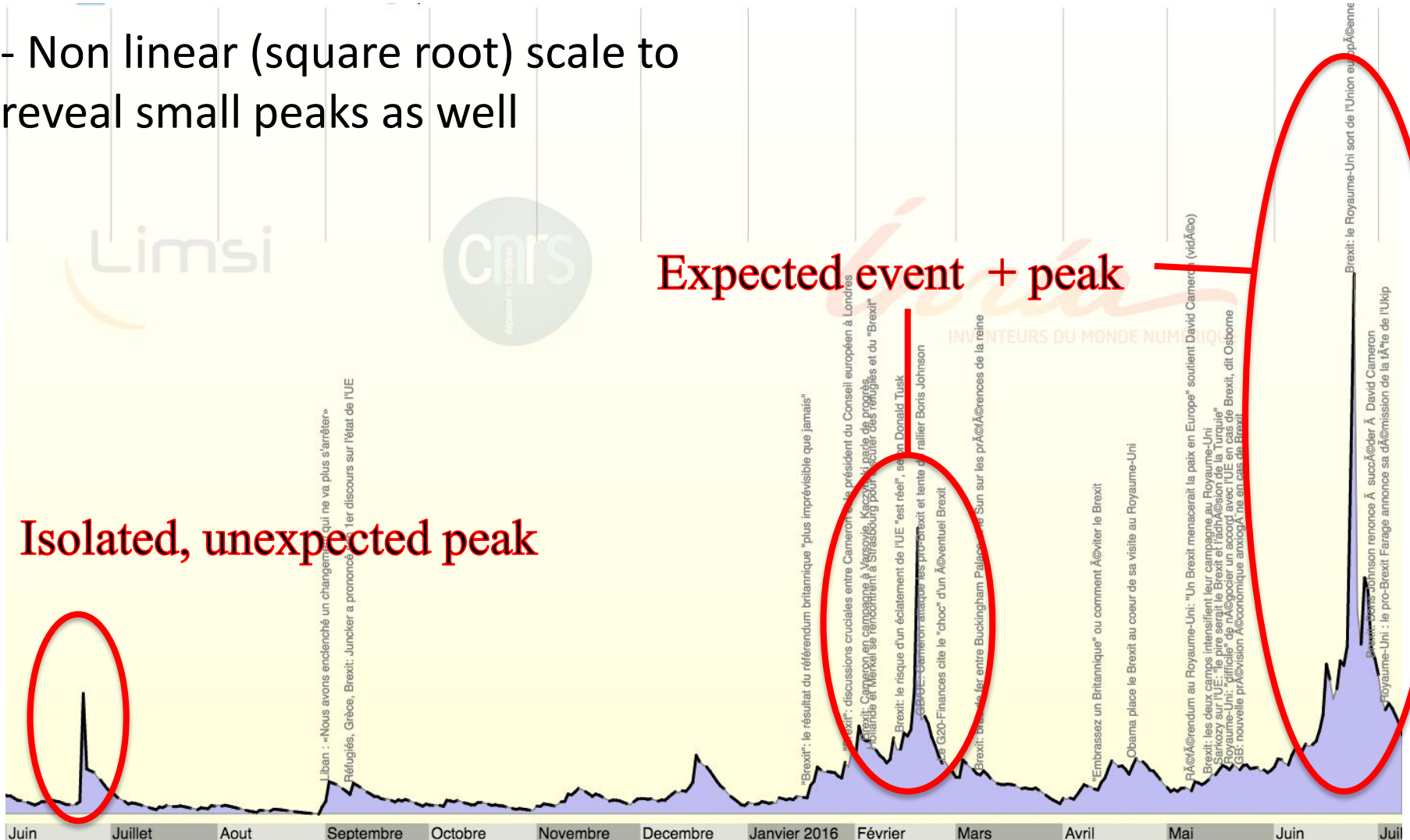


Smoothing options



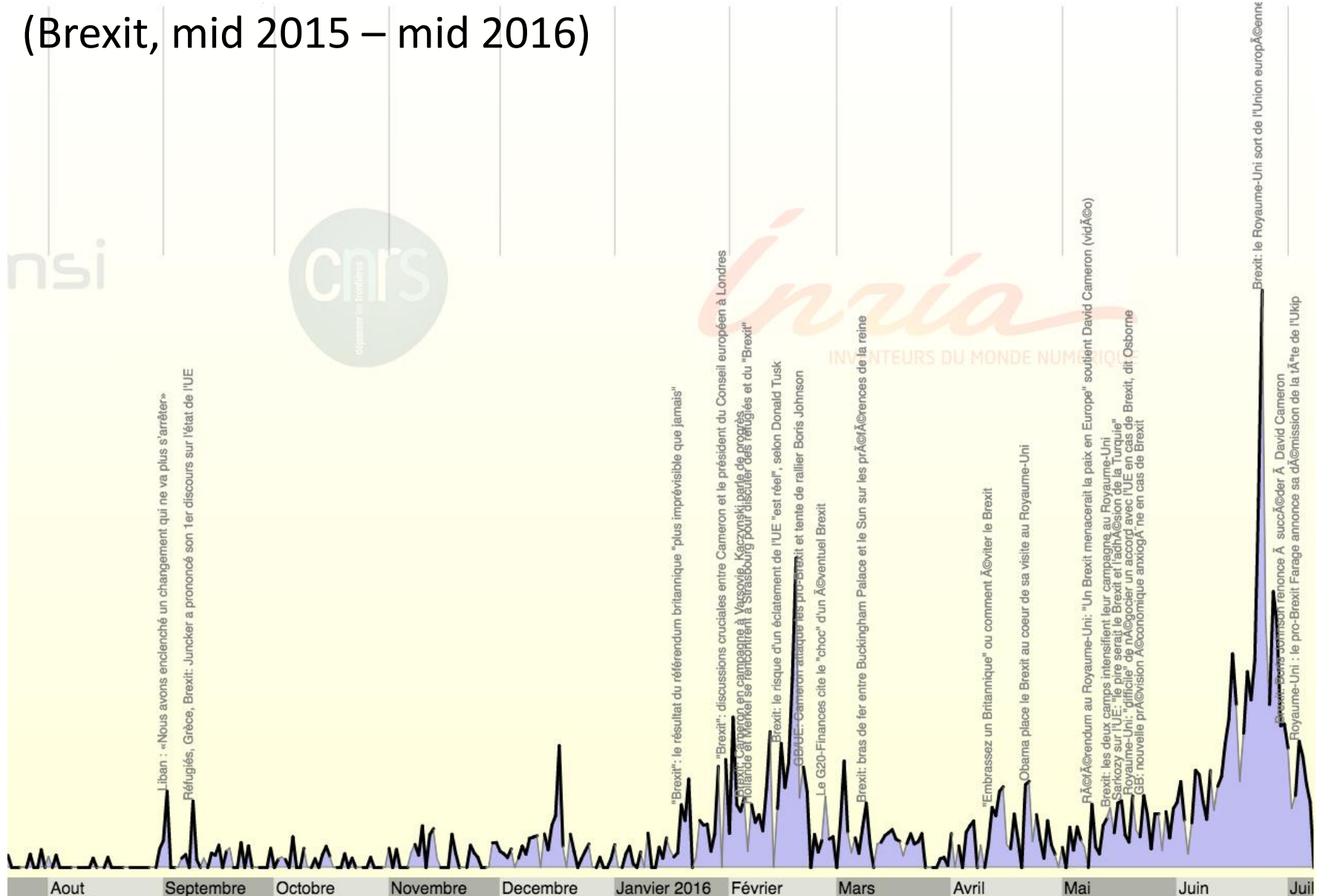
Bilateral filtering

- Smoothing but keeping edges
- Non linear (square root) scale to reveal small peaks as well



Raw data

(Brexit, mid 2015 – mid 2016)



3. Discussion, Future work

Data manipulation

- Are the results still reliable and interpretable?
 - Cons:
 - NLP tools make mistakes
 - Smoothing will hide some events
 - Saliency \neq importance
 - Granularity = day ; no more ; no less
 - Pros:
 - Data are aggregated
 - The smoothing is local
 - All data are processed the same way
 - Seeing raw data is allowed by the tool
 - The user can control the smoothness.
 - This is NOT a tool intended for end users.

Language dependence

- Study achieved on a French dataset
- Adaptation to other languages:
 - Tokenization and stemming (available for many languages)
 - Temporal normalization (Heideltime available for 13 languages, other tools may exist)

Future work

- **Use state-of-the-art date and event selection**
 - Machine learning to rank approach for salient dates (Kessler, Tannier et al., ACL 2012)
 - Event selection (Nguyen, Tannier et al., Coling 2014)
- **User-based evaluation**
 - Timeline accuracy can be automatically evaluated as multi-document summarization
 - Timeline ergonomics evaluation requires users
- **Event granularity**
- **Time efficiency**

Questions ?
