

Génération automatique de questions pour l'apprentissage de modèles de *Machine Reading Comprehension*

Jeremy Auguste

6 mai 2021

LIS, Aix-Marseille Université, CNRS

Point de départ

Question

Qui crée les nombreuses peintures sur les parois ?

Contexte

La grotte d'Altamira est située près de Santander, (Cantabrie) dans le Nord de l'Espagne. Elle est célèbre pour ses nombreuses peintures réalisées sur les parois par les Hommes de la Préhistoire. Les peintures ont été découvertes en 1879 par l'archéologue espagnol Marcelino Sanz de Sautuola.

Petit rappel sur le question-réponse par extraction

Question

Qui crée les nombreuses peintures sur les parois ?

Contexte

La grotte d'Altamira est située près de Santander, (Cantabrie) dans le Nord de l'Espagne. Elle est célèbre pour ses nombreuses peintures réalisées sur les parois **par les Hommes de la Préhistoire**. Les peintures ont été découvertes en 1879 par l'archéologue espagnol Marcelino Sanz de Sautuola.

Réponse

par les Hommes de la Préhistoire

Des corpus

- SQuAD (anglais - Wikipédia)
- FQuAD (français - Wikipédia)
- PIAF (français - Wikipédia)
- XQuAD, MLQA (multilingue - Wikipédia)

Corpus	Modèle	EM	F1
SQUAD 2.0	Humain	86.8	89.5
SQUAD 2.0	SA-Net on ALBERT	90.7	93.0
FQUAD 1.1	Humain	75.9	91.2
FQUAD 1.1	CAMEMBERB-large	82.1	92.2

Mais s'adaptent assez mal à d'autres corpus/domaines...

Les modèles appris sur un corpus généralisent mal à d'autre corpus :

Entraînement	Évaluation	EM	F1
FQUAD	FQUAD	69.8	85.5
FQUAD	PIAF	40.0	65.7
FQUAD	CALOR	29.2	61.7

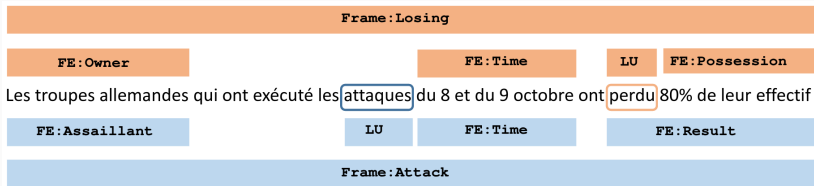
Problème

→ Comment construire un corpus d'apprentissage sur de nouvelles données ?

Approche proposée précédemment

Approches proposées par [Béchet et al., 2019] :

- Se fonder sur des annotations FrameNet (FN) pour produire des règles permettant de produire des questions.



Règle possible spécifique à la frame « Attack »

Quand est -ce que \$Assaillant a <attaqué> ?

→ \$Time

Approches proposées par [Béchet et al., 2019] :

- Se fonder sur des annotations FrameNet (FN) pour produire des règles permettant de produire des questions.

Résultats

Modèle affiné à partir de CAMEMBERT (camembert-large).
Évaluation sur les questions naturelles de CALOR.

Corpus train	Exact Match	F1
CALOR-FrameNet	71.10	83.02

Limites de l'approche fondées sur FrameNet :

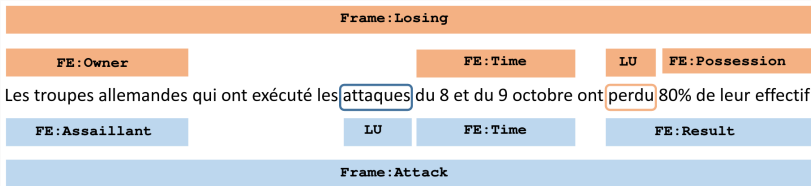
- Annotation spécifique aux données de CALOR (54 types de frames)
- Nombre de déclencheurs restreint (145 lemmes)
- Requiert la création de règles spécifiques pour chaque type de frame.

Proposition

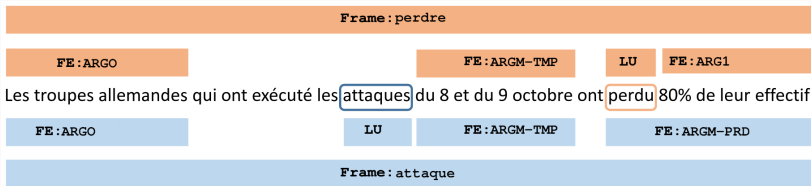
Remplacer l'utilisation de FrameNet par PropBank :

- Annotation plus simple car moins précise
- Peut couvrir davantage de thèmes

Passage de FrameNet à PropBank



FrameNet



PropBank

1. « Traduction » des annotations FrameNet en annotations PropBank sur CALOR.
2. Définition de règles pour transformer des annotations PropBank en questions.
 - Impossibilité de produire des règles aussi précises qu'avec FrameNet.
 - Certains choix (pronoms interrogatifs, conjugaison des verbes) deviennent encore plus difficiles à faire.
 - Questions souvent syntaxiquement peu correctes

Question

Comment se comporte un modèle de MRQA avec des questions issues de PropBank ? Équivalent à FrameNet ?

Exemples de questions générées

Question générée	Réponse
Qui cherche ces failles ? Context : « Voilà pourquoi Bertrand Meyer cherche ces failles »	Bertrand Meyer
Qui a été dit par certains chercheurs ? Context : « Certains chercheurs se disent prêts »	prêts
Qu'est-ce qui crée d'une mine ? Context : « d'une mine créée par les Royal Engineer. . . »	Royal Engineer. . .

Résultats en utilisant PropBank

Corpus train	Exact Match	F1
CALOR-FrameNet	71.10	83.02
CALOR-PropBank	61.90	75.77

- Perte importante sur les deux métriques.
- Il est nécessaire d'améliorer la qualité des questions générées avec PropBank.

Améliorer la génération des questions

Difficile de produire des règles à partir de PropBank :

- Utiliser à la place un modèle séquence-à-séquence
- Existence de modèles préentraînés pour la traduction ou le résumé automatique
- Il suffit de réaliser un affinage de ces modèles sur notre tâche.

Question

Comment représenter la réponse et le contexte (frames sémantiques) sous forme de séquence ?

Représenter les entrées comme des séquences

Informations à prendre en compte en entrée :

- La réponse attendue.
- Les éléments du contexte de la réponse nécessaires pour poser une question intéressante → frames sémantiques

Proposition

- Concaténer la réponse avec les différents éléments du contexte considéré
- Utiliser des tokens pour délimiter les différentes informations.

Exemple d'une séquence

[ANS] Flins-sur-Seine [LU] trouvés [ARG] outils du Paléolithique

Configurations de séquences

Context	[ANS] Flins-sur-Seine [CTX] outils du Paléolithique, avec dent et ossements de mammouth, trouvés à Flins-sur-Seine (Yvelines).
BasicPB	[ANS] Flins-sur-Seine [LU] trouvés [ARG] outils du Paléolithique
BasicPB/FN	[ANS] Flins-sur-Seine [LU] trouvés [ARG] outils du Paléolithique
NamedPB	[ANS:ARGM-LOC] Flins-sur-Seine [LU] trouvés [ARG1] outils du Paléolithique
NamedFN	[ANS:Location] Flins-sur-Seine [LU] trouvés [Sought_entity] outils du Paléolithique

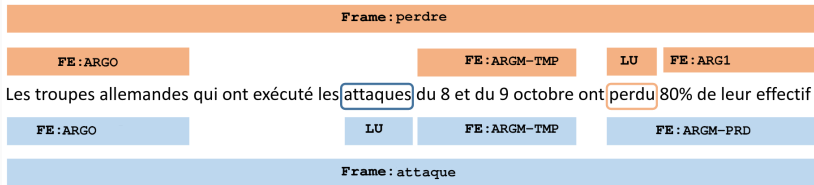
Utilisation du corpus FQUAD pour avoir des questions de référence.

Étapes de transformation du corpus :

1. Analyse automatique en frame sémantique (PropBank et FrameNet) des contextes et questions ;
2. Association des réponses avec leur frame sémantique la plus « pertinente ».
3. Construction des séquences d'entrées à partir des frames sémantiques sélectionnées.

Un corpus FQUAD-SEQ créée par configuration (c.-à-d. 5 versions différentes)

Choix de la frame sémantique pertinente



Question de référence dans FQUAD :

- Combien de troupes ont été **perdues** ?
→ Frame **perdre**

Frame pertinente :

→ **perdre** car présente dans le contexte et la question

Séquence construite (BasicPB) :

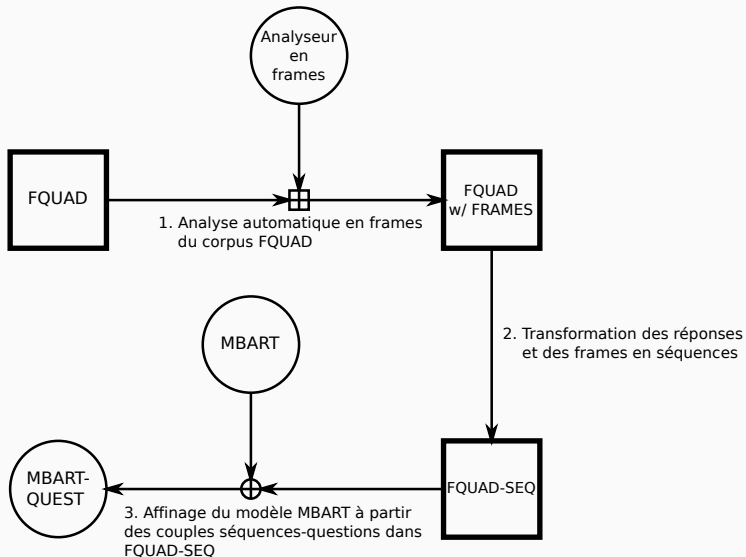
→ [ANS] 80% de leur effectif [LU] perdu [ARG] Les troupes allemandes [ARG] du 8 et du 9 octobre

Utilisation d'un modèle MBART :

- Existence d'un modèle préentraîné sur 25 langues
 - Permet le français vers le français
- Facile d'utilisation avec Huggingface
- Très gourmand en mémoire
 - Nécessité de geler les poids de la couche d'*embeddings*

Affinage d'un modèle MBART par type de configuration de séquences.

Récapitulatif



Évaluation des questions générées sur différents corpus MRQA transformés en séquences :

- CALOR-NAT : Possède déjà une annotation manuelle en frames sémantiques
→ environ 1 200 exemples
- PIAF
→ environ 2 600 exemples
- FQUAD (dev)
→ environ 4 000 exemples

BLEU (SacreBLEU)

Utilisé en traduction automatique.

Permet de déterminer à quel point une question générée ressemble à une ou plusieurs questions de référence.

Jugement humain

Réalisé sur 100 questions pour CALOR-NAT et PIAF par 1 humain.

Division en 3 catégories :

- Bonne question
- Mauvaise question
- Question partiellement correcte

Résultats (BLEU)

Configuration	CALOR-NAT	FQUAD	PIAF
Context	20.25 (± 1.42)	13.92 (± 0.76)	9.95 (± 0.37)
BasicPB	22.93 (± 6.77)	9.09 (± 3.35)	6.66 (± 1.64)
BasicPB/FN	22.93 (± 6.77)	8.76 (± 2.32)	5.74 (± 1.53)
NamedPB	26.64 (± 1.91)	11.22 (± 0.77)	7.60 (± 0.72)
NamedFN	16.26 (± 10.96)	7.26 (± 4.71)	4.95 (± 3.27)

- Résultats très bas mais surtout dû à la nature de la métrique
- PropBank est intéressant quand les annotations sont parfaites.
- Sur des annotations automatiques en frames, le contexte brut est plus intéressant.

Résultats (jugement humain)

Sur CALOR-NAT :

Configuration	Bonne	Partielle	Mauvaise
Context	49	13	38
BasicPB	48	17	35

Sur FQUAD :

Configuration	Bonne	Partielle	Mauvaise
Context	56	9	35
BasicPB	29	11	60

Entrée : [ANS:ARGM-LOC] Dans les villes de l' Empire romain
[LU] **construction** [ARG1] d' un ou plusieurs thermes
[ARG2] aux citoyens

Sortie 1 : Dans quelles villes **ont été construites** les thermes ?

Sortie 2 : Où **ont été construits** les thermes ?

Entrée : [ANS:ARGM-LOC] Dans les villes de l' Empire romain
[LU] construction [ARG1] d' un ou plusieurs thermes
[ARG2] **aux citoyens**

Sortie 1 : Dans quelles villes ont été construites les thermes ?

Sortie 2 : Où ont été construits les thermes ?

Entrée : [ANS:ARG2] de 50300 ans (à 2200 ans près [LU] datant
[ARG1] dans la strate n° 11

Sortie : De quand date la strate n°11 ?

Entrée : [ANS:ARG1] le troisième vase carolingien de ce type
[LU] connu [ARGM-LOC] en Grande-Bretagne

Sortie 1 : Quel est le troisième vase carolingien de ce type ?

Sortie 2 : Qu'est-ce qui est connu en Grande-Bretagne ?

Entrée : [ANS:ARGM-TMP] du début du II^e millénaire [LU]
exprime [ARG0] Un proverbe sumérien

Sortie : Quand un proverbe sumérien exprime-t-il Balzac ?

Utilisation des questions générées pour du MRQA

Utiliser la génération automatique de questions pour produire des annotations automatiques de questions-réponses sur de nouveaux corpus.

Objectifs

- Évaluer la qualité des questions générées pour une utilisation pratique.
- Permettre la production de modèles question-réponse pour de nouveaux domaines et types de documents.

À partir de corpus annotés en frames sémantiques, pour chaque *frame element* :

1. Elle devient une réponse possible.
2. On construit la séquence correspondante.
3. On utilise MBART-QUEST pour générer la question de la réponse.

Procédure appliquée pour les 5 configurations de génération de question.

Trois corpus annotés en frames sémantiques sont utilisés :

- CALOR-GOLD : annoté manuellement
→ environ 20 000 paires question-réponse
- CALOR-PRED : annoté automatiquement
→ environ 15 000 paires question-réponse

Apprentissage

Affinage du modèle CAMEMBERT sur la tâche de QA.

Production d'un modèle par corpus MRQA généré automatiquement

→ 10 types de modèles différents

Évaluation

Évaluation sur les questions naturelles de CALOR.

→ 2069 exemples

Métriques :

- **Exact Match** : La réponse prédite doit correspondre exactement avec la référence
- **F1** : Se fonde sur le nombre de mots en commun entre la prédiction et la référence → moins stricte

Configuration	Exact Match	F1
Règles-FrameNet	71.10	83.02
Règles-PropBank	61.90	75.77
Context	66.86	77.44
BasicPB	68.81	79.15
BasicPB/FN	68.83	78.71
NamedPB	70.18	79.92
NamedFN	70.35	80.41

→ PropBank et FrameNet presque équivalents

→ Résultats proches de ceux obtenus avec Règles-FrameNet

Résultats avec CALOR-GOLD

Configuration	Exact Match	F1
Règles-FrameNet	71.10	83.02
Règles-PropBank	61.90	75.77
Context	66.86	77.44
BasicPB	68.81	79.15
BasicPB/FN	68.83	78.71
NamedPB	70.18	79.92
NamedFN	70.35	80.41

→ PropBank et FrameNet presque équivalents

→ Résultats proches de ceux obtenus avec Règles-FrameNet

Configuration	Exact Match	F1
Règles-FrameNet	71.10	83.02
Règles-PropBank	61.90	75.77
Context	66.86	77.44
BasicPB	68.81	79.15
BasicPB/FN	68.83	78.71
NamedPB	70.18	79.92
NamedFN	70.35	80.41

→ PropBank et FrameNet presque équivalents

→ Résultats proches de ceux obtenus avec Règles-FrameNet

Configuration	Exact Match	F1
Règles-FrameNet	67.4	78.6
Context	65.48	76.58
BasicPB	66.87	77.75
BasicPB/FN	65.81	77.14
NamedPB	66.06	77.35
NamedFN	66.65	77.48

- Possibilité de se passer d'analyses sémantiques
- Nature des relations sémantiques inutile avec des annotations sémantiques automatiques

Configuration	Exact Match	F1
Règles-FrameNet	67.4	78.6
Context	65.48	76.58
BasicPB	66.87	77.75
BasicPB/FN	65.81	77.14
NamedPB	66.06	77.35
NamedFN	66.65	77.48

- Possibilité de se passer d'analyses sémantiques
- Nature des relations sémantiques inutile avec des annotations sémantiques automatiques

- Système de génération de question permettant de se passer de FrameNet
- Il est également envisageable se passer de PropBank **si** une autre méthode de sélection des réponses est disponible.



Béchet, F., Aloui, C., Charlet, D., Damnati, G., Heinecke, J., Nasr, A., and Herledan, F. (2019).

Calor-quest : generating a training corpus for machine reading comprehension models from shallow semantic annotations.

In MRQA : Machine Reading for Question Answering-Workshop at EMNLP-IJCNLP 2019-2019 Conference on Empirical Methods in Natural Language Processing.