

"Défi chatbot juridique" HackaTAL 2019, Toulouse

 **Juju** : *M. et Mme Botentouche
ont un bot juriste.*

Comment s'appelle-t-il ?

 **Claire** : ???

 **Juju** : *JUJU !! 😊*

Ségolène, Manying, Léon, Nicolas

Objectif

Assister automatiquement
M. Toutlemonde
pour trouver l'information juridique dont il a besoin

Problématiques

- Pb1 : traduction : langage(clair) \rightleftharpoons langage (intermédiaire) \rightleftharpoons langage (juridique)
- Pb2 : recherche : question \rightarrow réponse
hyp : réponse présente dans base de QA humaine
quest_clair $[\rightarrow$ quest_juridique \rightarrow rép_juridique $]\rightarrow$ rép_clair
- Pb3 : interface interactive en langage naturel avec gestion du dialogue à même de clarifier un besoin, router vers expert
- Pb4 : génération des réponses

pb1: Clair vs Intermédiaire vs Juridique

si j prend
mais congé
pdt mon
préavis es ce
q ca le
prolonge?

Puis-je
prendre mes
congés
pendant mon
préavis ?

[Art. 38](#).§ 1er. Le
travailleur peut
résilier le contrat
pendant la
suspension de son
exécution au sens
(des articles 28, 1° ,
2° et 5°) [[1](#) ...][1](#) et
31. <L 1983-11-
29/31, art. 2, 006>

En cas de congé
donné par le
travailleur avant ou

Données

- ~ Des textes en langage clair / intermédiaire / juridique
- ~ Paires de Question/Réponse
 - en langage intermédiaire et juridique (référence aux textes de lois)
 - catégorisées (taxonomie hiérarchisée et multi label)

Approche pb2 : ordonnancement

1. Représentation de la question posée et des questions candidates de notre base dans le même espace vectoriel
2. Mesure d'un score de similarité entre les vecteurs
3. Ordonnancement et renvoie de la réponse au top 1

Pré : réduction de l'espace de candidats via moteur de recherche ou classification

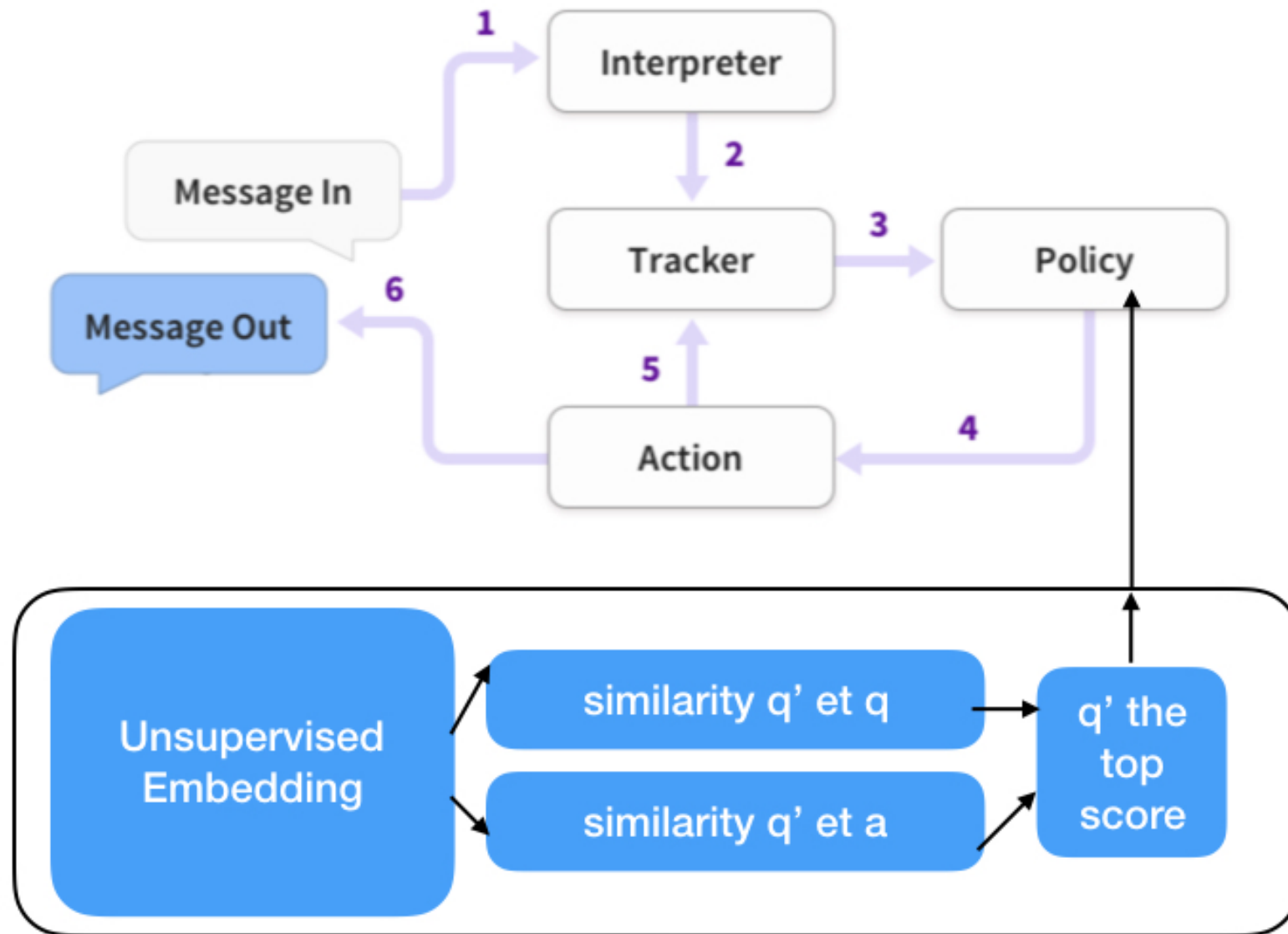
Repr. : Plongement de mots construit sur même corpus
Du mot à la phrase par SIF_Embedding

Approche pb1 : traduction des représentations

1. Construire des représentations spécifiques à chaque langage via corpus spécifique
2. Déterminer une transformation linéaire pour passer d'un espace à un autre
 1. identifier des "cognats" e.g. les n premières formes ayant la même vraisemblance dans les deux corpus
 2. utiliser leurs vecteurs pour constituer deux matrices C et J
 3. estimer T en minimisant $\sum_i (|| T c_i - j_i ||)$
3. Appliquer la transformation sur l'input pour projeter sur un même espace

Approche pb3 : chatbot

- Framework rasa :
 - composant nlu, dialogue management
 - Des règles puis du renforcement par "active learning" avec de la généralisation via prolongement de mots
- Génération de règles à partir de base de question/réponse en langage intermédiaire
 - intention à reconnaître : catégorie des questions
 - exemples d'intention : les questions
 - action à exécuter : les réponses



Protocole d'évaluation

Annotation par expert de 45 exemples

- Mesure d'évaluation : P, R, MAP, P@n
- Approche de base tf.idf
-

Résultat

- Un restaurant argentin
- Quelques bières et futs de rhum
- 2 prototypes complémentaires
 - bot auto-apprenant avec gestion du dialogue
 - ordonnancement par similarité de question avec KB
- Approche pour traduction
- Corpus de ref annoté
- Collaboration inter-disciplinaire

Discussion

- Experts TAListe et experts juriste doivent se connaître et parler un même langage
- Langage claire, langage juriste, langage intermédiaire ?
- Rôle du juriste ?
 - annoter des données ? description de la partie métier ?
valider ? analyser les erreurs ? Et puis après ?
- Rôle du taliste ?