



# RTR DIAMS GT RITUEL

Recherche d'Information,  
Traitements Utiles et  
Éclairés du Langage



# RITUEL : Objectifs

## **Réseautage scientifique...**

- ⇒ Des échanges des échanges des échanges....
- ⇒ Mutualisation des expériences au sein de la communauté scientifique
- ⇒ Diffusion des connaissances à destination des entreprises du numériques

## **... et plus si affinités**

- ⇒ Projets collaboratifs de recherche
  - ⇒ Méthodes pour répondre aux besoins des industriels et recherche de supports pour mettre en oeuvre des solutions
-

# RITUEL : OBJECTIFS

⇒ Echanges autour du Traitement Automatique des Langues Naturelles

## **RI : Recherche d'information**

⇒ Accroche auprès des entreprises : intérêt pour le TAL mais pas de compétences

⇒ Diffusion des connaissances à destination des entreprises du numériques

**TU : Traitements Utiles** Visée applicative claire

## **EL : et Eclaires du Langage**

⇒ Importance de l'interprétabilité des modèles (entreprise comme recherche)

⇒ Ancrage linguistique

⇒ Concerne également l'apprentissage profond

---

# Membres

## LIFO

- Anaïs Lefeuvre-Halftermeyer
- Billot Sylvie
- Ilaine Wang
- Gaëtan Caillaut

## LLL

- Anne-lyse Minard
- Emmanuel Schang
- Flora Badin
- Lotfi Abouda
- Marie Skrovec
- Helene Flamein

## LIFAT

- Jean Yves Antoine
- Agata Savary
- Denis Maurel
- Ilaine Wang
- Caroline Pasquer
- Mehdi Mirzapour

## Autres

- Marie Sophie Pausé
- Loïc Grobol (LATTICE)

## Non académiques

- Qual'Net. :
  - Marie Pommier
  - Stéphane Moinard
- BRGM
  - Claire Caillaut
  - Christelle Loiselet
  - Olivier Morel
- Aktan
  - Eric Clairambault
  - Gabrielle Brossard
- Kerpape
  - Mathieu Thebaud
  - Willy Allegre
- Garches
  - Samuel Pouplin



# Thématique : entités polylexicales

## Entités polylexicales [Pasquer 2019]

- Mots composés de plusieurs termes *avoir du pain sur la planche, court-circuiter, contrebraquer*
- Idiosyncrasie : non compositionnalité

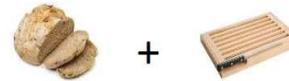
*faire valoir*<sub>Nom</sub>

↔

*faire*<sub>Verbe</sub> *valoir*<sub>Verbe</sub>

*avoir du pain sur la planche*

↔



- Très présentes dans les textes et aucune ressource descriptive complète : détection / identification

## Applications

- Traduction automatique *avoir un chat dans la gorge* ↔ *to have a frog in one's throat*
- Recherche d'information *champ libre* ≠



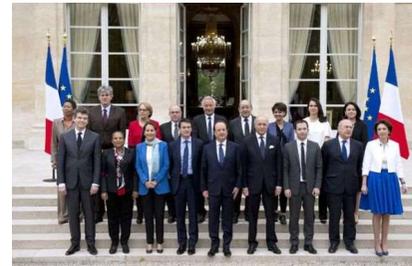
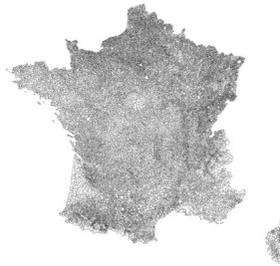
# Thématique : entités nommées

- Mots désignant une entité référentielle unique dans l'univers du discours : personnes, lieux, productions humaines, descriptions finies (montants, fonctions...)

- Entités polylexicales et imbrication *le [président du [conseil général du [département du Loiret]]]*

- Ambiguïté

France ?



## Applications

- Anonymisation
- Recherche d'information
- Terminologie scientifique

*La France a éliminé le Brésil*  
*phénylalanine hydroxylase*

# Thématique : coréférence

- Chaîne de coréférence : ensemble de termes désignant le même élément du discours
- Différents types :

<b>Coréférence directe</b>	... <i>cette équipe</i> ...
<b>Coréférence indirecte</b>	... <i>La Nationalmannschaft</i> ...
<b>Coréférence pronominale</b>	... <i>l</i> ...
<b>Anaphore associative</b>	... <i>leur entraîneur</i> ...
<b>Ana. associative pronominale</b>	... <i>ils</i> ... (joueurs de l'équipe)

En gagnant pour la quatrième fois la Coupe du monde en 2014, *l'équipe d'Allemagne* égale le nombre de victoires de l'Italie, mais reste toujours derrière le Brésil qui a conquis 5 Coupes du monde. *La RFA* ou *l'Allemagne* ~~co-détient~~ par ailleurs le record de victoires au championnat d'Europe avec l'Espagne qui ~~l'a rejointe en 2012~~. *La Nationalmannschaft* en juillet 2014 gagne pour la quatrième fois de suite une demi-finale de la Coupe du monde.

## Applications

- Traduction automatique

Les policiers et les juges des cours pénales devraient dire qu'il s'agit là d'une ordonnance de la cour et qu'elle doit être respectée.

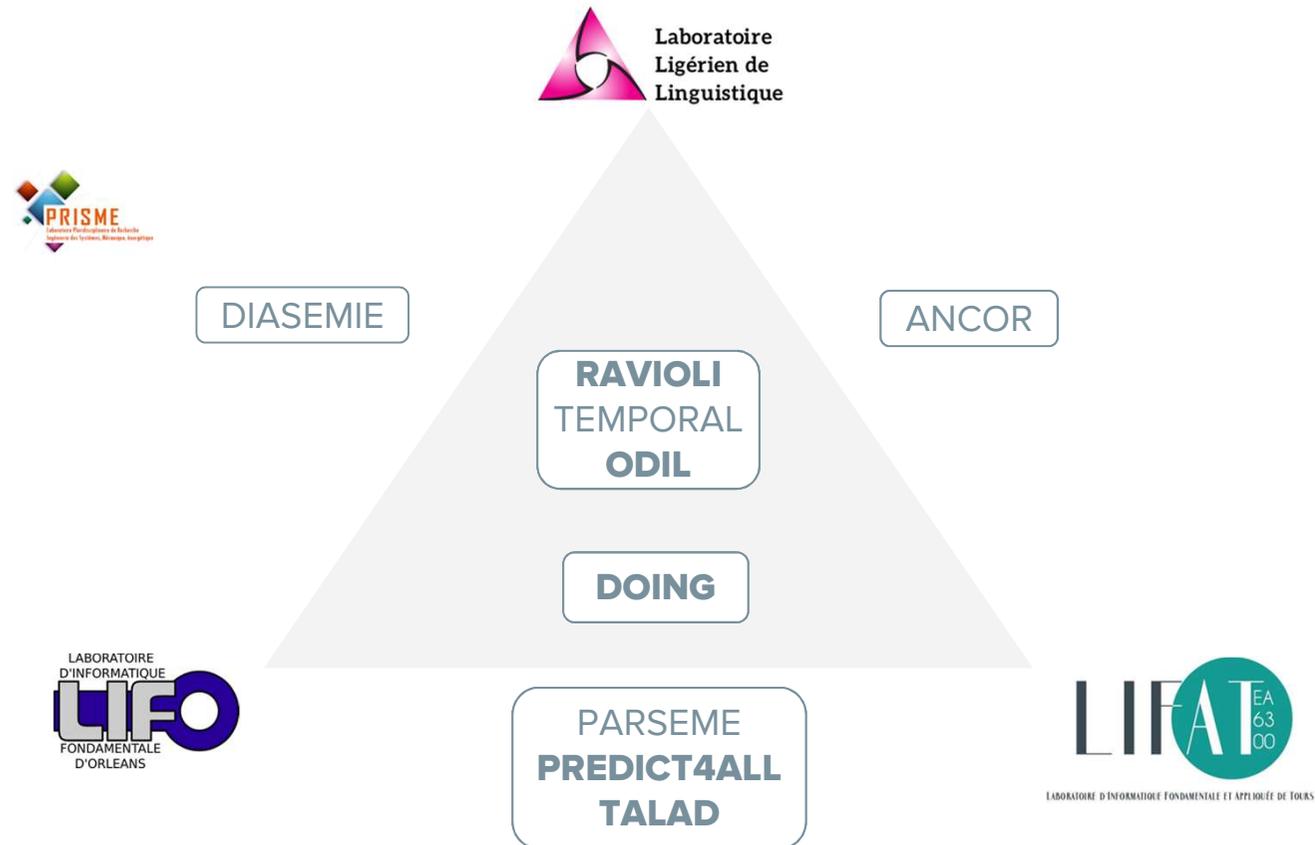
↳ [www2.parl.gc.ca](http://www2.parl.gc.ca)

The police and the criminal courts should say that this is an order of the court and it must be respected.

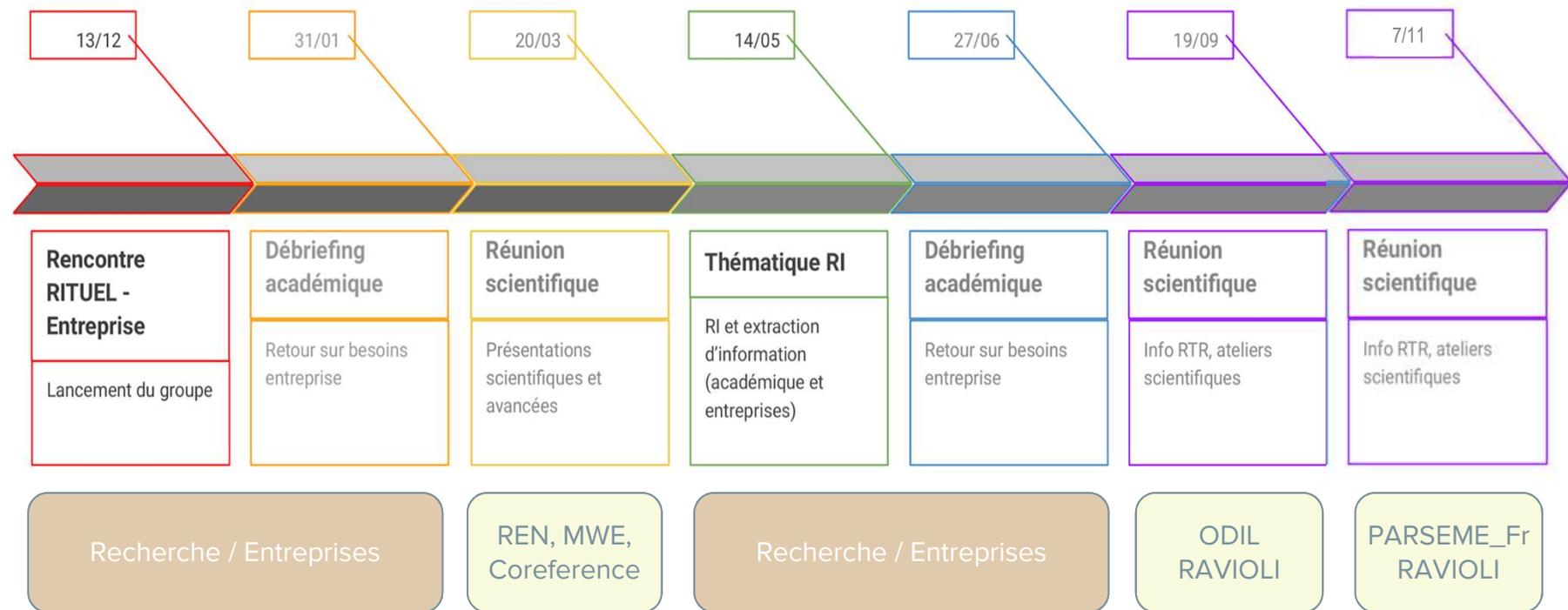
↳ [www2.parl.gc.ca](http://www2.parl.gc.ca)

- Recherche d'information *Emmanuel Macron a reconstruit Donald Trump. Il a déclaré...*
- Analyse du discours : nomination *Protectionnisme ... patriotisme économique*

# Ecosystème académique : collaborations



# Actions 2019 : réunions de travail



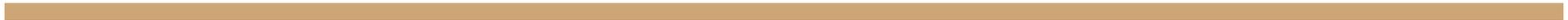
# Actions 2019 : entreprises

## Activité de conseil

- **Qual'Net** – Recherche d'information (Lucene) et correction orthographique
- **BRGM** – langues créoles dans le cadre du projet SURICATE (E. Schang)

## Collaborations

- **BRGM-LLL (stage Master SdL)** : analyse de formulaires de saisie
- **Aktan-LIFO-LIFAT (montage thèse CIFRE)** : apprentissage avec peu de données, repérage entités et relations sémantiques



# Actions 2019 : collaborations académiques

- **ANR AAC4All** – aide à la communication pour personnes handicapées (LIFAT + LIFO, Kerpape...) 
- **Thèse LIFAT/LLL/U. Dusseldorf** – Meta-grammaires et entités polylexicales 
- **Stages**
  - Evaluation des systèmes de résolution de coréférence (LIFO/LIFAT/LATTICE)
  - Génération par règles de corpus d'apprentissage pour la reconnaissance des Entités Nommées(LIFAT/LIFO)
  - Extractions de relations entre entités dans le domaine médical (LIFO/LLL/LIFAT)

DOING

---

# Actions 2020 : thème évaluation

**Journée scientifique sur le thème de l'évaluation** – 12/12/2019 LIFO Orléans

- Quelles sont les attentes en termes d'évaluation chez les industriels : mesures de qualité ?
  - Quelles sont les pratiques d'évaluation en recherche TAL ?
  - Quelles sont les limites méthodologiques des pratiques d'évaluation en recherche TAL
  - Evaluation technologique du TAL et évaluation ergonomique : quelles relations ?
-