



# *Représentations de La Défense dans la presse écrite et sur Twitter : analyse outillée*

Marina BAIDINA (Université Paris Nanterre)

Iris ESHKOL-TARAVELLA (Université Paris Nanterre, MoDyCo UMR 7114)

Jean-Luc MINEL (MoDyCo UMR 7114)

Camille DEBRAS (Université Paris Nanterre, CREA EA370)

Valentin-Gabriel SOUMAH (Université Paris Nanterre)

# Etat de l'art

- L'espace est présent dans les **recherches en linguistique** majoritairement à travers l'étude des expressions spatiales (Boons, 1987 ; Borillo, 1998 ; Laur, 1991 ; Vandeloise, 1986 ; Le Pesant, 2011, 2012 ; Gouvert 2008).
- Du point de vue de la **lexicologie** et plus particulièrement de l'**onomastique**, l'espace est étudié du point de vue des noms de lieux et dans le cadre de la toponymie (Bouvier 1999, Roseline le Squère 2006, Dominguès & Eshkol-Taravella 2016)
- De nombreux travaux en **TAL** s'intéressent aux lieux et à leurs spécificités dans l'optique d'un traitement automatique.
  - **Entités nommées** (Nadeau & Sekine, 2009 ; Chinchor, 1998; Ehrmann 2008 ; Nouvel *et al.*, 2015, 2013; Fort *et al.* 2009) + **REN** (Finkel *et al.*, 2005 ; Al-Rafou *et al.* 2015 ; Dupont 2017 ; Stern & Sagot 1996 ; Maurel *et al.*, 2011)
  - **Entités spatiales** (Lesbeguerries 2007 ; Loustau *et al.*, 2008 ; Moncla *et al.*, 2016 ; Zenasni 2016)
  - **Noms de lieu** (Dominguès & Eshkol-Taravella 2013 ; 2015 ; 2017)
  - **Humanités Numériques Spatialisées** créée en 2019 sous l'égide du Groupement de Recherche CNRS MAGIS (Méthodes et Applications pour la Géomatique et l'Information Spatiale) => « l'extraction d'information géographique et analyse spatiale à partir de textes en SHS ».

# Représentations du quartier

**Thématiques**

Approche non supervisée

**Désignations**

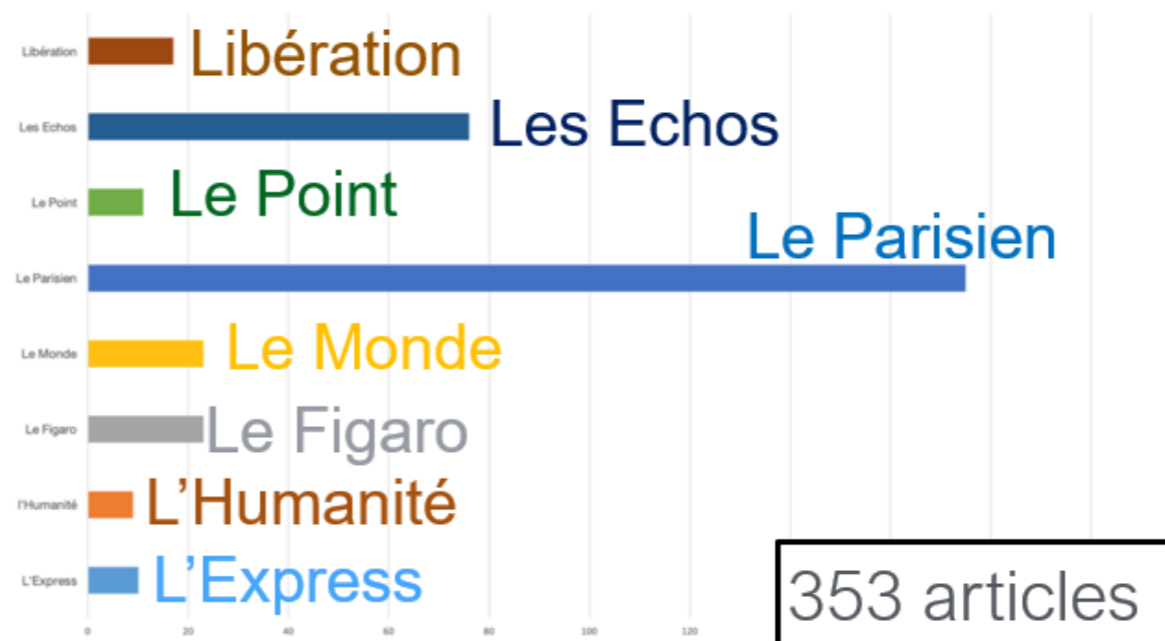
Approche supervisée

**Contexte d'emploi**

Approche supervisée

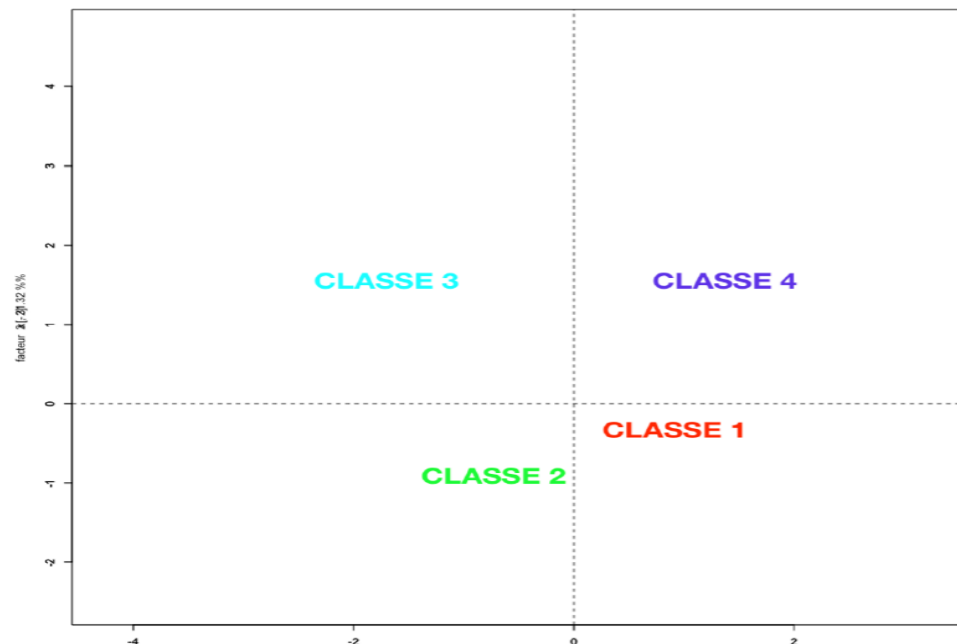
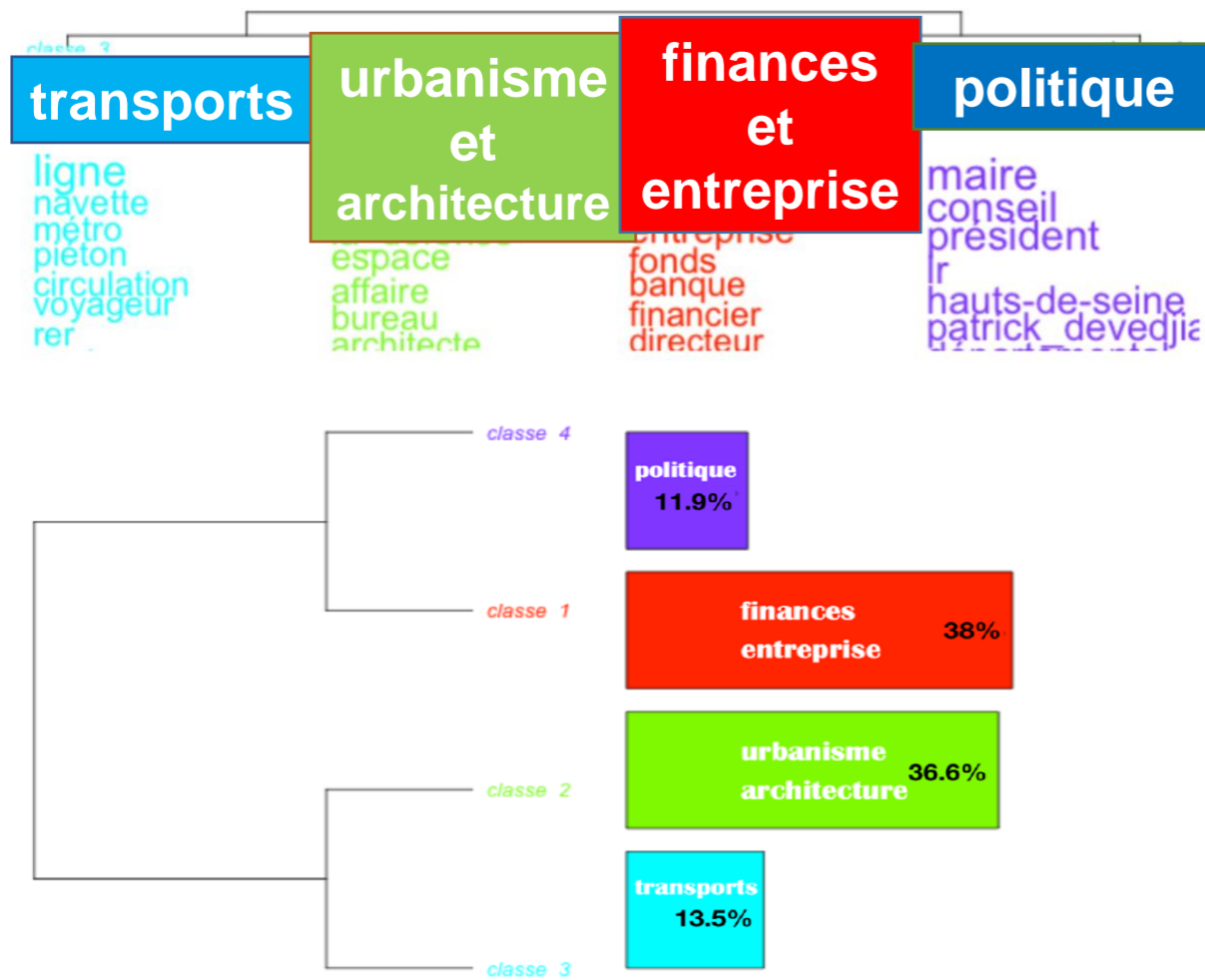
# Corpus : Presse

- Médias,
- Source officielle, institutionnelle,
- Données contrôlées
- Période de cinq ans entre 1er mars 2015 et 1er mars 2020
- La collecte du corpus des titres de presse depuis le site europresse.com (La Défense "3" quartier d'affaires)

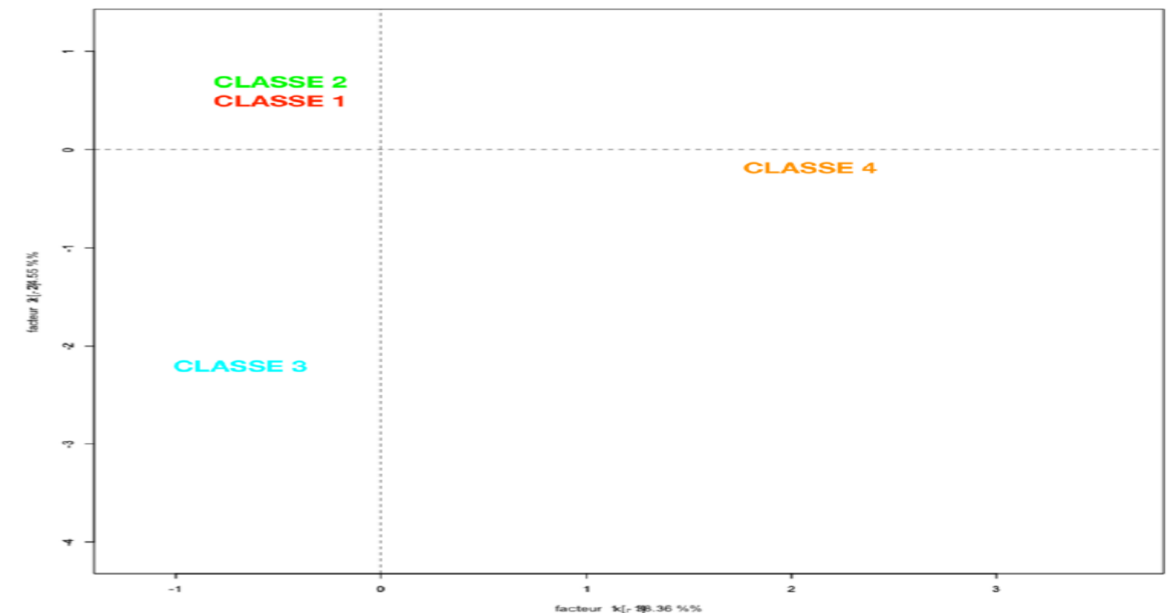
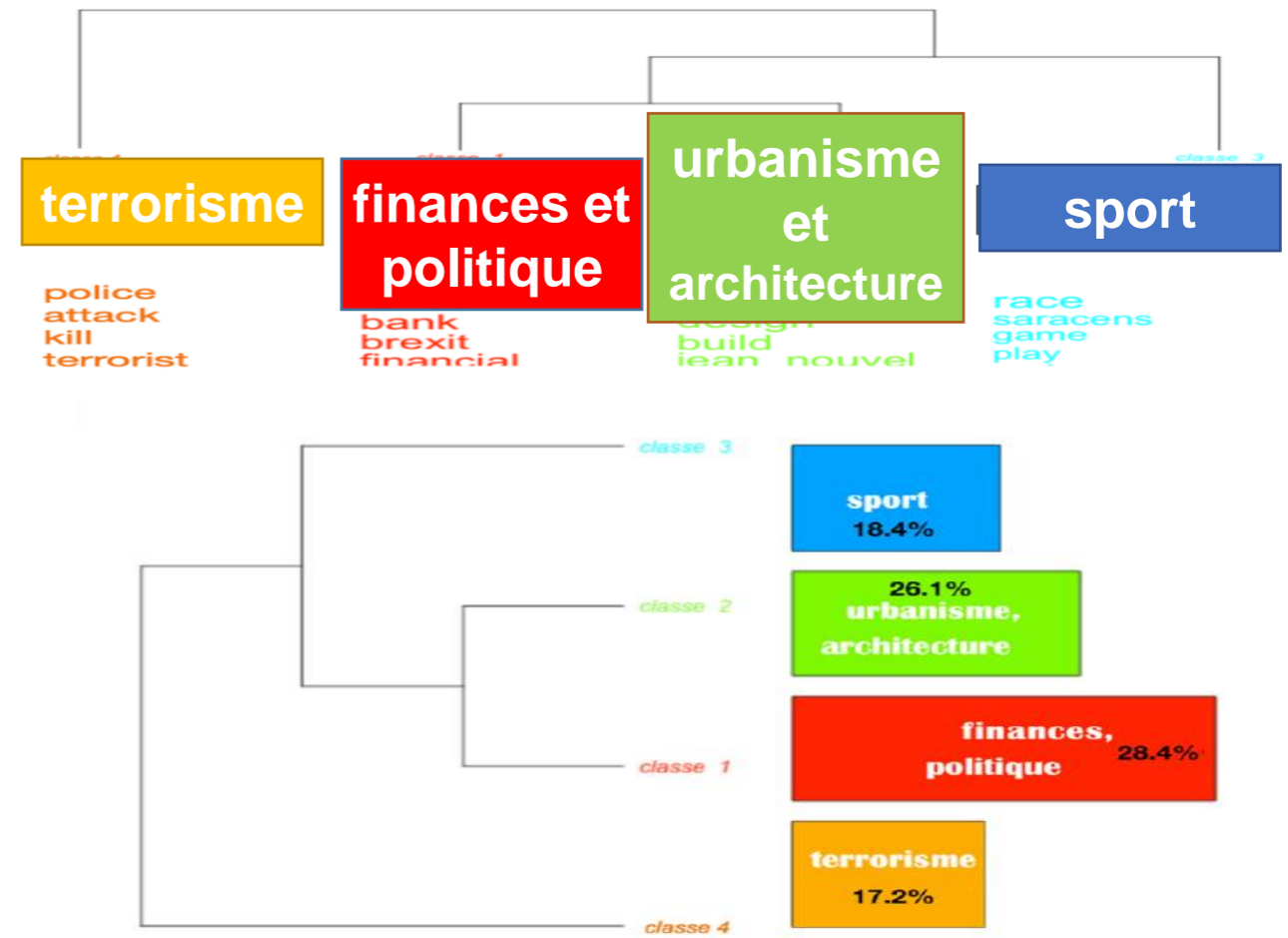


# Thématiques associées au quartier de La Défense

## Corpus français

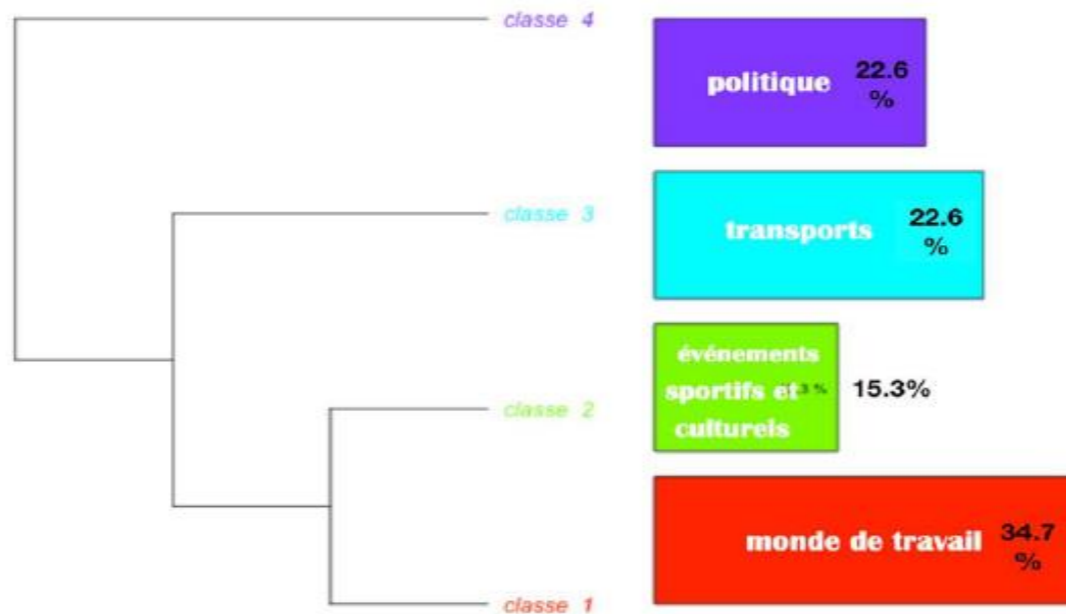
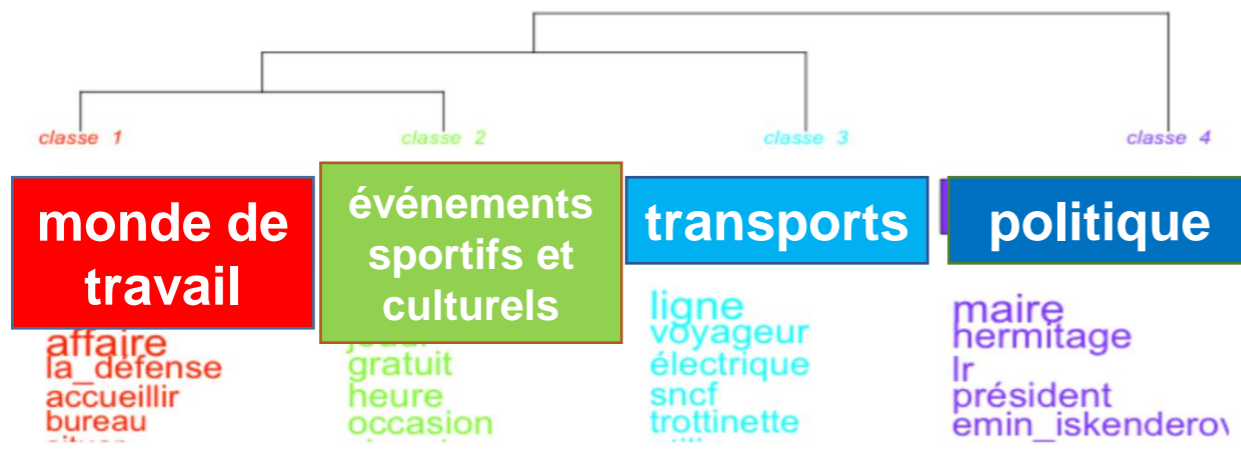


## Corpus britannique

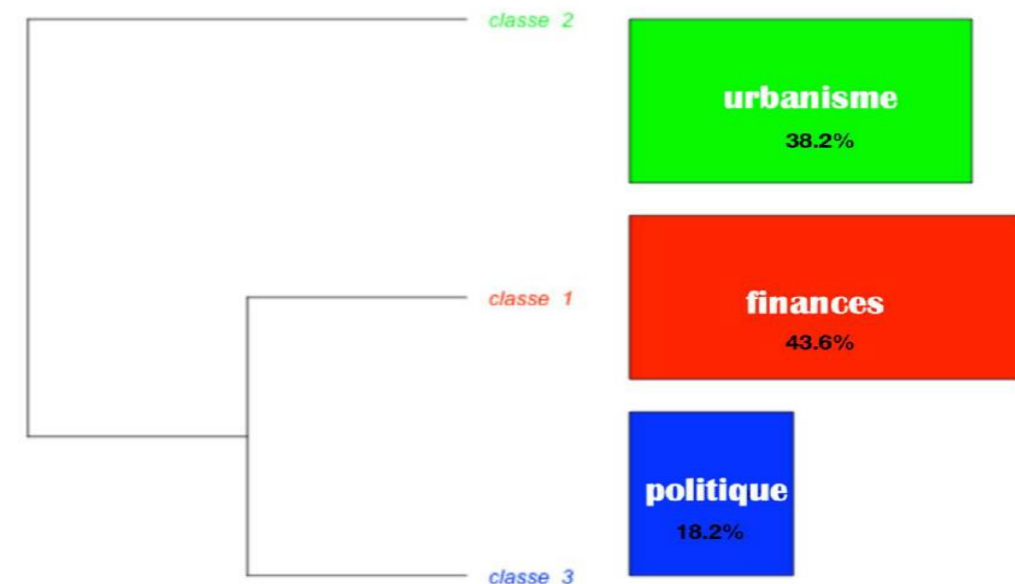
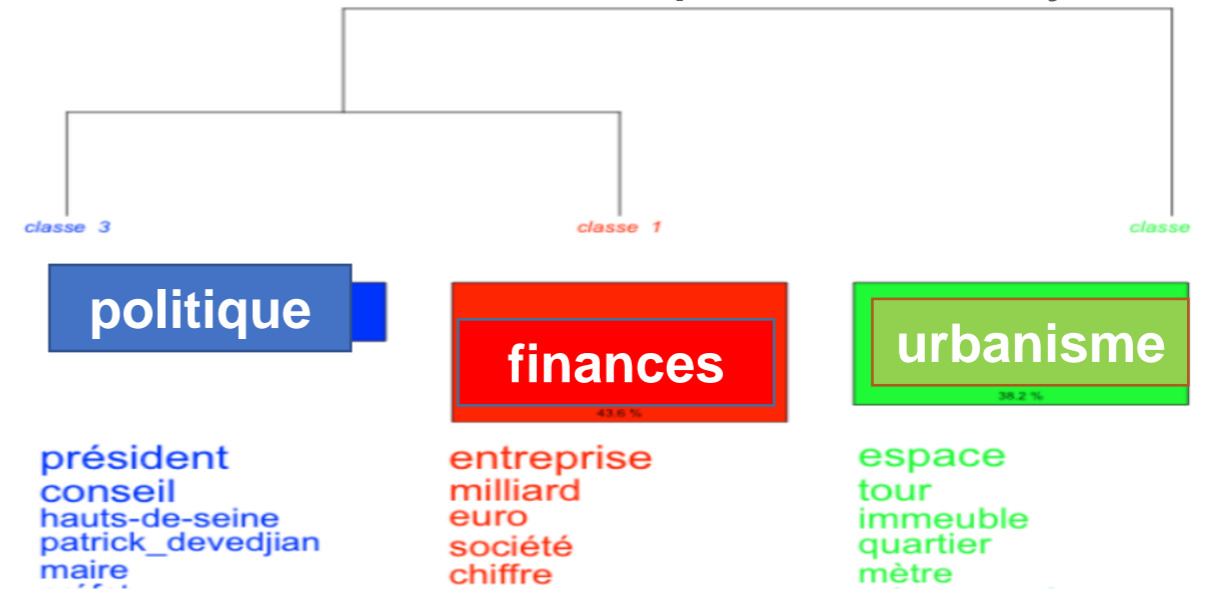


# Thématiques associées au quartier de La Défense

## Le Parisien



## Les autres titres de presse analysés

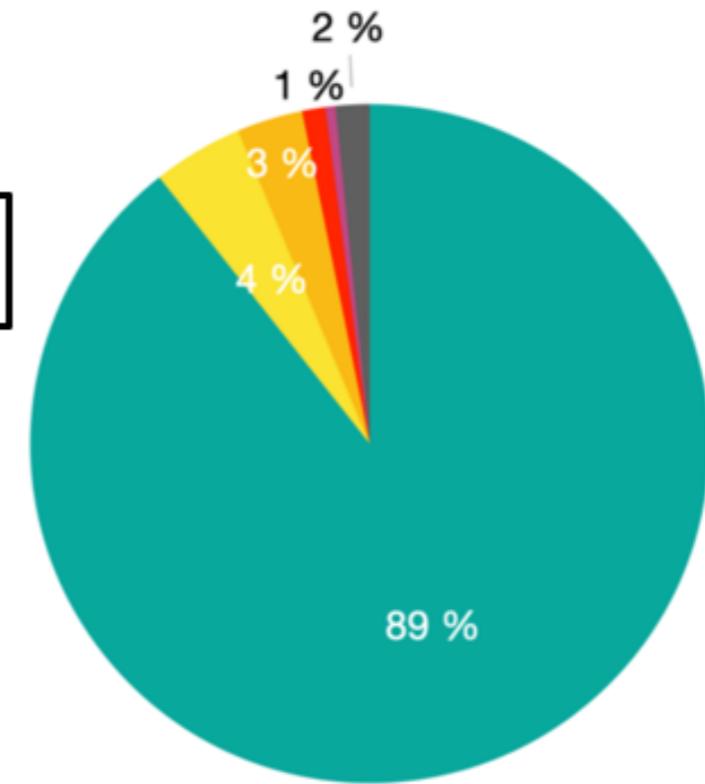


**le Parisien met d'abord en priorité les sujets qui concernent les employés de La Défense, plutôt que les entreprises qui y sont installées.**

- Réseau social,
- Données spontanées, non contrôlées
- à partir des deux hashtags #LaDefense et #ParisLaDefense

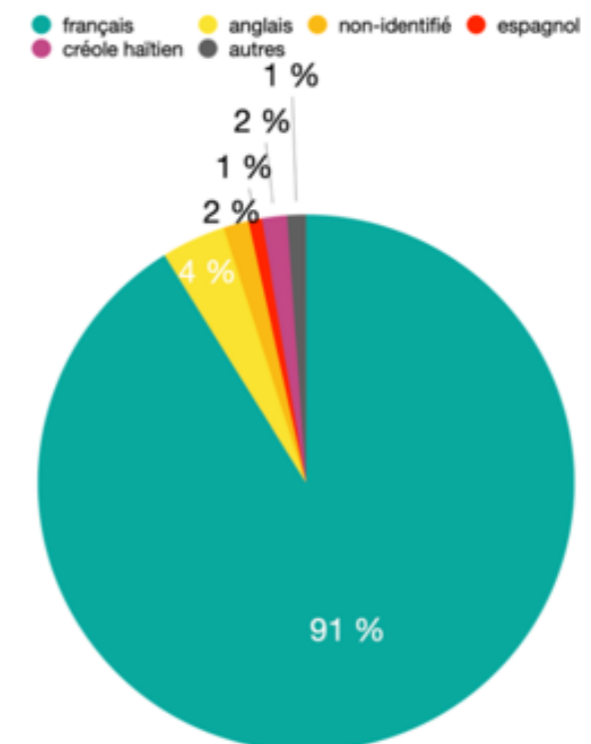
pendant le confinement 24 mars 2020 -11 mai 2020

2 988 tweets



après le confinement : 12 mai 2020 — 30 juin 2020

17 643 tweets

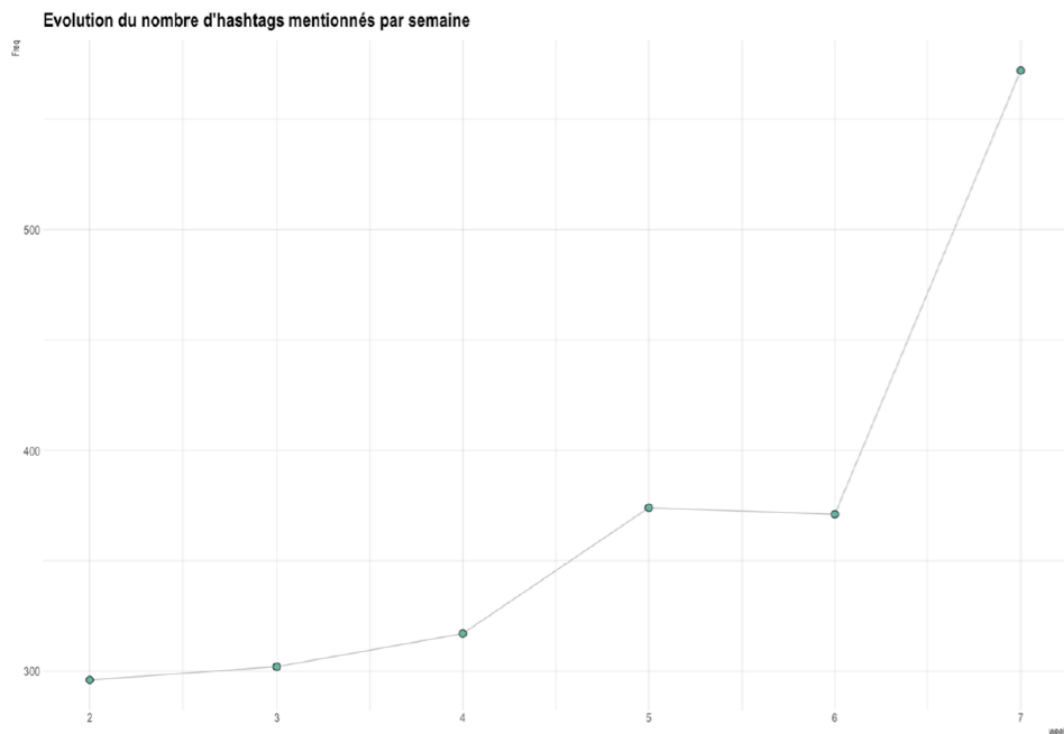


# Analyse quantitative pendant le confinement

## Analyse des hashtags

cinq hashtags les plus fréquents

Hashtag (toutes classes confondues)	Nombre d'occurrences
Courbevoie	113
RestezChezVous	110
Puteaux	103
OnApplaudit	102
Paris	82

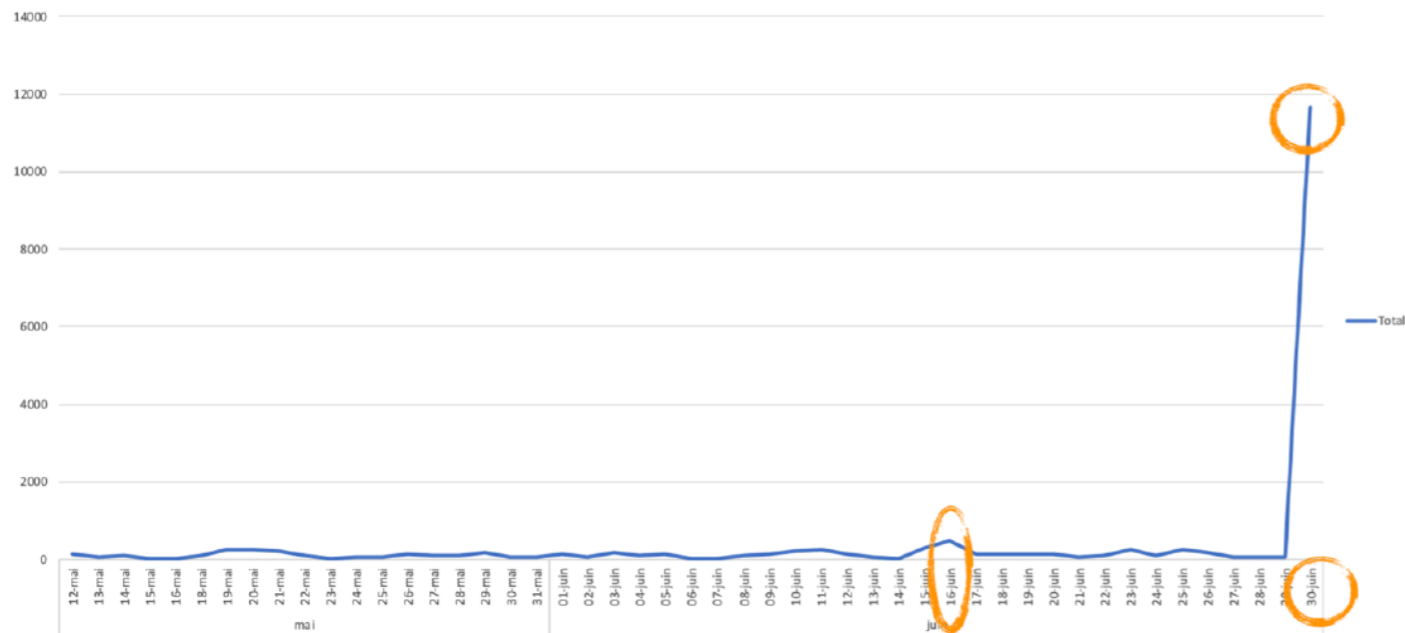


Semaine	Hashtag
24/03 – 29/03	URGENT
30/03 – 05/04	Ehpad
06/04 – 12/04	Paris
13/04 – 19/04	Courbevoie
20/04 – 26/04	COVID19
27/04 – 03/05	coronapistes
04/05 – 10/05	pistescyclablestemporaires
11/05	deconfinementjour1



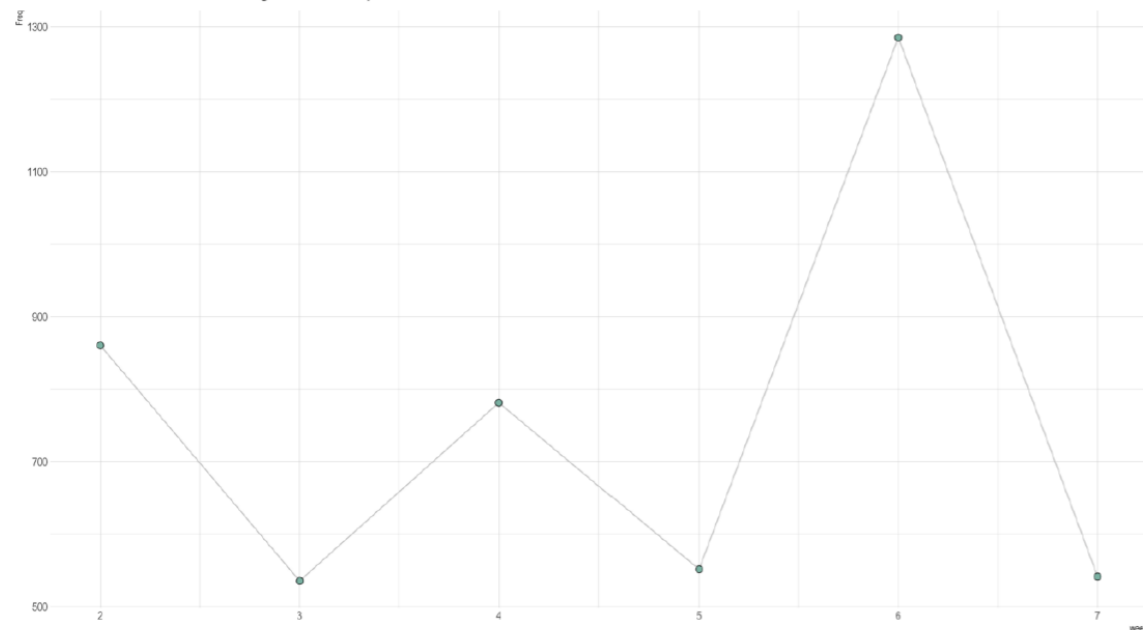
- **l'ouverture des pistes cyclables** passant par le quartier d'affaires de La Défense a créé beaucoup d'intérêt et d'enthousiasme chez les utilisateurs de Twitter :
  - **fréquence de tokens : *veloidf* et *cyclable*,**
  - **le tweet le plus retweeté** de la liste est un tweet envoyé par le compte «Collectif Vélo Île-de-France » qui **parle de l'ouverture des pistes cyclable**
  - **le nombre de tweets** envoyés a **augmenté le jour où l'ouverture des pistes cyclables** temporaires a été annoncée ;

Progression du nombre de tweets après la période de confinement



16 juin : #Extatiques l'annonce de l'exposition Extatique ayant lieu au sein du quartier de La Défense et #PoliciersenColere : une mobilisation des Policiers en Colère au quartier d'affaires de La Défense le 16 juin  
 30 juin : #attentat, #4temps, #RAID font référence à une opération policière qui a eu lieu au sein du quartier

Evolution du nombre d'hashtags mentionnés par semaine



Semaine	Hashtag
12/05 – 17/05	Courbevoie
18/05 – 24/05	Vélo
25/05 – 31/05	Puteaux
01/06 – 07/06	Eclair
07/06 - 14/06	TheLink
15/06 – 21/06	BercyCharenton
22/06 – 28/06	Extatiques
29/06 - 30/06	4temps

un pic à la sixième semaine de la collecte, la dernière semaine avant le second tour des élections municipales.

Les hashtags sont moins homogènes et semblent relever des domaines suivants :

- **Vie de résidents du quartier** : #vélo, #eclair, #Extatiques
- **Annonces** : #TheLink, #Extatiques
- **Vie politique parisienne** : BercyCharenton (faisant référence au même tweet cité auparavant : A.Hidalgo est la Maire du béton : #BercyCharenton, #TEPMénilmontant..)
- **Evénements** : #4temps, #Extatiques
- **Localisation** : #Puteaux, #Courbevoie

=> après le déconfinement la vie du quartier a repris son cours ; cette période a été caractérisée par un nombre d'événements variant en ampleur : allant d'un grand orage à une opération policière

# **Representations du quartier de *La Défense***

**Désignations**

**Contexte  
d'utilisation**

**Approche supervisée**

# Typologie des désignations

## • Procédé de nommage

- Norme
- Type de Lieu
- Référence Géographique
- Métaphore
- Classement

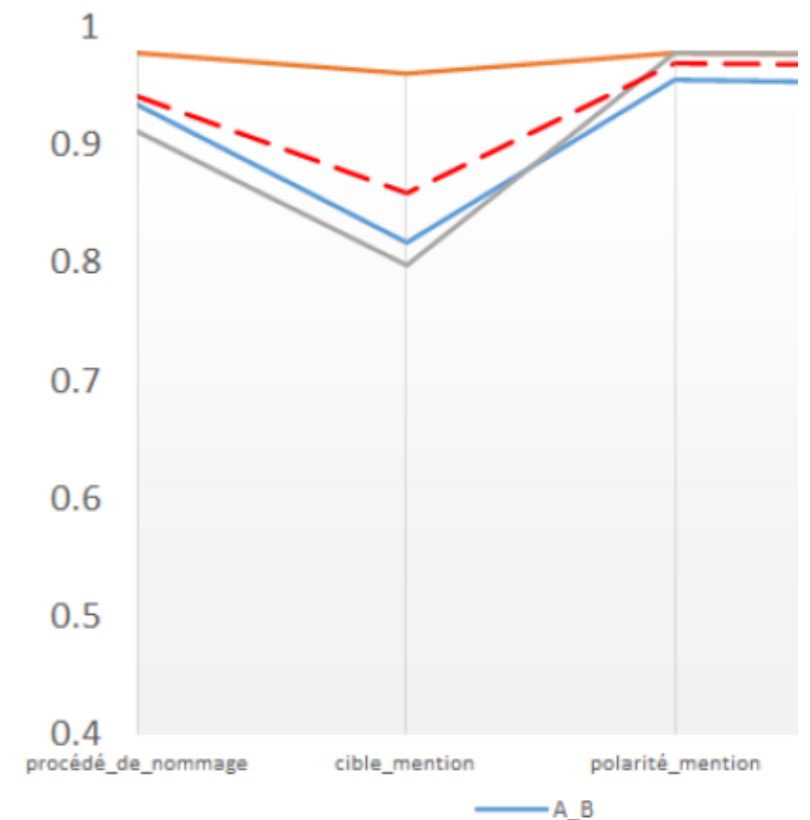
## • Cible

- Localisation
- Économie
- Urbanisme
- Architecture
- Taille
- Disparité
- Changement
- Espace de Vie

## • Polarité

- Positif
- Neutre
- Négatif

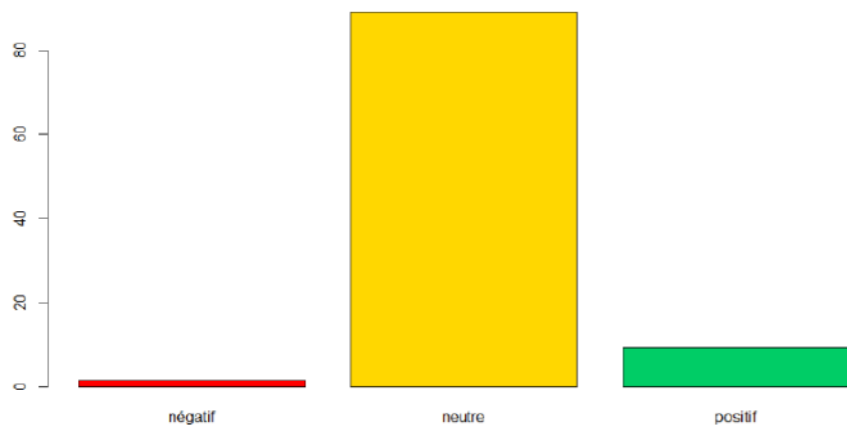
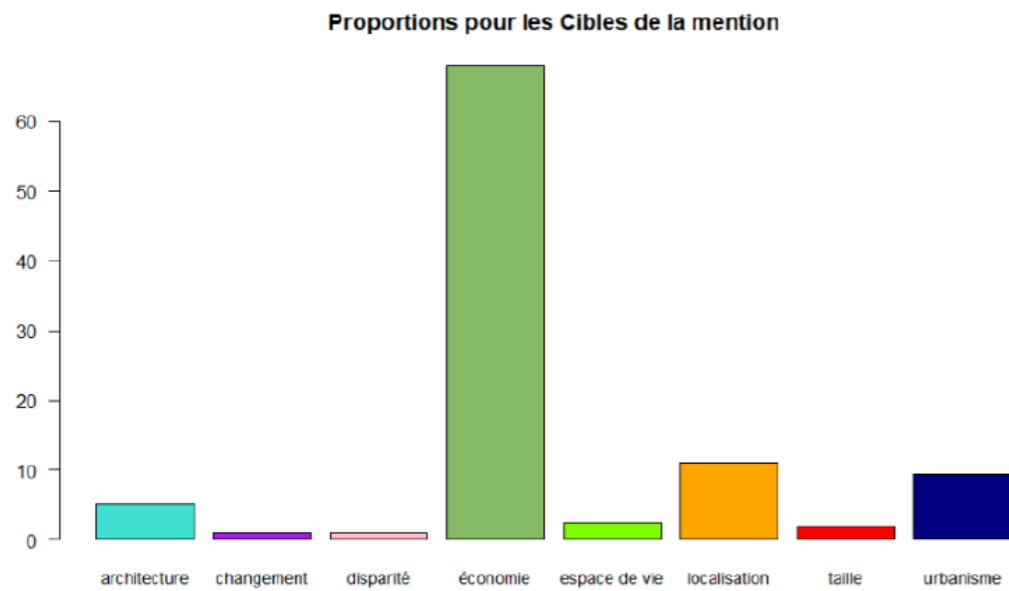
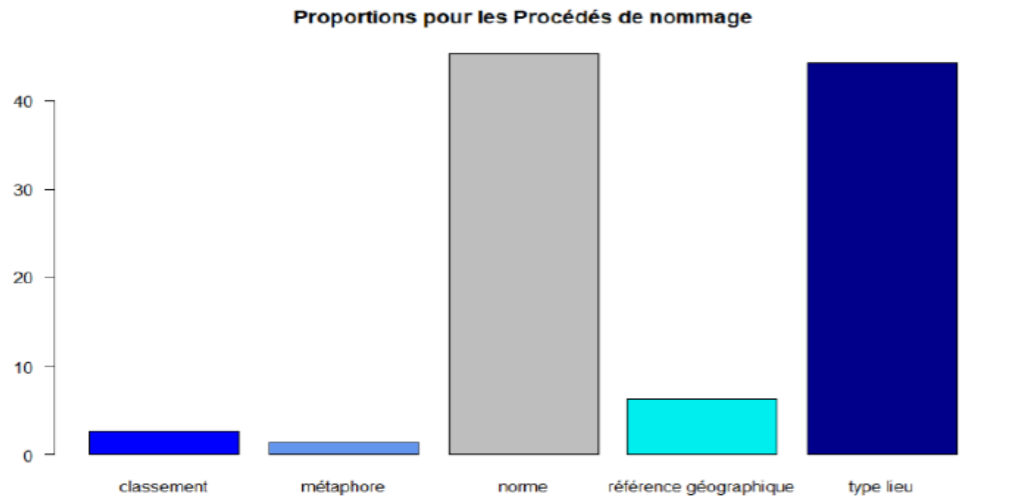
*Kappas de Cohen et de Fleiss entre les annotateurs*



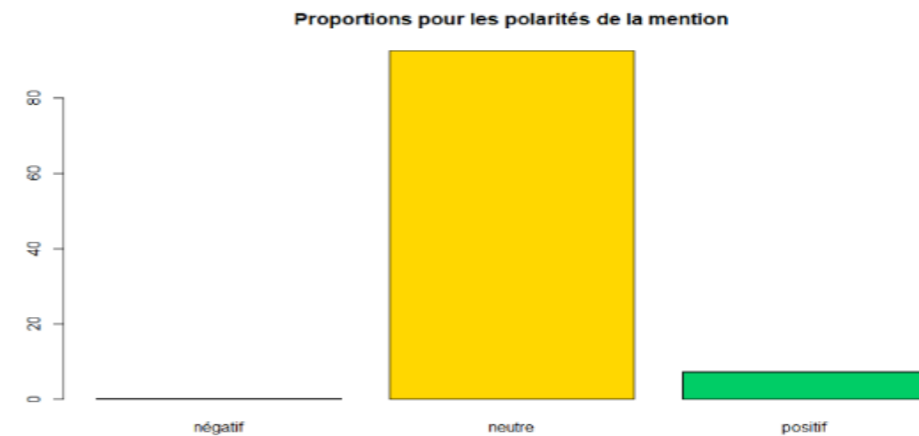
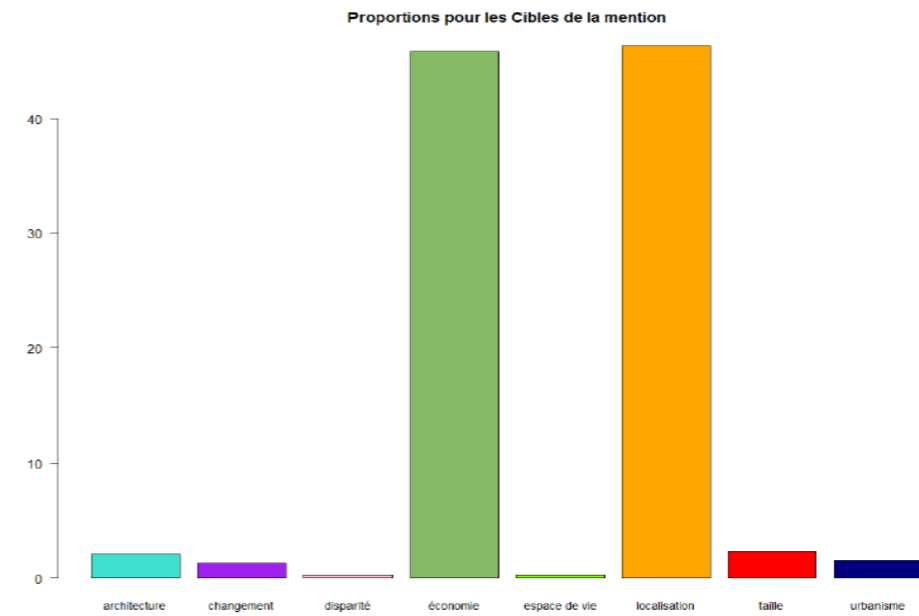
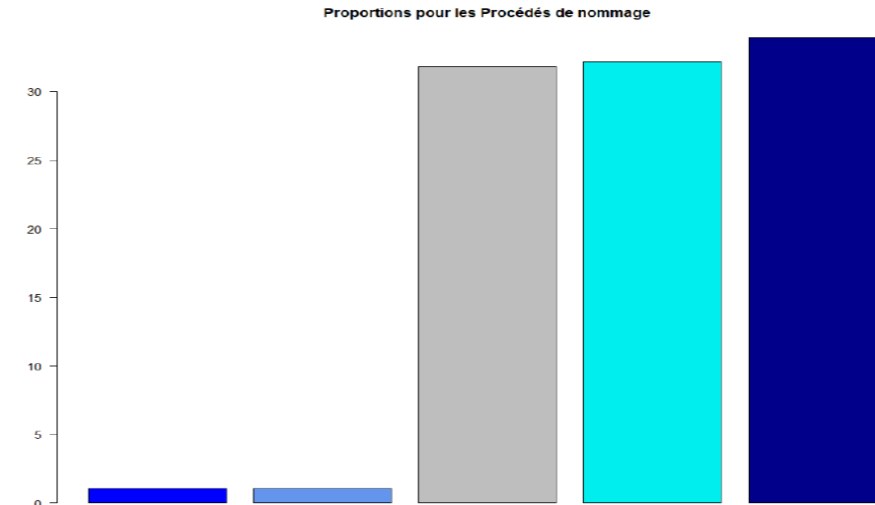
Désignation	Procédé de nommage	Cible	Polarité
Cette pépite de la région parisienne	Métaphore, Géographique	Référence, Localisation, Economie	Positif

# Analyse quantitative de désignations du quartier

- Corpus français



- Corpus britannique



# Typologie du contexte

- **Cible**

- Economie
- Politique
- Urbanisme
- Immobilier
- Quotidien
- Evénement
- Tourisme

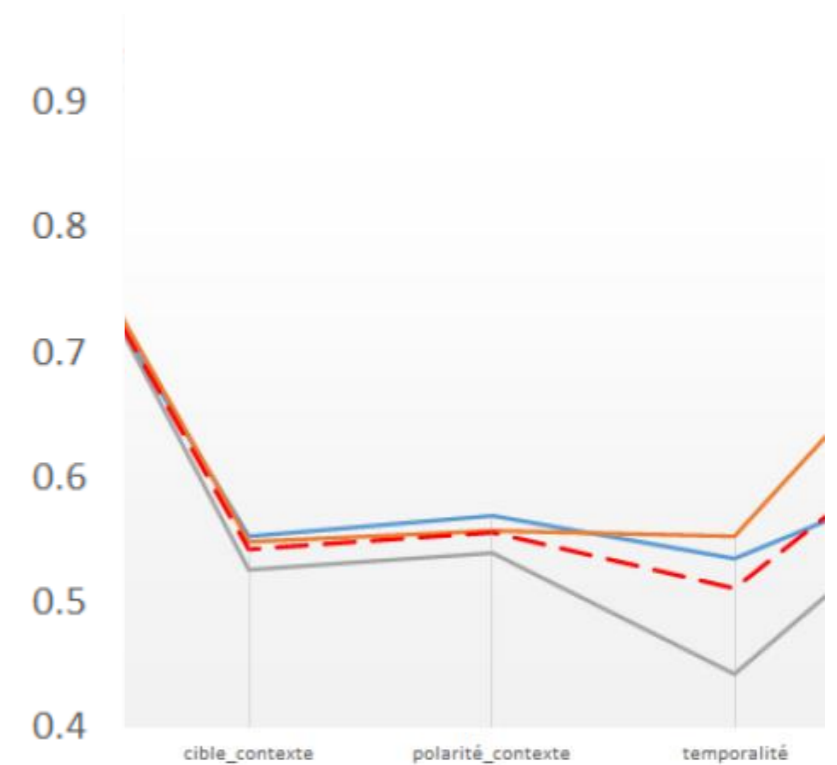
- **Polarité**

- Positif
- Neutre
- Négatif

- **Temporalité**

- Passé
- Présent
- Futur

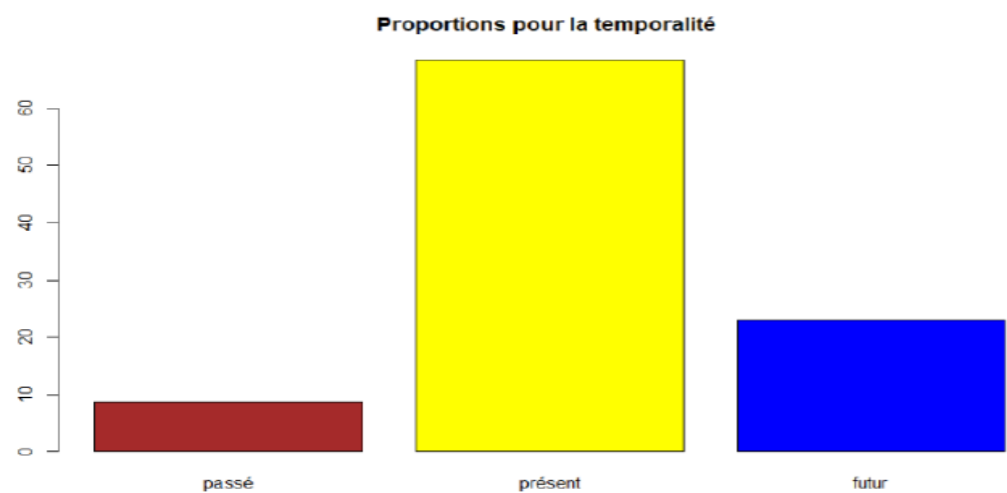
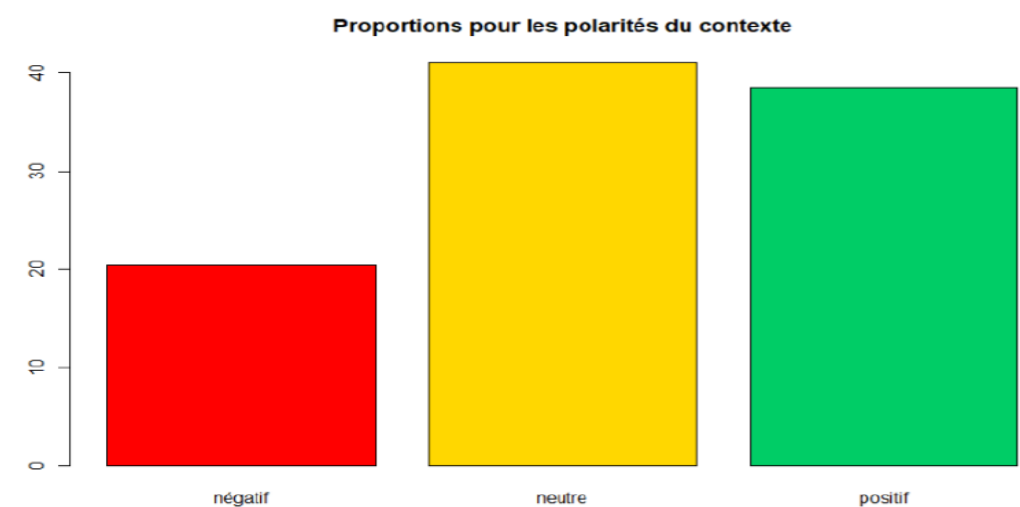
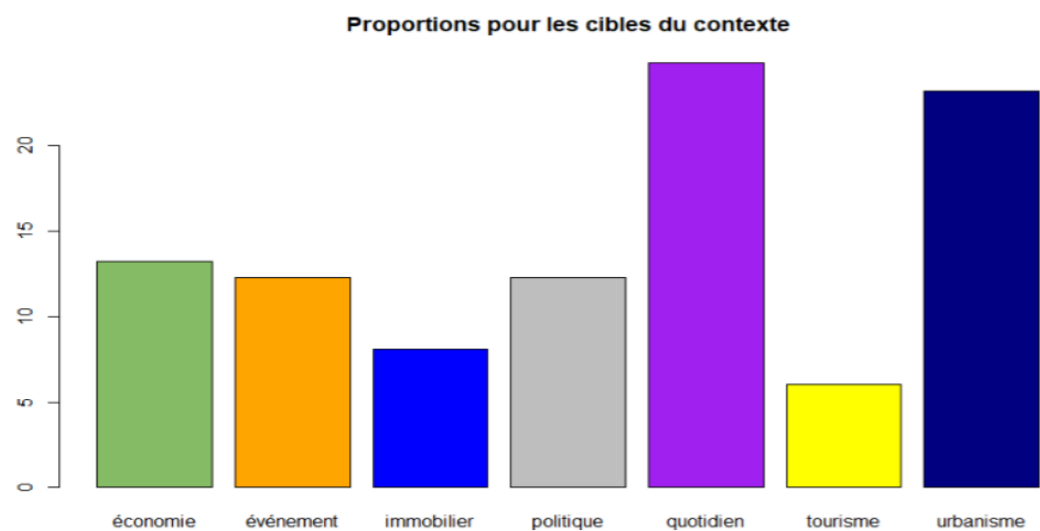
1 Kappas de Cohen et de Fleiss entre les annotateurs



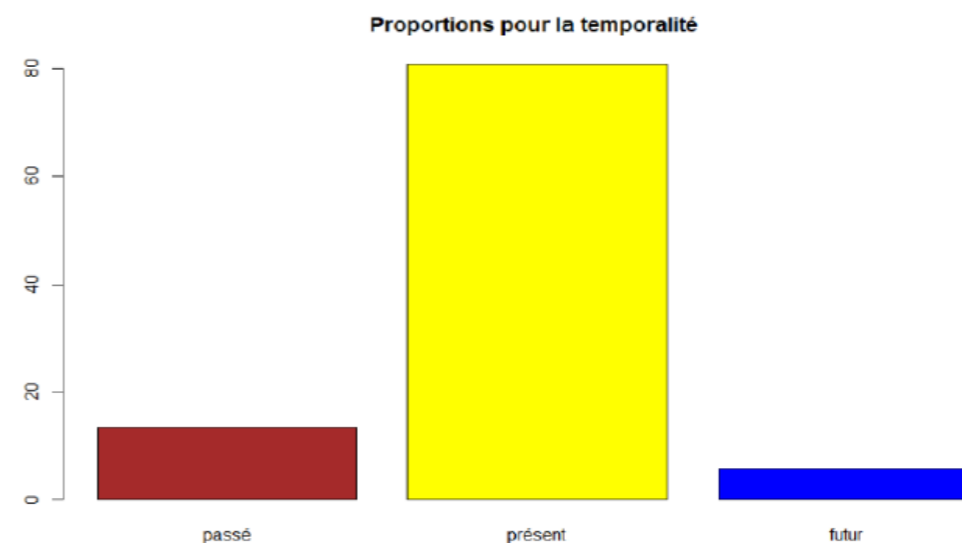
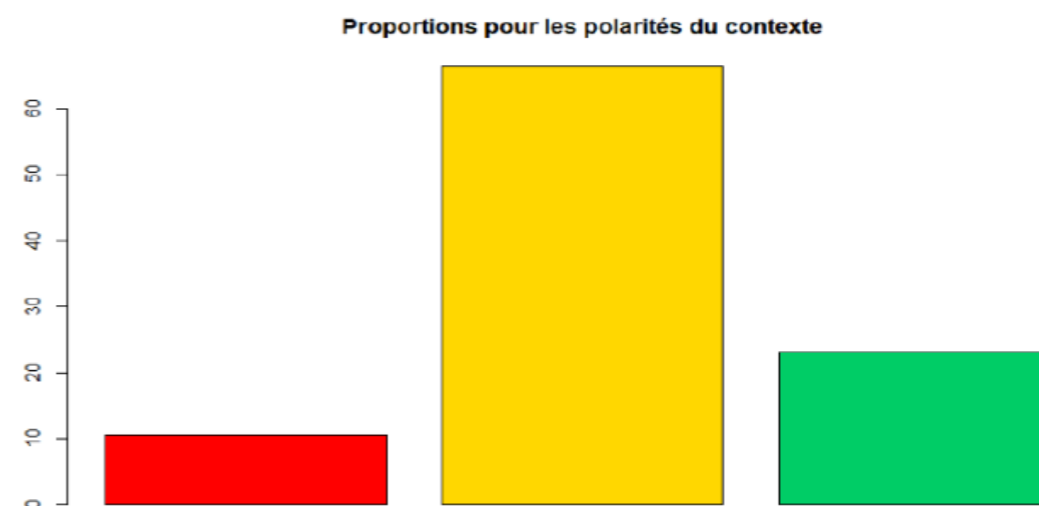
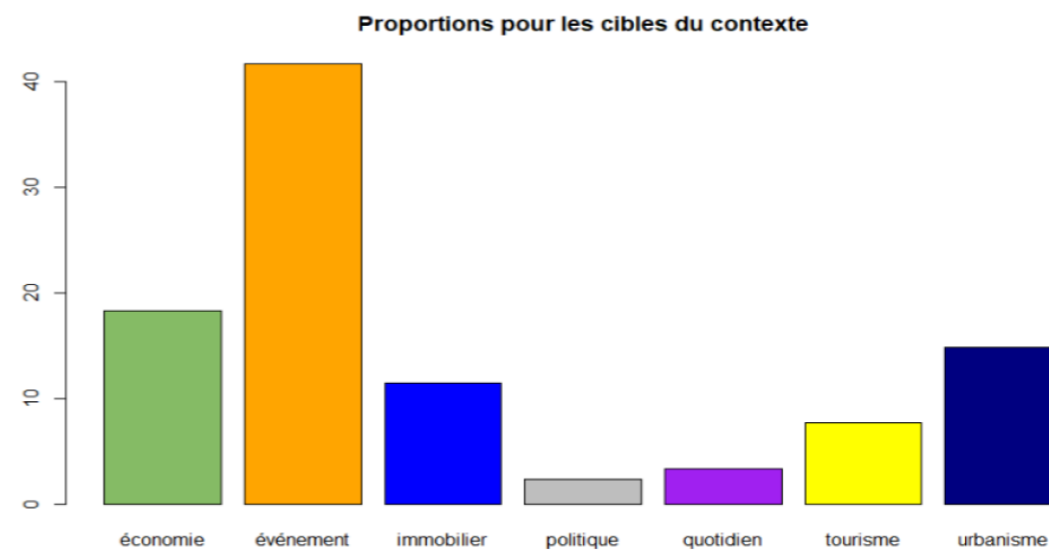
Phrase	Cible	Temporalité	Polarité
Bien <u>heureusement</u> , de <u>nouvelles lignes de transport</u> (Grand Paris Express, RER E) vont desservir <b>La Défense</b> <u>dans les prochaines années</u> .	Quotidien	Futur	Positif

# Analyse quantitative du contexte

- Corpus français



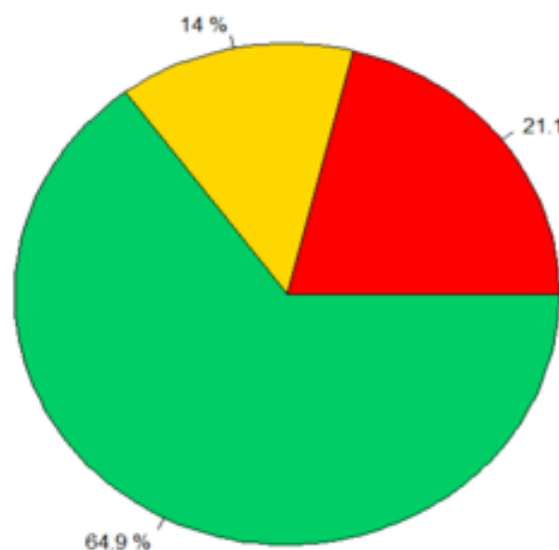
- Corpus britannique



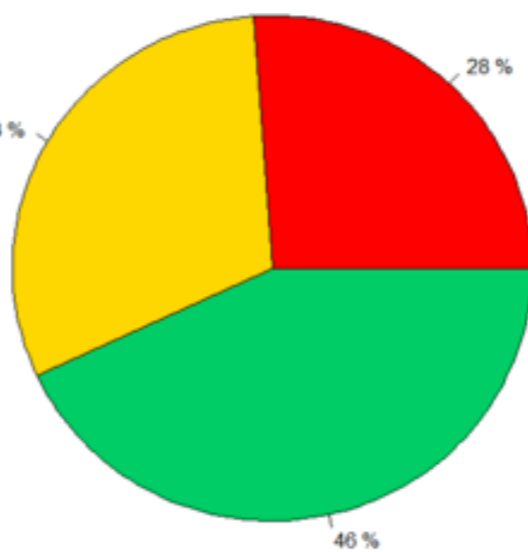
# Analyse quantitative du contexte

- Corpus français

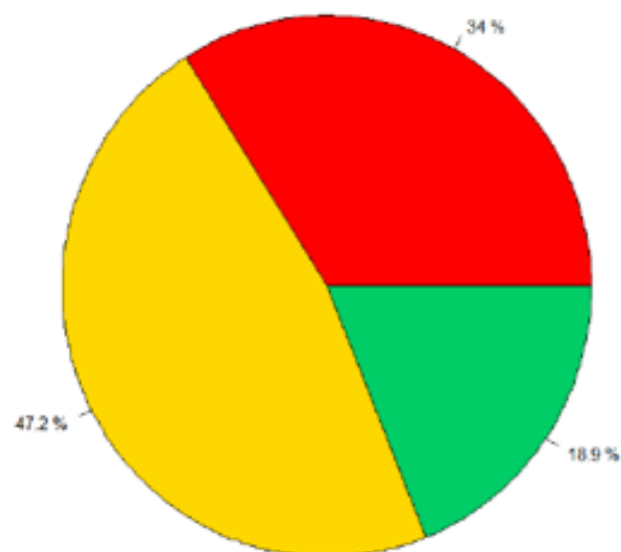
Polarité / Economie



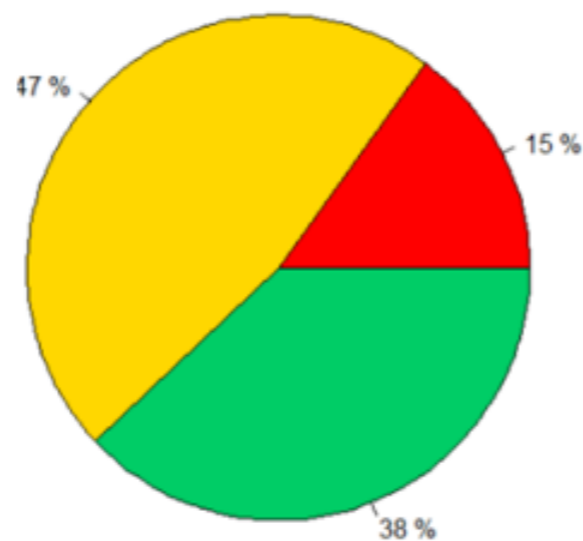
Polarité / Vie quotidienne



Polarité / Événement

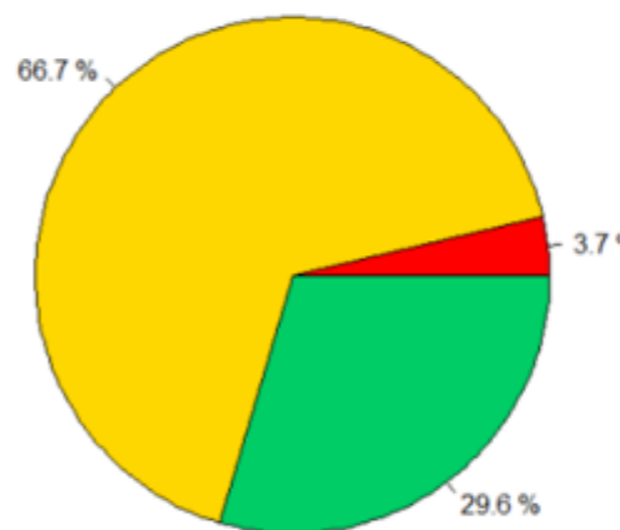


Polarité / Urbanisme

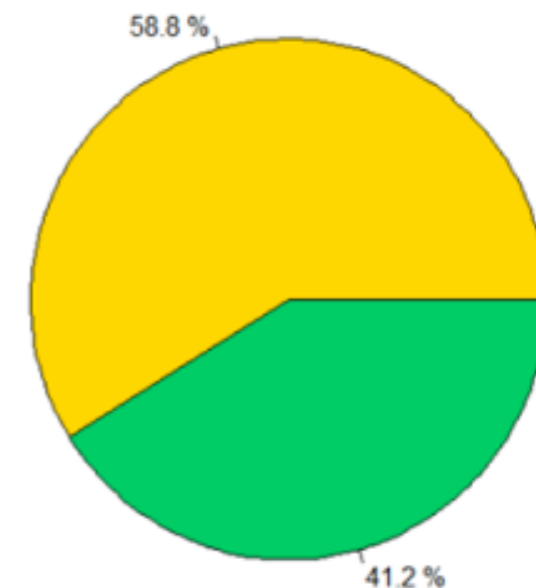


- Corpus britannique

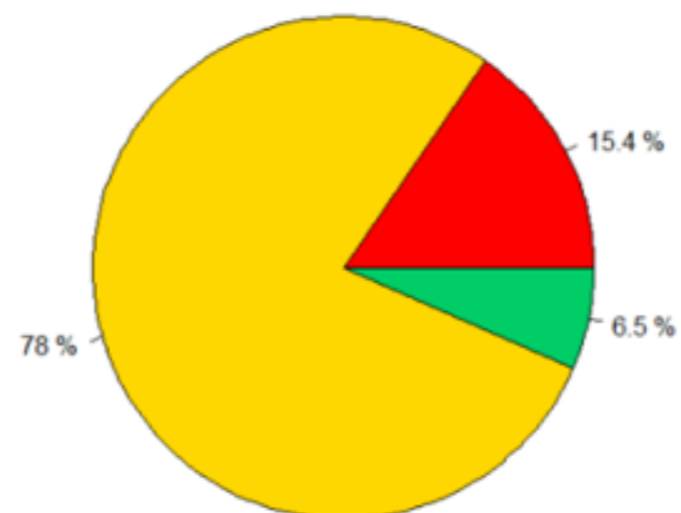
Polarité / Économie



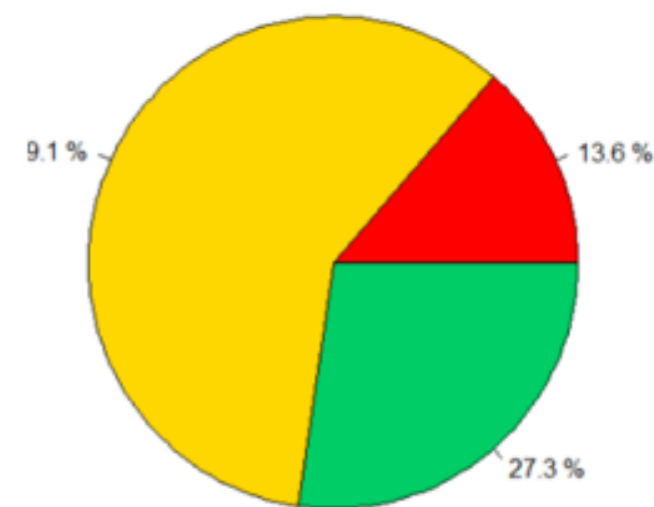
Polarité / immobilier



Polarité / Événement



Polarité / Urbanisme





# Annotation automatique de désignations

- ▶ Transformation en BIO Schema
- ▶ Modèle : CamemBERT/RobERTa/W2V
- ▶ Validation croisée : 5 plis (0.80 / 0.20)

	Recall	Precision	F-Measure
<b>O</b>	<b>0.978</b>	<b>0.965</b>	<b>0.971</b>
<b>B-La_Defense</b>	0.817	0.746	0.779
<b>I-La_Defense</b>	0.901	0.899	0.899

	Recall	Precision	F-Measure
<b>Classement</b>	1.00	0.96	0.98
<b>Métaphore</b>	<b>0.81</b>	<b>0.85</b>	<b>0.83</b>
<b>Référence géographique</b>	0.99	0.97	0.98
<b>Type de lieu</b>	0.93	0.95	0.94
<b>Norme</b>	1.00	1.00	1.00

	Recall	Precision	F-Measure
<b>Architecture</b>	1.00	0.81	0.89
<b>Changement</b>	0.88	0.93	0.90
<b>Disparité</b>	1.00	0.93	0.97
<b>Espace de vie</b>	0.93	0.95	0.94
<b>Localisation</b>	0.94	0.79	0.86
<b>Taille</b>	0.88	0.93	0.90
<b>Urbanisme</b>	<b>0.72</b>	<b>0.87</b>	<b>0.79</b>
<b>Economie</b>	0.98	0.96	0.97

	Recall	Precision	F-Measure
<b>Neutre</b>	<b>0.96</b>	<b>0.97</b>	<b>0.96</b>
<b>Négatif</b>	0.50	0.57	0.53
<b>Positif</b>	0.72	0.69	0.70

# Annotation automatique du contexte

- ▶ Modèle : CamemBERT
- ▶ Validation croisée : 5 plis (0.80 / 0.20)

	Recall	Precision	F-Measure
Politique	0.77	0.72	0.74
Immobilier	0.52	0.56	0.54
Tourisme	0.56	0.58	0.57
Urbanisme	0.72	0.77	0.76
Vie quotidienne	0.75	0.77	0.76
Economie	0.81	0.76	0.79
Événement	0.78	0.69	0.73

	Recall	Precision	F-Measure
Futur	0.65	0.65	0.65
Passé	0.64	0.51	0.57
Présent	0.82	0.85	0.84

	Recall	Precision	F-Measure
<b>Neutre</b>	0.68	0.72	0.70
<b>Négatif</b>	0.71	0.68	0.69
<b>Positif</b>	0.70	0.68	0.69

# Conclusion

- Représentations de *La Défense*
  - Thématiques
  - Désignations
  - Contexte d'emploi
- Analyse comparative
  - Presse française / britannique
  - Presse / Tweeter
- Approche non supervisée / supervisée
- Techniques de la linguistique de corpus et du TAL
- Représentations d'un lieu
  - les pratiques du lieu et les habitudes qui y sont liées,
  - une caractéristique propre à l'endroit,
  - le profil du locuteur / le statut du communicant / l'origine de la source d'information
  - l'objectif de l'act de parole



---

**MERCI !!!**

