

Event Nominals: Annotation Guidelines and a Manually Annotated Corpus in French

Béatrice ARNULPHY

Xavier TANNIER

Anne VILNAT

LIMSI-CNRS, Univ. Paris-Sud
firstname.surname@limsi.fr

Context

Verbal Events

They are easier to detect and more studied than nominal events.
*Bombs have **exploded*** *Planes have **crashed** in...*

Nominal Events

Events are given names when important enough.
*September-11 **attacks*** *G20 **summit***

Nominal events are important in information extraction, but more complex and less studied

Objectives

Corpus

A corpus is needed to study, to learn and to test systems for nominal event extraction. Such a corpus did not exist for French.

Guidelines

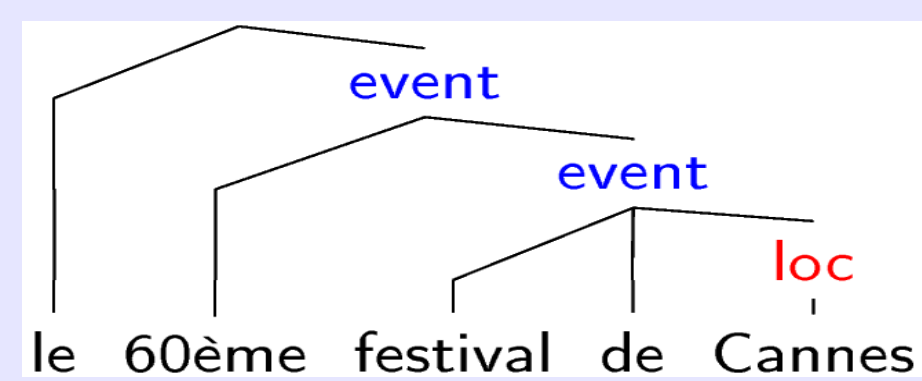
Guidelines help at building the corpus, but also at defining strict and simple definition of what is exactly a nominal event.

- Definition
- Ambiguities
- Typology
- Boundaries

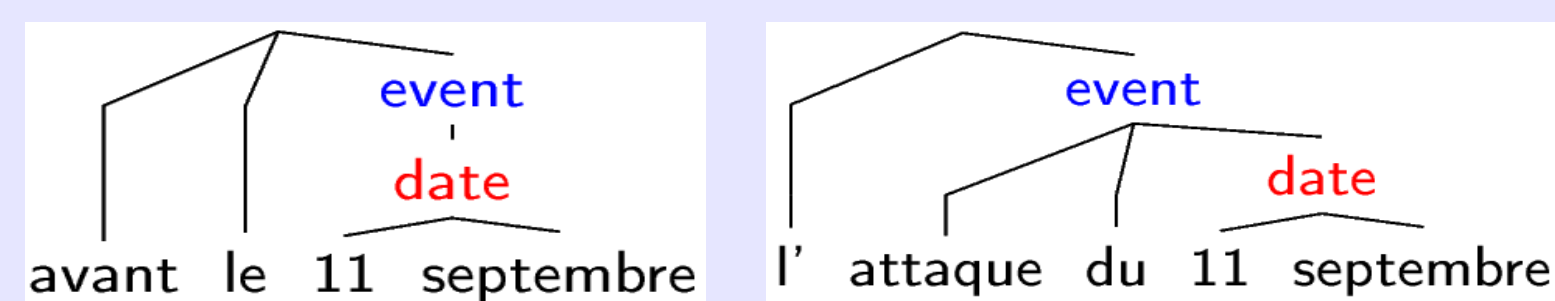
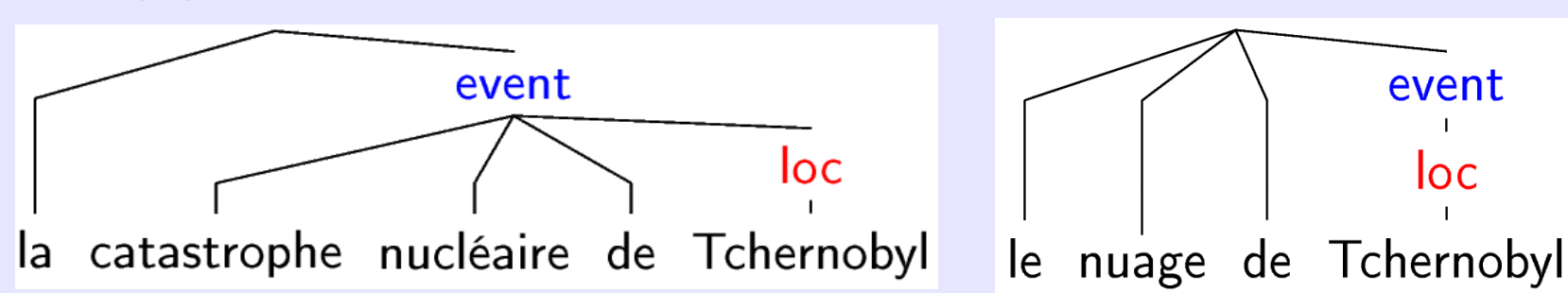
Guidelines

Quaero Named Entity Project

Our event annotation overlaps with the other entities or between events



Metonymy : Another named entity can be tagged as event



Modality

- Factual event
- Hypothetical event
- Nonfactual event
- Abstract event

Frequency

- Unique event
- Recurring event
- Instantiation of a recurring event

Anchorage to time (utterance time)

- Before
- Now
- After

Typology

En 2003, 65 soldats du feu sont morts en
<event type="hypothetical" frequency="unique" temp="before">
service
</event>.

Le
<event type="factual" frequency="instance" temp="before">
60ème
<event type="factual" frequency="recurring" temp="now">
Festival de Cannes
</event>
a eu lieu du 16 au 27 mai 2007.

La
<event type="abstract" frequency="unique" temp="now">
crise
</event>
suit une période de confiance excessive.

Annotation Tips

- **Substitute** with non-ambiguous non-event noun to disclaim the eventual reading.
- Take inspiration of **eventive** and **non-eventive uses** of the same word.
- If most of the nouns in an **enumeration** is non-ambiguously eventive, the ambiguous word should be an event.

Event Boundaries

- Annotation **according to syntax**.
- Nominal dependencies (adjectives...) + spatial and temporal complements → **Inside the scope of annotation**
- Relative and infinitive clauses → **Outside**

Corpus and Observations

Corpus

We chose to annotate a corpus of news, because of their high density of nominal events.

	Le Monde	FR-TimeBank	Total
Texts	83	109	192
Words	31,449	16,197	47,646
Events	1,107	737	1,844

As a comparison : in number of non-stative nominal events

- 3695 IT-TimeBank (Russo et al., 2011)
- 1579 (EN) corpus from (Creswell et al., 2006)
- 663 FR-TimeBank (Bittar, 2010)
- 1792 (EN) Time-Bank 1.2 (Pustejovsky et al., 2003)

Observations

Rates of singular and plural occurrences

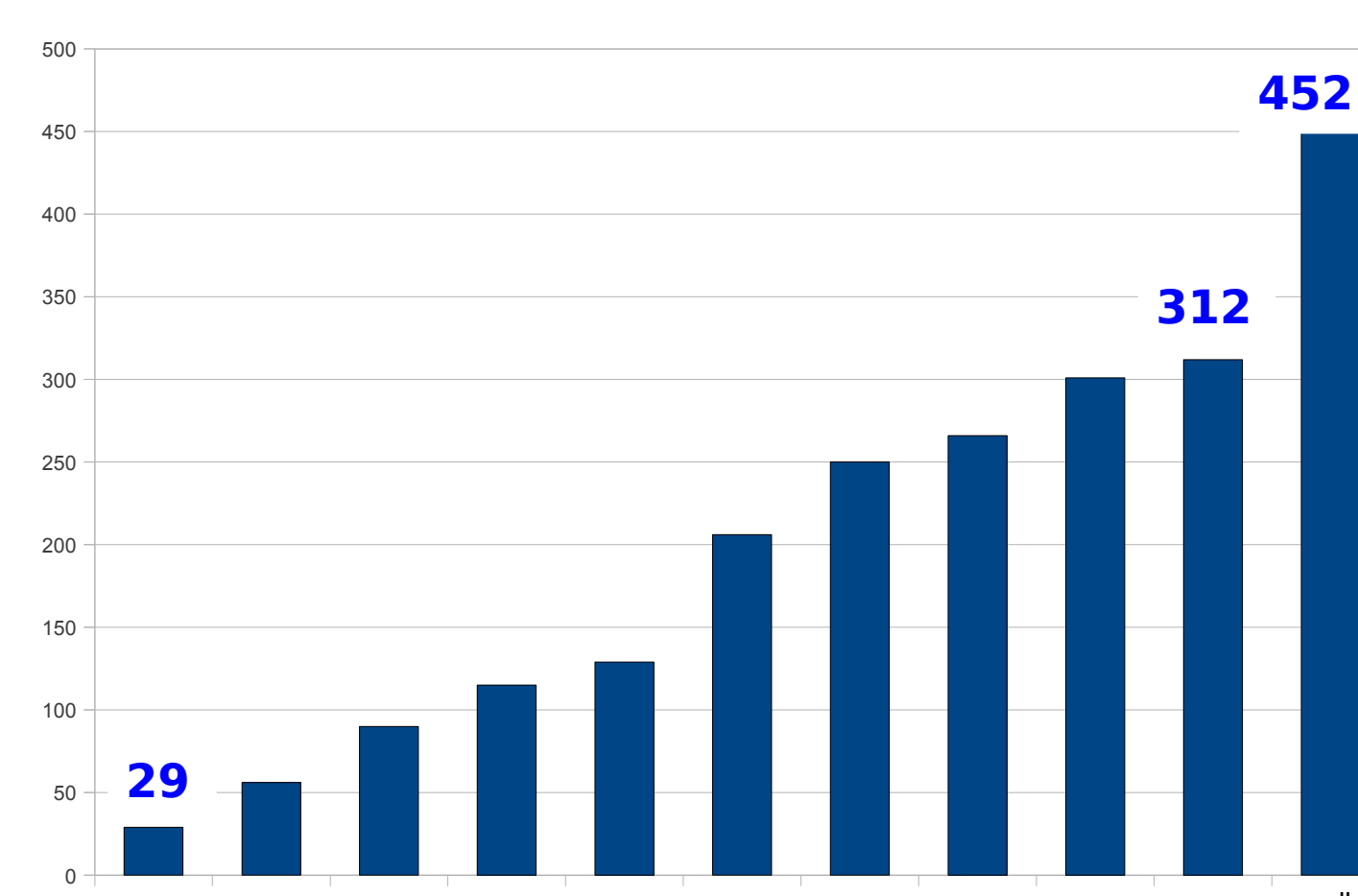
	Event nouns	All nouns
Singular	80.1%	83.4%
Plural	19.9%	16.6%

Rates of different types of determiners introducing nouns

	Event nouns	All nouns
Definite article	27.9%	19.9%
Indefinite article	14.3%	6.2%
Demonstrative	4.0%	1.7%
Possessive	6.1%	3.3%

Progression of the number of eventive nouns

according to the rate of occurrences of these nouns that have an eventive reading



For example, 29 nouns have an eventive reading in less than 10% of their occurrences. 312 nouns have an eventive reading in less than 100% of their occurrences.

Nouns having (sometimes or always) an eventive reading together with the rate of their eventive reading in the corpus

Example	Translation	Rate
disparition	disappearance	100%
meurtre	murder	100%
démission	resignation	100%
campagne	campaign/country	88.0%
peine	punishment/sadness	88.2%
vote	vote	80.0%
commentaire	comment	66.7%
bombe	bomb	50.0%
signe	sign	44.4%
mort	death/dead	37.5%
prix	price/award	22.2%
conseil	advice/council	10.7%