

# De CALOR-QUEST à CALOR-DIAL

Elie Antoine, Jeremy Auguste, Cindy Aloui, Frédéric Béchet, Delphine Charlet,  
Géraldine Damnati, Laure Dupont, Benoît Favre, Clara Grégoire, Johannes  
Heinecke, Frédéric Herlédan, Gabriel Marzinotto, Alexis Nasr, Ludivine Robert,  
Lina Maria Rojas Barahona, Zakaria Taouil

Aix Marseille Université, Orange Labs

Lannion 15 mars 2022

# de CALOR-FRAME à CALOR-DIAL

Un peu d'histoire ...

- **CALOR-FRAME-1**

- ▶ annotation de corpus en cadre sémantique pour l'extraction d'information

- **CALOR-FRAME-2**

- ▶ extension du corpus et annotations en co-référence en complément des frames

- **CALOR-QUEST**

- ▶ utilisation de CALOR-FRAME dans un cadre applicatif
- ▶ **la compréhension automatique de documents** (*Machine Reading Comprehension* - MRQA)
- ▶ principe : utiliser les annotations en frame pour générer des questions sur lesquelles entraîner un système de MRQA

- **CALOR-DIAL**

- ▶ annotation et génération de questions multi-hop de type *dialogue*

## Etudes prévues

- **Tâche 1** : *Etude sur les capacités de généralisation cross-domaine d'un modèle MRQA*
  - ① *Etude sur la robustesse et la couverture sémantique des modèles sur le corpus **Autogestion** du projet **Archival***
  - ② *Etude sur la robustesse aux changements de types de questions : questions de type **fact checking** et **FAQ***
- **Tâche 2** : **CALOR-DIAL**, *un corpus de dialogue sur les textes encyclopédiques*
  - ① **Multi-Hop QA** : *générer et collecter des questions nécessitant l'examen de plusieurs segments d'un même document*
  - ② **Conversational QA** : *rajout de la notion de **contexte de dialogue** sur des séquences enchaînées de questions*

# Tâche 1 : Etude sur les capacités de généralisation cross-domaine d'un modèle MRQA

## Contexte double

- 1 Analyse sémantique de surface (Cadres et rôles sémantiques)
  - ▶ études sur l'utilité des représentations sémantiques explicites à l'heure du tout *end-to-end*
  - ▶ *le end-to-end est-il la fin des représentations explicites ?*
  - ▶ études sur la robustesse et l'adaptation des analyseurs sémantiques à des changements de domaine
  - ▶ *est-il possible d'avoir un analyseur sémantique **générique** suffisamment robuste ?*
- 2 Compréhension de documents et génération de question à partir de textes
  - ▶ tâches *miroirs* du Traitement Automatique des Langues traitées par des modèles *end-to-end*
  - ▶ *l'annotation sémantique explicite est-elle utile pour adapter ces modèles à de nouveaux domaines ?*
  - ▶ limites des méthodes actuelles
  - ▶ *comment se comportent les modèles état-de-l'art dans des contextes réalistes avec des questions plus abstraites, portant sur plusieurs parties d'un document, et s'enchaînant ?*

# Compréhension de documents et génération de questions

- Processus *miroirs*

- ▶ document + question → réponse *Machine Reading QA* (MRQA)
- ▶ document + réponse → question *Question Generation* (QG)

- Méthodes

- ▶ avant l'ère du *deep learning*
  - ★ *pipelines* d'outils linguistiques : **analyse syntaxique** pour QG, **recherche d'information** pour MRQA
- ▶ depuis que les *Transformers* ont envahi le monde
  - ★ approches *end-to-end* : modèle de type **BERT** pour MRQA et **BART** pour QG

# Compréhension de documents et génération de questions

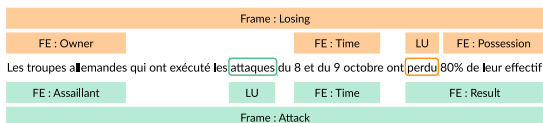
- Evaluation

- ▶ Compréhension de documents (MRQA)
  - ★ taux de réponses identiques avec la référence (**Exact Match** - EM)
  - ★ taux de mots en commun entre la prédiction et la référence (**Précision/Rappel/F-mesure**)
- ▶ Génération de questions (QG)
  - ★ ressemblance avec des questions de référence (**BLEU**)
  - ★ par rapport à une tâche (ex : MRQA)

# Compréhension de documents et génération de questions

- Limites des approches **end-to-end**
  - ▶ nécessité de disposer de corpus important d'exemples d'*entrée/sortie*
  - ▶ annotation manuelle difficile et couteuse, impossible à réaliser pour chaque nouveau cadre applicatif
- Adaptation non-supervisée d'un modèle de MRQA
  - ▶ générer des couples (*question, réponse*) sur le corpus cible
  - ▶ réapprendre un modèle de MRQA sur les questions générées
- Problématique : **Comment générer des couples (*question, réponse*)?**
  - ▶ à base d'annotation syntaxique/sémantique ?
  - ▶ uniquement à partir du texte ?
  - ▶ en mélangeant les deux ?

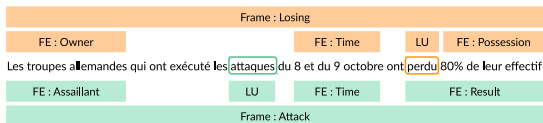
# Un monde parfait CALOR-Frame / CALOR-Quest



- 1 Sélection d'un ensemble de 53 *Frames* représentant des sens fréquents
- 2 Annotation manuelle du corpus **CALOR** avec ces *Frames*
- 3 Génération de couples (*question, réponse*) à partir de patrons sur les *Frames*
- 4 Apprentissage d'un modèle de MRQA sur le corpus CALOR et les questions/réponses générées
- 5 Collectes de questions *naturelles* portant sur les mêmes relations sémantiques que les *Frames*
- 6 Evaluation du modèle de MRQA sur ces questions *naturelles*



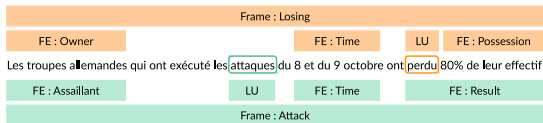
# Production de questions à partir de l'analyse en frame



## ● Méthode

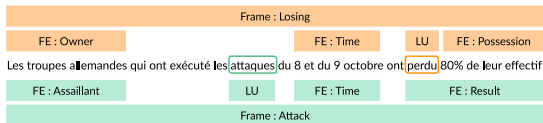
- ▶ sélection d'une *Frame*  $F$  et un *Frame Element*  $E$  dans une phrase → production de questions dont la réponse est  $E$
- ▶ la frame  $F$  et les FE constituent le contexte  $C$  de ces questions :  $(F, E, C)$
- ▶ en faisant varier  $F$ ,  $E$  et  $C$  pour une même phrase, on peut obtenir un ensemble de questions pour lesquelles on dispose des réponses avec leur classe sémantique (le type du FE)

# Production de questions à partir de l'analyse en frame



- Par exemple, 18 combinaisons possibles (F,E,C) sur la phrase précédente :
  - ▶ (Losing, Owner, {Time, Possession}),
  - ▶ (Losing, Owner, {Possession}),
  - ▶ (Losing, Owner, {Time}),
  - ▶ (Losing, Time, {Owner, Possession}),
  - ▶ ...

# Production de questions à partir de l'analyse en frame



- Chaque triplet (F,E,C) correspond à de nombreuses questions Q.
  - ▶ Par exemple pour  $Q(\text{Losing}, \text{Owner}, \{\text{Time}, \text{Possession}\})$  :
    - ★ *Qui a perdu 80% de ses effectifs du 8 au 9 octobre ?*
    - ★ *Quelles troupes ont été décimées à 80% dans les attaques du 8 et 9 octobre ?*
  - ▶ Avec pour réponse le même segment de texte : les troupes allemandes

# Production de questions à partir de patrons

- Patrons génériques et spécifiques écrits à la main pour les frames de CALOR-FRAME

Activity_start	spécifique	Qu'est-ce qui a commencé \$Time [ \$Place ] [ pour \$Purpose ] ? → \$Activity .
	générique	quoi commencer [ \$Agent ] [ \$Circumstances ] [ \$Co-timed_event ] [ \$Containing_event ] [ \$Event_description ] [ \$Explanation ] [ \$Manner ] [ \$Means ] [ \$Place ] [ \$Purpose ] [ \$Time ] ? → \$Activity .
Leadership	spécifique	Quand est-ce que \$Leader a dirigé \$Governed [ \$Place ] ? → \$Time .
	générique	quand diriger [ \$Leader ] [ \$Governed ] [ \$Place ] [ \$Role ] [ \$Duration ] [ \$Activity ] ? → \$Time .

- Exemple pour les Frames Activity\_start et Leadership

# Collecter des questions naturelles

Exemple présentés aux annotateurs :

**Frame** : Hiding\_objects

## Contexte

- **Agent** : un chef de milice gauloise
- **Hidden\_object** : un trésor
- **Hiding\_place** : dans sa ferme de Bassing

## But

- objet de la question : Place
- réponse : Moselle

## Questions collectées

- *Dans quelle région un chef de milice gauloise cache-t-il un trésor ?*
- *Où se trouve la ferme de Bassing, dans laquelle un chef de milice gauloise cache un trésor ?*

# Le corpus *CALOR-QUEST*

Le corpus *CALOR-QUEST* est composé d'un ensemble de questions *naturelles*  $Q_N$  et d'un ensemble de questions *automatiques*  $Q_A$  portant sur des textes du corpus *CALOR-Frame*. A chaque question est associé un segment de texte représentant la réponse à la question.

collection	nb docs	nb questions naturelles ( $Q_N$ )	nb moyen $Q_N$ par doc	nb questions automatiques ( $Q_A$ )	nb moyen $Q_A$ par doc	nb moyen appariements $Q_A/Q_N$
V_antiq	61	274	4,5	4 672	76,6	4,2
archblog	42	250	6,0		214,4	5,8
WP_arch	96	302	2,4	36 259	377,7	4,1
CT_1GM	16	241	15,1	7 502	468,9	2,5
WP_1GM	123	319	2,6	50 971	414,4	5,1
<b>total</b>	<b>296</b>	<b>1 136</b>	<b>3,8</b>	<b>99 404</b>	<b>335,8</b>	<b>4,1</b>
picasso	7	198	28,3		652,6	4,4
prehistoire	42	182	4,3		75,3	2,9
savanturier	67	237	3,5		61,7	4,3

# Evaluation

## Entraîner le modèle MRQA sur CALOR-QUEST

- Apprentissage
  - ▶ Découpe du corpus CALOR en paragraphes d'environ 120 tokens
  - ▶ Apprentissage sur un ensemble de 14K questions générées avec CAMEMBERT
- Evaluation

Corpus	Exact-Match	F-mesure
CALOR-Quest	67.4	78.6
SQuAD ( <i>en</i> )	80.8	88.1
FQuAD ( <i>fr</i> )	66.1	77.6

(pour SQuAD et FQuAD, modèle Multilingual BERT, résultats obtenus en interne)

# Retour à la réalité : ARCHIVAL (CRE2021)

- **ARCHIVAL**

- ▶ Projet ANR (FMSH, Orange, IRISA, AMU)
- ▶ Thème : Valorisation d'archives multimédia
- ▶ Corpus : fond *autogestion* de la FMSH

- Compréhension / Mise en relation

- ▶ Question-Réponse comme paradigme de la mise en relations de *documents*
- ▶ Principe : **un document A est en relation avec un document B si les questions générées à partir de A trouvent une réponse dans B**

- Corpus Archival comme terrain d'expérimentation pour les études cross-domaine

- ▶ annotation en Frame
- ▶ MRQA : validation de la méthodologie CALOR-QUEST



# Défis posés par le corpus ARCHIVAL

- Différences par rapport à CALOR

- ▶ Corpus linguistiquement plus *difficile* que Wikipedia
- ▶ Pas d'annotation manuelle en *Frames*
  - ★ annotation automatique → utilisation d'**ETANOL**
- ▶ Questions *réalistes*, non limitées au modèle de *Frames*
  - ★ nécessité d'utiliser un modèle d'annotation sémantique large
  - ★ difficile d'avoir des patrons de générations de questions pour tout les cas

# Défis posés par le corpus ARCHIVAL

- Problèmes/Questions

- ▶ Comment se comporte un modèle appris sur Wikipedia sur des textes *difficiles*?
- ▶ Quelles sont les performances d'un modèle d'annotation sémantique sur ces textes?
- ▶ Comment générer des questions basées sur l'annotation sémantique en l'absence de patrons?

# Défis posés par le corpus ARCHIVAL

- Problèmes/Questions

- ▶ Comment se comporte un modèle appris sur Wikipedia sur des textes *difficiles*?
- ▶ Quelles sont les performances d'un modèle d'annotation sémantique sur ces textes ?
- ▶ Comment générer des questions basées sur l'annotation sémantique en l'absence de patrons ?

Essayons la *magie*!!

→ les modèles **séquence-à-séquence**!!

# Les modèles séquences à séquences pour la génération de question

- Principe : *traduction d'une phrase affirmative en phrase interrogative*
  - ▶ possibilité d'ajouter dans la séquence d'entrée des éléments *conditionnant* la production de la séquence de sortie
    - ★ réponse attendue
    - ★ analyse sémantique
  - ▶ modèles de langage préentraînés permettant la génération après *fine tuning* (BART, BARthez)
  - ▶ possibilité d'utiliser les corpus de MRQA (SQuAD, FQuAD) pour *fine tuner* ces modèles
- Problèmes
  - ▶ Comment sélectionner les phrases sur lesquelles des questions vont être générées ?
  - ▶ Comment sélectionner les éléments dans ces phrases sur lesquels on aimerait poser des questions ?
  - ▶ Que mettre dans la séquence d'entrée pour *conditionner* la question en sortie ?

# Méthode

- 1 Annotation automatique avec **ETANOL** (FrameNet, SRL) des contextes et des questions du corpus **FQuAD**
- 2 Pour chaque triplet question/réponse/contexte :
  - 1 Trouver un rôle sémantique associé à la réponse d'une question donnée grâce à l'annotation effectuée
  - 2 Générer un exemple d'entraînement avec une séquence d'entrée contenant le contexte, la réponse et l'annotation sémantique, et la question comme séquence de sortie
- 3 Affinage d'un modèle de génération pré-entraîné *BARTez* sur le corpus obtenu

Travail réalisé par [Jeremy Auguste](#) et [Elie Antoine](#)

# Méthode

## Quelle représentation pour la séquence d'entrée ?

- 3 représentations
  - 1 **Rep+Ctx** : seuls la réponse (l'empan du rôle sémantique détecté) et le contexte sont donnés
  - 2 **Rep+Ctx+SRL** : le rôle de la réponse, le déclencheur et les autres rôles sémantiques sont explicitement ajoutés suivant un schéma SRL (*PropBank*)
  - 3 **Rep+Ctx+FrameNet** : étiquettes des éléments de cadre de FrameNet pour caractériser la réponse et le contexte
- Exemples
- Exemples

# Evaluation sur ARCHIVAL

## Annotation en questions du corpus ARCHIVAL selon 3 niveaux de difficulté

- Question *facile* (niveau 1)
  - ▶ *Contexte* : « Pour l'heure, l'autogestion s'accompagne souvent d'une certaine étroitesse de vue. »
  - ▶ *Question* : De quoi l'autogestion s'accompagne-t-elle souvent ?
- Question *moyenne* (niveau 2)
  - ▶ *Contexte* : « Le marxisme a accouché tout aussi bien du réformisme que du léninisme et du stalinisme (sans parler du maoïsme). »
  - ▶ *Question* : Quelles ont été les conséquences du marxisme ?
- Question (niveau 3)
  - ▶ *Contexte* : « Curieuse découverte en réalité que celle-ci. Aucun élément « scientifique » ne vient étayer la religion nouvelle de Bakounine ! »
  - ▶ *Question* : À quoi est comparé le communisme ?

# Résultats

Performances du modèle appris uniquement sur les questions générées sur la tâche de question/réponse selon le niveau de difficulté et la représentation en entrée des contexte permettant de générer les questions du corpus *Autogestion*

Modèle <i>M1</i>	Niveau 1 (416 questions)				Niveau 2 (352 questions)				Niveau 3 (74 questions)			
	EM	F1	Prec	Rappel	EM	F1	Prec	Rappel	EM	F1	Prec	Rappel
Rep+Ctx	<b>27.88</b>	44.58	<b>57.71</b>	36.32	<b>16.19</b>	<b>34.87</b>	<b>45.32</b>	<b>28.33</b>	5.41	<b>24.82</b>	<b>27.29</b>	<b>22.75</b>
Rep+Ctx+SRL	25.96	45.30	57.02	37.58	14.77	31.99	41.11	26.18	<b>8.11</b>	22.04	24.42	20.08
Rep+Ctx+FrameNet	20.91	<b>43.17</b>	49.20	<b>38.46</b>	11.93	31.81	38.53	27.08	6.76	17.37	18.42	16.44



# Résultats

Avec l'ajout du corpus FQuAD

Modèle <i>M2</i>	Niveau 1 (416 questions)				Niveau 2 (352 questions)				Niveau 3 (74 questions)			
	EM	F1	Precision	Recall	EM	F1	Precision	Recall	EM	F1	Precision	Recall
FQuAD	38.46	63.57	<b>81.39</b>	52.15	24.72	<b>54.76</b>	<b>70.37</b>	44.82	18.92	<b>52.93</b>	<b>66.47</b>	43.98
FQ+Rep+Ctx	42.55	63.61	79.93	52.82	<b>29.26</b>	53.00	70.32	42.52	<b>22.97</b>	42.22	51.96	35.56
FQ+Rep+Ctx+SRL	<b>43.27</b>	<b>66.46</b>	79.49	<b>57.10</b>	26.70	50.76	58.30	44.94	21.62	39.17	40.79	37.67
FQ+Rep+Ctx+FrameNet	41.83	63.94	79.77	53.36	28.98	54.42	64.35	<b>47.15</b>	<b>22.97</b>	50.15	52.06	<b>48.37</b>

# Discussions

- “*Proof Of Concept*” sur CALOR-QUEST
  - ▶ ça marche!!!! dans des conditions (très) contrôlées
  - ▶ contrôle complet de la sémantique → peu réaliste pour le passage à l'échelle
- Application à la “*vie réelle*”
  - ▶ résultats mitigés sur *ARCHIVAL*
  - ▶ assez bonne généralité de modèles *généralistes* (**FQuAD**) sur les questions simples
  - ▶ gains (légers) grâce à l'adaptation par génération de questions

# Discussions

- Mais en fait ...
  - ▶ MRQA et QG sont appris sur les mêmes corpus
  - ▶ méthode d'adaptation par génération de question → [autosupervision](#) ?
  - ▶ sur quelles informations nouvelles le processus d'adaptation peut-il se baser ?
    - [les annotations sémantiques](#)
      - ★ mais sont-elles fiables ?
      - ★ est-ce qu'on les utilise correctement ?

# Et maintenant ?

- **Perspective 1**

- ▶ renforcer l'utilisation des annotations sémantiques pour se rapprocher de *CALOR-QUEST*
  - ★ contrôle sémantique des questions générées et des réponses obtenues
  - ★ changer de modèle sémantique ? **AMR** ?

- **Perspective 2**

- ▶ **Vers une magie plus sombre ?**
  - ★ *prompting models*

# Exemples avec modèle T0pp (Huggingface)

- Context

- ▶ *Peu de jours avant sa mort , au moment où il venait d' accepter de participer à la rédaction de ces Cahiers , prenant la parole à Bruxelles au Colloque organisé en l' honneur de Proudhon , Georges Gurvitch rappelait avec netteté le rôle international joué par le syndicalisme révolutionnaire et son essentielle revendication — l' autogestion ouvrière .*
- ▶ Task: Generate a question
  - ★ What was the role of syndicalism in the revolutions of 1848?
- ▶ Task: Generate a question on a person
  - ★ What was the name of the person who was a major influence on Gurvitch?
- ▶ Task: Generate a question on an event
  - ★ What was the name of the event that Georges Gurvitch spoke at?
- ▶ Task: Generate a question on a place
  - ★ Where did Gurvitch speak a few days before his death?
- ▶ Task: Generate a question on "Proudhon"
  - ★ What was the name of the Syndicalist leader who was a major influence on Gurvitch?

## Etudes réalisées

- **Tâche 1** : *Etude sur les capacités de généralisation cross-domaine d'un modèle MRQA*
  - ① *Etude sur la robustesse et la couverture sémantique des modèles sur le corpus **Autogestion** du projet **Archival***
  - ② *Etude sur la robustesse aux changements de types de questions : questions de type **fact checking** et **FAQ***
- **Tâche 2** : **CALOR-DIAL**, *un corpus de dialogue sur des questions portant sur des textes encyclopédiques*
  - ① **Multi-Hop QA** : *générer et collecter des questions nécessitant l'examen de plusieurs segments d'un même document*
  - ② **Conversational QA** : *rajout de la notion de **contexte de dialogue** sur des séquences enchaînées de questions*

## Tâche 2 : **CALOR-DIAL**, *un corpus de dialogue sur des questions portant sur des textes encyclopédiques*

### Réalisation

- création d'une interface d'annotation *multihop/conversational QA*
- annotation manuelle du corpus **CALOR**
  - ▶ 2 annotatrices : Clara Grégoire / Ludivine Robert
  - ▶ rédaction d'un guide d'annotation
- annotation automatique à partir de **CALOR-Frame**+coréférence
- premières expériences MRQA sur **CALOR-DIAL**

# Interface d'annotation *multihop/conversational* QA

Pamiat Merkuria (croiseur, 1902)

Le Souvenir du Mercure ou Mémoire du Mercure ( en russe : Память Меркурия , Pamiat Merkuria ) est un croiseur protégé de la classe Bogatyr de la Marine Impériale russe , construit en 1901 - 1902 .

Première Guerre mondiale

Pamiat Merkuria (croiseur, 1902)

Le Souvenir du Mercure ou Mémoire du Mercure ( en russe : Память Меркурия , Pamiat Merkuria ) est un croiseur protégé de la classe Bogatyr de la Marine Impériale russe , construit en 1901 - 1902 .

Ses sister-ships sont l' Otchakov et l' Oleg .

Histoire

Le ministère de la Marine ordonne sa construction qui se déroule en 1901 - 1902 au chantier naval de Nikoliev sur la mer Noire .

**Il est mis à l'eau le 20 mai 1902 et armé en 1905 sous le nom de Kagoul .**

**Cependant à cause de la confusion avec le navire du même nom ( l'ancien Otchakov ) rebaptisé Kagoul en 1907 ( à cause des mauvais souvenirs de la révolution de 1905 ) le navire prend le nom de Pamiat Merkuria ( Souvenir du Mercure ) , le même année .**

Il est à quai à Sebastopol du 6 janvier 1913 au 1er mai 1914 pour des travaux de fond .

Service pendant la Première Guerre mondiale

Le Souvenir du Mercure en 1914

Le croiseur navigue en mer Noire pour la défense des côtes et participe au blocus de la Turquie , alliée de l' Empire allemand .

C est un escortant cinq pré- Dreadnoughts au large du cap Sarytch que le Souvenir du Mercure rencontre les croiseurs allemands SMS Goeben ( sous pavillon turc et renommé Yavuz Sultan ) et SMS Breslau ( sous pavillon turc et renommé Hiddir ) , le 18 novembre 1914 .

Ces derniers dévient rapidement de leur course après qu' une casemate ( 15 cm ) de SMS Goeben a été touchée , provoquant un début d' incendie .

Le Souvenir du Mercure et le Kagoul rencontrent de nouveau à deux reprises le SMS Breslau en janvier 1915 , accompagné du croiseur turc Hariride , mais ces derniers s' échappent et il n' y a aucun dommage de part et d' autre .

La flotte de la mer Noire bombarde , le 10 mai 1915 , les forteresses turco - allemandes du Bosphore .

Le Souvenir du Mercure est repéré par le SMS Goeben ( sous pavillon turc ) de retour d' une patrouille au large d' Eregli à 115 miles ( 185 km ) à l' est du Bosphore .

Le croiseur allemand se lance à ses trousseaux , tandis que le croiseur russe se tourne à toute vitesse pour se diriger vers lui et atteindre le corps du bâtiment .

Les pré- Dreadnoughts russes atteignent le navire allemand à trois reprises qui se dégage de la bataille profitant de sa vitesse plus importante .

Le croiseur russe n' est pas atteint .

Lorsque le Souvenir du Mercure est arrêté pour remise en état , de décembre 1916 à avril 1917 , ( après la révolution de Février ) , il est équipé de seize canons de 130 mm , au lieu des six précédents de 152 mm .

Le navire est posé en face de Constanza en Roumanie le 1er novembre 1917 , pour détruire le dépôt d' essence du port , qui était en possession des Allemands qui occupaient la ville .

Toutefois une fausse alerte sous - marine oblige le Souvenir du Mercure à partir sans même avoir tiré .

Il retourne sur place trois jours plus tard et cette fois - ci détruit quinze des trente-sept réservoirs .

Guerre civile

Vue du Souvenir du Mercure

Question N°0 - Qu'est-ce que le Pamiat Merkuria ?

Question(dialogue)	Qu'est-ce que le Pamiat Merkuria ?
Question(hors contexte)	
Type question	paraphrase-simple-multihop
Réponse courte	Un navire de type Croiseur.
Réponse avec contexte	Le Pamiat Merkuria est un navire de type Croiseur.
Support 0	TITRE : Pamiat Merkuria
Support 1	SPAN : un croiseur protégé de la classe Bogatyr de la Marine Impériale russe
Support 2	SPAN : le navire prend le nom de Pamiat Merkuria
Support 3	Le Souvenir du Mercure ou Mémoire du Mercure ( en russe : Память Меркурия ,
Support 4	Il est mis à l'eau le 20 mai 1902 et armé en 1905 sous le nom de Kagoul .
Support 5	Cepeni

Question N°1 - Comment s'appelait-il initialement ?

Question(dialogue)	Comment s'appelait-il initialement ?
Question(hors contexte)	Comment s'appelait le Pamiat Merkuria initialement ?
Type question	paraphrase-co-reference-multihop
Réponse courte	Le Kagoul
Réponse avec contexte	Initialement le Pamiat Merkuria se nommait le Kagoul.
Support 0	TITRE : Histoire
Support 1	SPAN : sous le nom de Kagoul
Support 2	Il est mis à l'eau le 20 mai 1902 et armé en 1905 sous le nom de Kagoul .

Question N°2 - Pourquoi a-t-il changé de nom ?

Question N°3 - Quel autre navire a-t-il souvent croisé ?

Question N°4 - Quels sont les navires qui ont permis sa réparation ?



# Interface d'annotation *multihop/conversational* QA

## Question N° 1 - Comment s'appelait-il initialement ?

Question(dialogue)	Comment s'appelait-il initialement ?
Question(hors contexte)	Comment s'appelait le Pamiat Merkouria initialement ?
Type question	paraphrase-coreference-multihop
Réponse courte	Le Kagoul
Réponse avec contexte	Initialement le Pamiat Merkouria se nommait le Kagoul.
Support 0	TITRE : Histoire
Support 1	SPAN : sous le nom de Kagoul
Support 2	Il est mis à l' eau le 20 mai 1902 et armé en 1905 sous le nom

# Guide d'annotation

L'annotation d'un document consiste à y identifier :

- des questions *dialogue* : formulation minimum de la question qui suit une logique conversationnelle
  - ▶ Par qui ?
- des questions *hors contexte* : formulation et contextualisation complète de la question
  - ▶ Qui a créé le port d'Empùries en -580 ?
- des types de questions :
  - ▶ `litteral` vs `paraphrase`
    - ★ est-ce que la question reprend la verbalisation ou est-ce une paraphrase ?
  - ▶ `ellipse` vs `simple` vs `coréférence`
    - ★ est-ce qu'il y a une coréférence ou une ellipse dans la question ?
  - ▶ `simple` vs `multihop`
    - ★ faut-il accéder à plusieurs parties du texte pour répondre à la question ?
    - ★ faut-il résoudre des coréférences dans le texte ?
    - ★ `litteral-ellipse-multihop`

# Guide d'annotation

- des réponses courtes : entité nommée ou syntagme qui répond à la question
  - ▶ Par des colons grecs phocéens
- des réponses avec contexte : verbalisation et contextualisation complète de la réponse ;
  - ▶ Le port d'Empuries a été créé par des colons grecs phocéens
- des titres : identifier un point de repère dans le texte / début de paragraphe qui va supporter la réponse
  - ▶ TITRE : Histoire et archéologie
- des spans : encadrer précisément la réponse si possible (plusieurs spans pour les réponses multi hop)
  - ▶ SPAN : par des colons grecs phocéens
- des supports : identifier largement l'endroit du texte qui permet d'inférer, justifier la réponse
  - ▶ Empúries est un port et une cité
  - ▶ La ville a été fondée vers -580 par des colons grecs phocéens

# Exemples

Question (dialogue)	En quelle année a été créé le port d' Empúries ?
Question (hors contexte)	
Type question	litteral-simple-multihop
Réponse courte	En -580.
Réponse avec contexte	Le port d'Empuries a été créé en -580.
Support 0	TITRE : Histoire et archéologie
Support 1	SPAN : vers -580
Support 2	Empúries est un port et une cité
Support 3	La ville a été fondée vers -580

# Exemples

Question (dialogue)	Qui est Vercingétorix ?
Question (hors contexte)	
Type question	litteral-simple-multihop
Réponse courte	Le plus connu des chefs gaulois et un des chefs de la résistance des Gaulois contre la conquête de la Gaule par Jules César.
Réponse avec contexte	Vercingétorix est le plus connu des chefs gaulois et un des chefs de la résistance des Gaulois contre la conquête de la Gaule par Jules César.
Support 0	TITRE : Vercingétorix
Support 1	SPAN : le plus connu des chefs gaulois
Support 2	SPAN : un des chefs de la résistance des Gaulois contre la conquête de la Gaule par Jules César
Support 3	Vercingétorix est le plus connu des chefs gaulois
Support 4	Il est un des chefs de la résistance des Gaulois contre la conquête de la Gaule par Jules César

# Exemples

Question (dialogue)	Quand est-il né ?
Question (hors contexte)	Quand est né Vercingétorix ?
Type question	litteral-coreference-multihop
Réponse courte	En 72 av. J.-C.
Réponse avec contexte	Vercingétorix est né en 72 av. J.-C.
Support 0	TITRE : Vercingétorix
Support 1	SPAN : en 72 av. J.-C.
Support 2	Vercingétorix
Support 3	Il est né en 72 av. J.-C.

# Exemples

Question (dialogue)	Où ?
Question (hors contexte)	Où est né Vercingétorix ?
Type question	litteral-ellipse-multihop
Réponse courte	En Auvergne.
Réponse avec contexte	Vercingétorix est né en Auvergne.
Support 0	TITRE : Vercingétorix
Support 1	SPAN : en Auvergne
Support 2	Vercingétorix
Support 3	Il est né en 72 av. J.-C. en Auvergne

# Le corpus CALOR-DIAL

- Description du corpus
  - ▶ Nb de dialogue réalisés : 234
  - ▶ Nb de questions annotées : 1663
  - ▶ Nb moyen de questions par fil : 7.3
- Répartition des tailles de dialogues :

taille dial	2	3	4	5	6	7	8	9	10	11	12	13	14	17
nb question	3	6	18	36	42	35	23	17	16	22	7	7	1	1



# Le corpus CALOR-DIAL

- Repartition des questions par type : litteral/paraphrase
  - ▶ litteral : 947 questions
  - ▶ paraphrase : 748 questions
- Repartition des questions par type : simple/coreference/ellipse
  - ▶ simple : 474 questions
  - ▶ coreference : 995 questions
  - ▶ ellipse : 226 questions
- Repartition des questions par type : simple/multihop
  - ▶ simple : 980 questions
  - ▶ multihop : 715 questions

# Le corpus CALOR-DIAL

- Repartition des questions avec combinaison des types :

litteral-simple-simple	249 questions
litteral-coreference-multihop	218 questions
litteral-simple-multihop	75 questions
litteral-ellipse-multihop	43 questions
litteral-coreference-simple	275 questions
paraphrase-ellipse-simple	56 questions
litteral-ellipse-simple	87 questions
paraphrase-coreference-simple	222 questions
paraphrase-simple-simple	91 questions
paraphrase-coreference-multihop	280 questions
paraphrase-ellipse-multihop	40 questions
paraphrase-simple-multihop	59 questions

# Le corpus CALOR-DIAL

- Repartition des questions par corpus :
  - ▶ prehistoire : 298 questions
  - ▶ antiquite : 332 questions
  - ▶ histclio : 103 questions
  - ▶ histwiki : 488 questions
  - ▶ archwiki : 497 questions

# Le corpus CALOR-DIAL

## Utilisations possibles

- MRQA *classique*
  - ▶ pas de contexte dialogique si on prend les questions hors-contexte
  - ▶ possibilité de définir 4 niveaux de difficulté
    - ① littéral-simple
    - ② paraphrase-simple
    - ③ simple-multihop
    - ④ paraphrase-multihop
- MRQA *conversationnelle*
  - ▶ enchaînement des questions avec coréférences et élipSES
  - ▶ 12 niveaux de difficultés !

# Utilisations possibles de CALOR-DIAL

- Génération de texte : passage des réponses courtes aux réponses avec contexte

Archéologue et préhistorien.	=> Édouard Piette était un archéologue et préhistorien.
Il aurait voulu être avocat.	=> Edouard Piette aurait voulu être avocat avant de s'intéresser à l'archéologie.
Des statuettes et fragments de statuettes de femmes.	=> Les types d'objet qu'il a découvert sont des statuettes et fragments de statuettes de femmes.
L'authenticité.	=> Edouard Piette a reconnu l'authenticité des peintures d'Altamira.
En -580.	=> Le port d'Empuries a été créé en -580.
Des colons grecs phocéens.	=> Le port d'Empùries a été créé par des colons grecs phocéens.
Marseille.	=> Les phocéens ont aussi fondé Marseille.
La Paléopolis.	=> La Paléopolis est le site archéologique quasi impossible à fouiller.

# Utilisations possibles de CALOR-DIAL

- Génération de texte : passage de questions avec coréférence/ellipse vers questions hors-contexte

Quels sont les types d'objets qu'il a découvert ?

=> Quels sont les types d'objets dont Édouard Piette a fait la découverte ?

Qu'a-t-il reconnu des peintures d'Altamira ?

=> Qu'a reconnu Edouard Piette des peintures d'Altamira ?

Par qui ?

=> Qui a créé le port d'Empùries en -580 ?

Quelle autre ville a été fondée par les Phocéens ?

=> Quelle autre ville qu'Empùries a été fondée par les Phocéens ?

# Annotation automatique en questions conversationnelles + multi-hop

- Génération de questions conversationnelles
  - ▶ à partir de **CALOR-Frame**+coréférence
  - ▶ méthode
    - ★ chaînes de coréférences reliant des annotations en *Frames*
    - ★ génération de questions mélangeant coréférences et mentions
- Génération de questions multi-hop
  - ▶ fusion des contextes et rôles sémantiques selon les chaînes de coréférence
  - ▶ génération de questions à partir des entrées fusionnées

# Premières expériences de MRQA

- Utilisation des questions *hors contexte*
- Résultats avec modèle MRQA appris sur **CALOR-QUEST**

Difficulté 1 (601 questions) :				Difficulté 2 (362 questions) :				Difficulté 3 (332 questions) :				Difficulté 4 (368 questions) :			
EM	F1	Precision	Recall	EM	F1	Precision	Recall	EM	F1	Precision	Recall	EM	F1	Precision	Recall
31.78	54.56	60.28	49.84	19.06	42.09	49.58	36.57	12.05	37.68	51.41	29.74	5.98	26.09	38.14	19.82

- Résultats avec modèle MRQA appris sur **FQuAD**

Difficulté 1 (601 questions) :				Difficulté 2 (362 questions) :				Difficulté 3 (332 questions) :				Difficulté 4 (368 questions) :			
EM	F1	Precision	Recall	EM	F1	Precision	Recall	EM	F1	Precision	Recall	EM	F1	Precision	Recall
33.94	70.12	81.71	61.41	27.62	59.84	74.69	49.91	10.54	47.68	62.94	38.38	5.71	36.43	49.65	28.77



# Premières expériences de MRQA

- Résultats sur l'application du modèle **CALOR-QUEST**
  - ▶ manque de généralité des *frames*
- Utilisation du modèle **FQuAD**
  - ▶ généralisation OK sur CALOR-DIAL pour les questions simples
  - ▶ *baseline* pour les questions multi-hop

# Conclusions et discussion

- Nombreuses pistes ouvertes !!
  - ▶ contrôle sémantique pour le MRQA et la génération de questions
  - ▶ prise en compte des coréférences et des ellipses
  - ▶ de l'extraction de réponse à la génération de réponses en multihop
  - ▶ changement dans le modèle sémantique (AMR ?)
  - ▶ ...