Traduction automatique contextuelle avec sélection du mot de contexte pertinent

Dorsaf Haouari, Jian-yun Nie

RALI, université de Montréal

Plan

- Contexte
- État de l’art et problématique
- Solution proposée
- Expériences
- Conclusion

Contexte

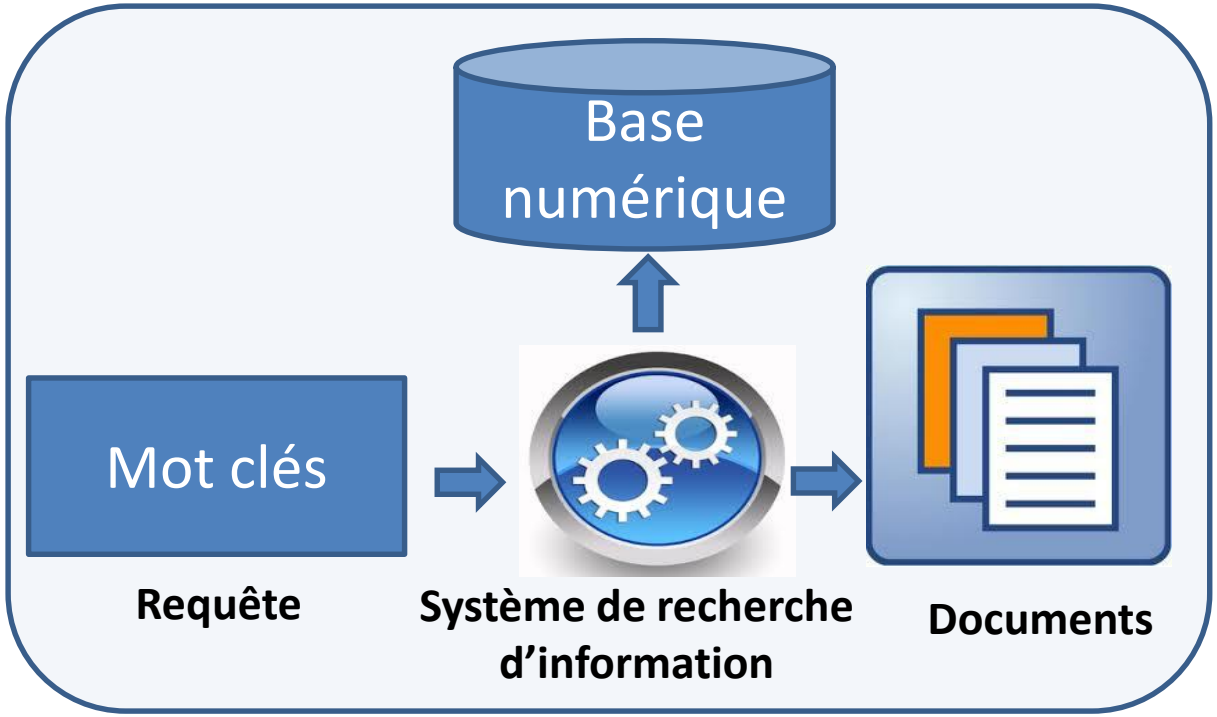
- Bibliothèque numérique:
  - Large volume de données
  - Différents domaines
  - Différentes langues

- Besoin de trouver le document pertinent

  => système de recherche d'information

Contexte

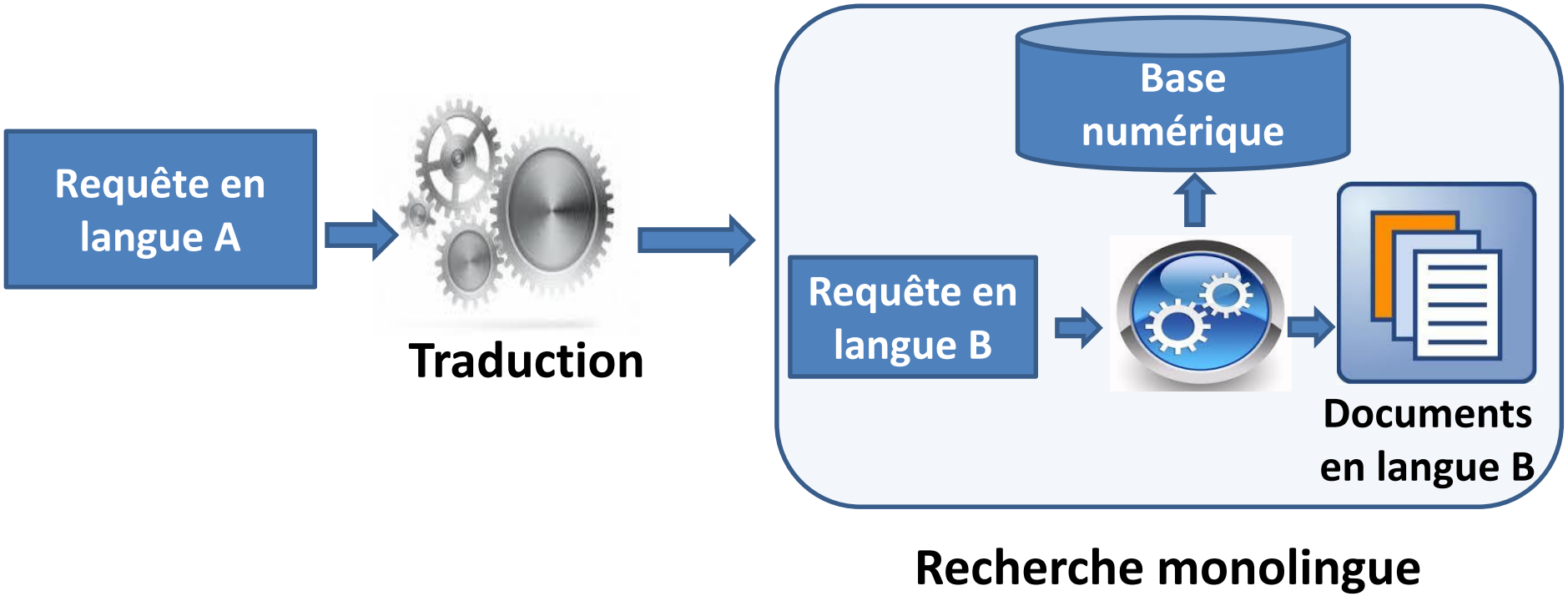
- Recherche d'information:



=> Recherche d'information translinguistique

Contexte

- Recherche d'information translinguistique:



État de l’art et problématique

- Modèle de traduction de mots
  - IBM
  - Unité de traduction = mot

reprise resumption 0.147378
reprise of 0.038884
reprise session 0.017293
reprise declare 0.00142493

État de l'art et problématique

- Limite des modèles de traduction de mots
  - pas de prise en compte du contexte des mots sources

- Mauvaises traductions:
  - Pomme de terre => apple of soil (potatoe)
  - Il livre une commande: sens de livraison (deliver)
  - Un livre de mathématique : sens de l'article livre (book)

État de l'art et problématique

- Modèle de traduction de segments
  - Unité de traduction = segment

```
ces sans-papiers ||| these illegal immigrants ||| 0.1 0.003766740.333333 0.00876794 2.718 ||| 0-0 1-1 1-2 ||| 10 3 1

ces sans-papiers ||| these stateless people ||| 1 0.00346736 0.333333 0.000622986 2.718 ||| 0-0 1-1 1-2 ||| 1 3 1

ces sans-papiers ||| those who have no papers but ||| 1 0.0012611 0.333333 1.7533e-10 2.718 ||| 0-0 1-1 1-3 1-4 1-5 ||| 1 3 1

ces tigres de papier , ces ||| these paper tigers , these ||| 1 0.00745598 1 0.104598 2.718 ||| 0-0 3-1 1-2 4-3 5-4 ||| 1 1 1

ces tigres de papier , ||| these paper tigers , ||| 1 0.0109508 1 0.207562 2.718 ||| 0-0 3-1 1-2 4-3 ||| 1 1 1

ces tigres de papier ||| these paper tigers ||| 1 0.0139421 1 0.279537 2.718 ||| 0-0 3-1 1-2 ||| 1 1 1
```

État de l'art et problématique

- Limites des modèles de traduction de segments
  - Pas de prise en compte des dépendances distantes entre les mots sources

[Le premier concerne] [un renforcement] [supplémentaire] [de la transparence]

- Exemple de dépendance distante:

  grappe puissante d'ordinateurs

État de l'art et problématique

- Limites des modèles de traduction de segments

  - Défauts connus en entretien d'embauche

    - Quelles sont les connexions à tenir en compte?
    - Choix minutieux des dépendances

      Entretien+ défauts => défauts de maintenance X

      Entretien+embauche, défauts+entretien+embauche ✓

Solution proposée

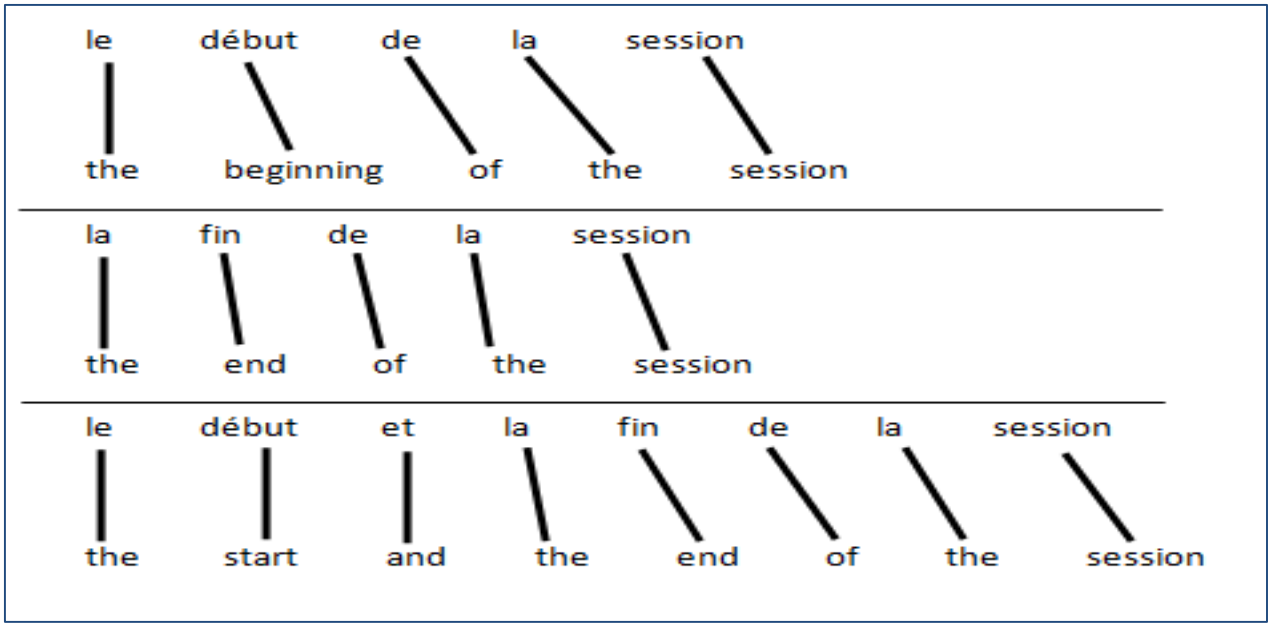
- Niveaux d'adaptation du modèle de traduction:
    - traduction en tenant en compte des dépendances entre les mots qu'elles soient adjacentes ou distantes
    - Traduction en sélectionnant les mots de contexte pertinents

Modèle de traduction contextuel

- Entrainement

Vous avez souhaité un débat à ce sujet dans les prochains jours, au cours de cette période de session.	You have requested a debate on this subject in the course of the next few days, during this part-session.
Les coûts de cette organisation du marché de la fécule de pommes de terre ont été contrôlés.	The costs of the market organization for potato starch are being kept under control.
Émissions spécifiques de CO dues aux voitures particulières neuves	Carbon dioxide emissions from new cars
L'humanité a besoin d'explorer à fond cette possibilité de production d'énergie.	Mankind needs to explore this possibility for energy production thoroughly.
Facebook, parmi d'autres services et plateformes en ligne, constitue une ouverture virtuelle importante au reste du monde.	Facebook, as well as on-line services and platforms, are an important virtual passage to the rest of the world.

Modèle de traduction contextuel



$$P(t \mid s, s_c) = \frac{\#(t, s)_{s_c}}{\#s, s_c}$$

t: traduction
s: mot source
sc: mot de contexte

$$P(beginning \mid début, session) = \frac{\#(beginning, début)_{session}}{\#début, session}$$

Modèle de traduction contextuel

- Extrait d'une table de traduction contextuelle

t	s	s_c	$P(t\|s,s_c)$
poissons	fish	consumption	0.53
poissons	fish	consumption	0.29
pêche	fish	consumption	0.07
quantité	fish	consumption	0.03
....			

pêcher	fish	laws	0.33
peut	fish	laws	0.16
poissons	fish	laws	0.16
....			

Modèle de traduction contextuel

- Choix des traductions:
  - Sc correspond à tout mot voisin du mot source dans la requête à traduire

$$P_c(t \mid s) = \frac{1}{\# s_c} \sum_{s_c} P(t \mid s, s_c)$$

$$t^* = \arg\max_{t_i} P_c(t_i \mid s)$$

Modèle de traduction contextuel

- Sélection des mots de contexte pertinents:
- Hypothèse:
  - Un mot de contexte est pertinent si la distribution de probabilité de traduction contextuelle est différente de la distribution de la probabilité de traduction non contextuelle
  - Exemple:
    - $P(t|pomme) \neq P(t|pomme, terre)$
    - $P(t|langue) \neq P(t|\ langue, médecin)$
    - $P(t|viande) \approx P(t|viande, s_c)$
    - $P(t|forêt) \approx P(t|forêt, s_c)$

Modèle de traduction contextuel

- Sélection des mots de contexte pertinents:
  - Méthodes de comparaison entre les distributions de probabilité:
    - KL-divergence $$D_{KL}(P \| Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$
    - Jensen-Shanon divergence $$JSD(P \| Q) = \frac{1}{2} D(P \| M) + \frac{1}{2} D(Q \| M) \quad \text{avec} \quad M = \frac{1}{2}(P + Q)$$
    - L'entropie croisée $$H(P,Q) = H(P) + D_{KL}(P \| Q) \quad \text{avec } H(P) \text{ est l'entropie de P}$$

Expériences

- Traduction des requêtes de l'anglais vers le français et calcul du Map

- Résultats:

  MC= modèle contextuel

  MHC= modèle non contextuel (IBM)

EN->fr	MHC	MC
Trec 6	**0,24**	0,21
Trec 7	**0,23**	0,21
Trec 8	**0,28**	0,26

Expériences

- Résultats:

$$P_i(t \mid s) = \lambda P_c(t \mid s) + (1-\lambda) P(t \mid s)$$

Lamda	Trec 6	Trec 7	Trec 8
0	0,242	0,236	0,2897
0.1	0,243	0,2404	**0,2998**
0.2	0,2474	0,2403	**0,2998**
0.3	**0,2503**	0,2399	0,2961
0.4	0,2416	0,2431	0,2954
0.5	0,2415	0,2454	0,2919
0.6	0,2405	0,245	0,2919
0.7	0,2438	0,2479	0,2931
0.8	0,2438	0,2604	0,2838
0.9	0,2439	**0,2605**	0,2883
1	0,2107	0,2136	0,2618

Expériences

- Sélection du mot de contexte

	KLD	KLD-Inverse	EC	EC-Inverse	JSD
Trec6	0,2022	0,1998	0,2001	0,214	0,1981
Trec7	0.157	0.1726	0,1624	0.1831	0.1576
trec8	0.2668	0.2693	0.2502	0.2521	0.2667

EN->fr	MHC	MC
Trec 6	**0,24**	0,21
Trec 7	**0,23**	0,21
Trec 8	**0,28**	0,26

Expériences

- Trec 6

$$P_i(t \mid s) = \lambda P_c(t \mid s) + (1-\lambda) P(t \mid s)$$

	MC+MHC	MC EC+ MHC	MC ECInverse+ MHC	MC JSD+ MHC	MC KLD+MHC	MC KLDInverse + MHC
0	0,242	0,242	0,242	0,242	0,242	0,242
0,1	0,243	0,2441	0,2433	0,2448	0,2448	0,2433
0,2	0,2474	**0,2501**	0,2632	**0,2508**	**0,2575**	0,2571
0,3	**0,2503**	0,2358	0,2612	0,2325	0,2392	0,2571
0,4	0,2416	0,2429	0,2645	0,2401	0,2468	**0,261**
0,5	0,2415	0,248	0,2659	0,2452	0,2505	0,2578
0,6	0,2405	0,2466	0,2664	0,2442	0,2495	0,258
0,7	0,2438	0,2466	0,2654	0,2434	0,2486	0,2539
0,8	0,2438	0,2466	0,2661	0,2434	0,2486	0,2552
0,9	0,2439	0,2466	**0,2666**	0,2434	0,2486	0,2557
1	0,2107	0,2001	0,214	0,1981	0,2022	0,1998

Expériences

- Trec 7

$$P_i(t \mid s) = \lambda P_c(t \mid s) + (1-\lambda) P(t \mid s)$$

	MC+MHC	MC EC+ MHC	MC ECInverse+ MHC	MC JSD+ MHC	MC KLD+MHC	MC KLDInverse+ MHC
0	0,236	0,236	0,236	0,236	0,236	0,236
0,1	0,2404	0,2497	0,2397	0,2508	0,2507	0,2398
0,2	0,2403	**0,2627**	0,2425	**0,2594**	**0,2592**	**0,2426**
0,3	0,2399	0,2571	0,2459	0,239	0,2387	0,2425
0,4	0,2431	0,2527	0,2452	0,2357	0,2362	0,2376
0,5	0,2454	0,2431	0,2477	0,2228	0,2233	0,2401
0,6	0,245	0,2367	**0,2481**	0,2156	0,2161	0,2393
0,7	0,2479	0,2345	0,248	0,2137	0,2141	0,2393
0,8	0,2604	0,2299	0,23	0,1983	0,2141	0,2214
0,9	**0,2605**	0,2298	0,2293	0,1982	0,2139	0,2206
1	0,2136	0,1624	0,1831	0,1576	0,157	0,1726

Expériences

- Trec 8

$$P_i(t \mid s) = \lambda P_c(t \mid s) + (1-\lambda) P(t \mid s)$$

	MC+MHC	MC EC+ MHC	MC ECInverse+ MHC	MC JSD+ MHC	MC KLD+MHC	MC KLDInverse +MHC
0	0,2897	0,2897	0,2897	0,2897	0,2897	0,2897
0,1	**0,2998**	0,2995	0,2897	0,3008	0,3013	0,2897
0,2	**0,2998**	**0,3237**	0,289	**0,3281**	**0,3285**	0,3083
0,3	0,2961	0,3119	0,2879	0,3129	0,3128	0,3072
0,4	0,2954	0,3096	0,2975	0,3173	0,3176	0,3151
0,5	0,2919	0,3049	0,3023	0,3128	0,3134	0,3193
0,6	0,2919	0,3021	**0,3107**	0,3134	0,314	**0,3288**
0,7	0,2931	0,289	0,2966	0,3003	0,3008	0,3148
0,8	0,2838	0,2706	0,2791	0,2839	0,2845	0,2972
0,9	0,2883	0,2706	0,2791	0,2825	0,2831	0,2974
1	0,2618	0,2502	0,2521	0,2667	0,2668	0,2693

Expériences

- Moses: décodeur

	Trec6	Trec7	trec8
Moses	**0.2729**	0.2440	**0.3361**
Expériences précédentes	0.2666 (MC EC inversé + MHC)	**0.2627** (MC EC + MHC)	0.3288 (MC KLD inverse + MHC)

Conclusion

- Application du modèle de traduction contextuel pour la recherche d'information
  - Dépendance distante
  - Sélection du mot de contexte pertinent

- Propositions futures:
  - Intégration du modèle contextuel dans un décodeur:
    - Modèle de langue
    - Modèle de cohérence