COMPRÉHENSION AUTOMATIQUE DE LA PAROLE SANS DONNÉES DE RÉFÉRENCE

TALN 2015, Caen

Emmanuel Ferreira, Bassam Jabaian et Fabrice Lefèvre

23 juin 2015

CERI-LIA, Université d'Avignon, France

Motivations et approche

- **Motivations :** méthodes état de l'art en compréhension automatique de la parole reposent sur grande quantité de données annotées
  - dépendance aux données, réel obstacle pour portabilité vers une nouvelle tâche/langue
- **Approche :** méthode limitant ce besoin par un mécanisme d'apprentissage sans données de référence (zero-shot learning)
  - combine description ontologique minimale de la tâche et espace sémantique continu
  - stratégie d'adaptation en ligne à l'aide d'une supervision faible et ajustable
- **Évaluation :** tâches de compréhension de la parole de référence (Dialog State Tracking Challenges 2 et 3).
  - modèle simple et peu coûteux avec, dès le démarrage, des performances état de l'art

Compréhension automatique de la parole

- **Principe :** extraire une séquence de m étiquettes sémantiques (concepts) $C = c_1, c_2, ..., c_m$ d'une phrase utilisateur de n mots, $W = w_1, w_2, ..., w_n$
  - transcrite ou issue d'un reconnaisseur de parole, avec beaucoup d'erreurs
- Standard d'annotation sémantique, fondé par Univ. Cambridge et utilisé dans les campagnes d'évaluation DSTC2 et DSTC3
- c_i définie par un triplet type acte/champ/valeur
  - "hello i am looking for a french restaurant in the south part of town" $\rightarrow$ `[hello(), inform(food=french), inform(area=south)]`

Etiquettes sémantiques

- Combinaisons de `acttype(champ=valeur)` déterminées sur la base d'un **inventaire ontologique**
- 4 grands groupes de type d'acte de dialogue, indépendants de la tâche :
  1. transmettre de l'information (`inform`),
  2. requêtes (`request, reqalts, reqmore`),
  3. relatifs aux confirmations (`confirm, affirm, negate, deny`),
  4. gestion de l'interaction (`hello, thankyou, bye`)
- Ensemble des couples champ/valeur lié à la tâche de dialogue. 1 couple $\leftrightarrow$ 1 entrée dans la BD utilisée pour répondre aux requêtes des utilisateurs (e.g. contraintes de recherche)

Apprentissage sans données de référence

- **Cas particulier d'apprentissage :** certaines classes peuvent ne pas être présentes dans l'ensemble d'exemples du corpus d'apprentissage
- **Notre interprétation :** prédire la séquence d'étiquettes d'une phrase utilisateur sans avoir vu au préalable un exemple de phrase utilisateur dans le **contexte de l'interaction**
- Source de connaissance sémantique doit être exploitée pour extrapoler ces classes liées aux étiquettes directement à partir de leur définition

Méthode

Combinaison de :

1. Espace sémantique continu F de dimension d pouvant coder les propriétés des étiquettes sémantiques
2. Base de connaissances K, dictionnaire d'exemples dans F, relie l'espace sémantique à l'espace de sortie du système
3. Analyseur sémantique extrait liste ordonnée des meilleures hypothèses de séquence d'étiquettes sémantiques
   - à partir d'un transducteur à états finis représentant l'ensemble des hypothèses pour une phrase utilisateur, scorées par les informations issues de F et de K.

Méthode (aperçu)

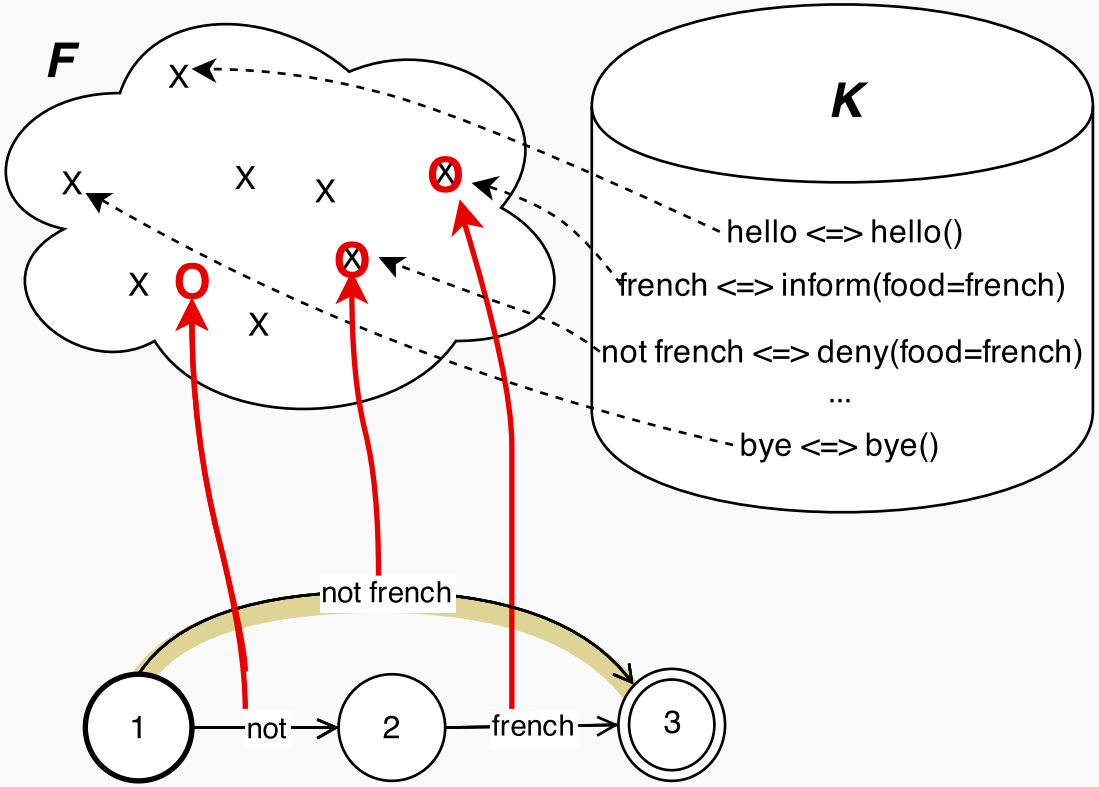


Illustration d'un décodage sémantique basé sur une technique d'apprentissage sans données de référence

Espace sémantique continu

- Représentation vectorielle compacte de mots (word embedding) apprise avec réseaux de neurones profonds
  - régularités notables sur les propriétés syntaxiques et sémantiques des mots modélisés
  - travaux ont monté l'intérêt de ces représentations sur différentes tâches de traitement automatique des langues naturelles
- Représentation définissant un espace sémantique continu avec possibilités de généralisation
  - apprentissage non-supervisé sur une très grande quantité de données (de large couverture, type dump wikipedia)
  - pas de données spéciales liées à la tâche
  - existence de techniques permettant d'adapter/transférer l'espace sur une tâche ou langue spécifique

Base de connaissances sémantique...

- Relations entre étiquettes sémantiques et des exemples de forme de surface qui leur sont associées
  - "what food is served?" $\leftrightarrow$ `request(food)`, "yes" $\leftrightarrow$ `affirm()`, "french food" $\leftrightarrow$ `inform(food=french)`...
- Matrice d'affectation représentant les informations ontologiques du domaine visé :

colonne étiquette sémantique

ligne vecteur d'exemple de dimension d dans F

valeur cellule valeur d'affectation, $c_{i,j}$, indique s'il existe une relation entre le $i^{ème}$ vecteur de F et la $j^{ème}$ étiquette sémantique

Base de connaissances sémantique

IMPORTANT pas besoin d'exhaustivité lors de la définition de ces exemples (contrairement à une approche à base de règles expertes - grammaire), le recours à l'espace sémantique *F* permettra la généralisation **après coup** pendant l'analyse sémantique

Analyseur sémantique

Phase de décodage, pour chaque nouvelle phrase :

1. Inventaire des segments (sous-séquences de mots contiguës)
   - ex.“yeah downtown” $\rightarrow$ “yeah”, “downtown” et “yeah downtown”.
2. Segments projetés dans F et comparés (similarité cosinus) aux vecteurs associés aux exemples dans K
3. Construction d’un transducteur à états finis : segments et hypothèses sémantiques sont les entrées/sorties des arcs, pondérées par les similarités cosinus avec k plus proches hypothèses sémantiques associées à chaque segment
   - repondération règle l’influence de la longueur des segments considérées
4. ”Plus court chemin” appliqué sur le transducteur

$\rightarrow$ Liste ordonnée d’hypothèses scorées de séquences d’étiquettes sémantiques

Processus de décodage

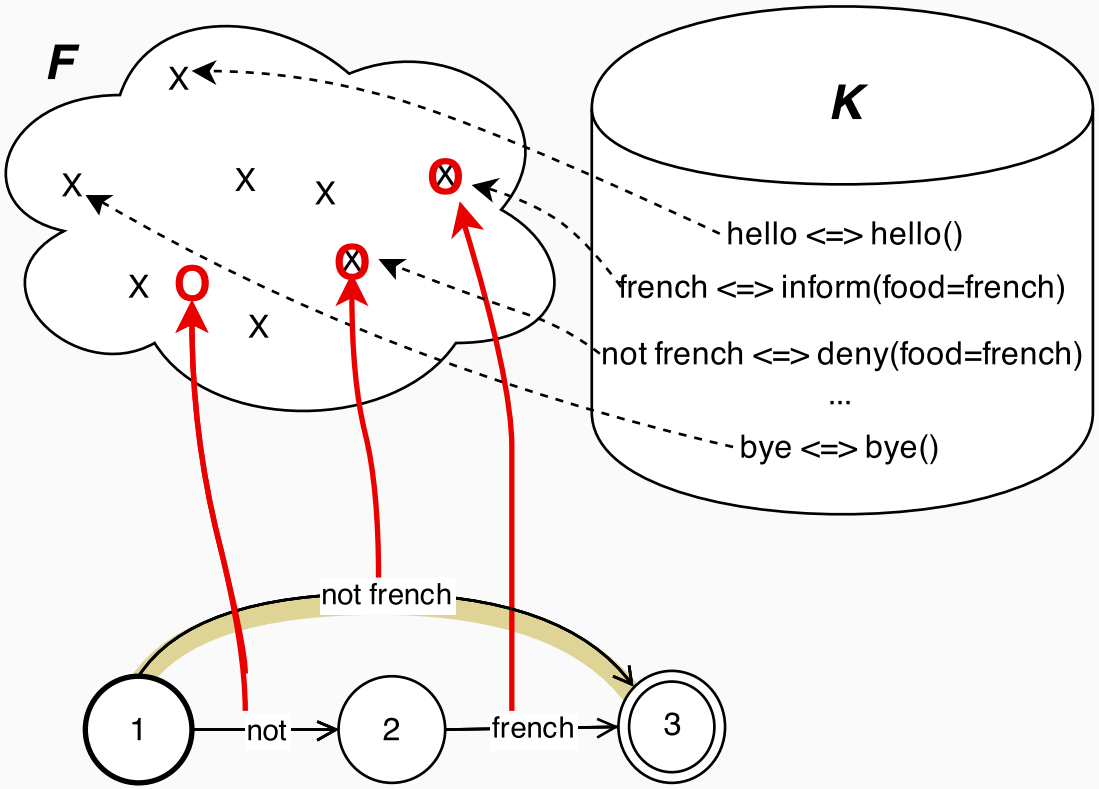
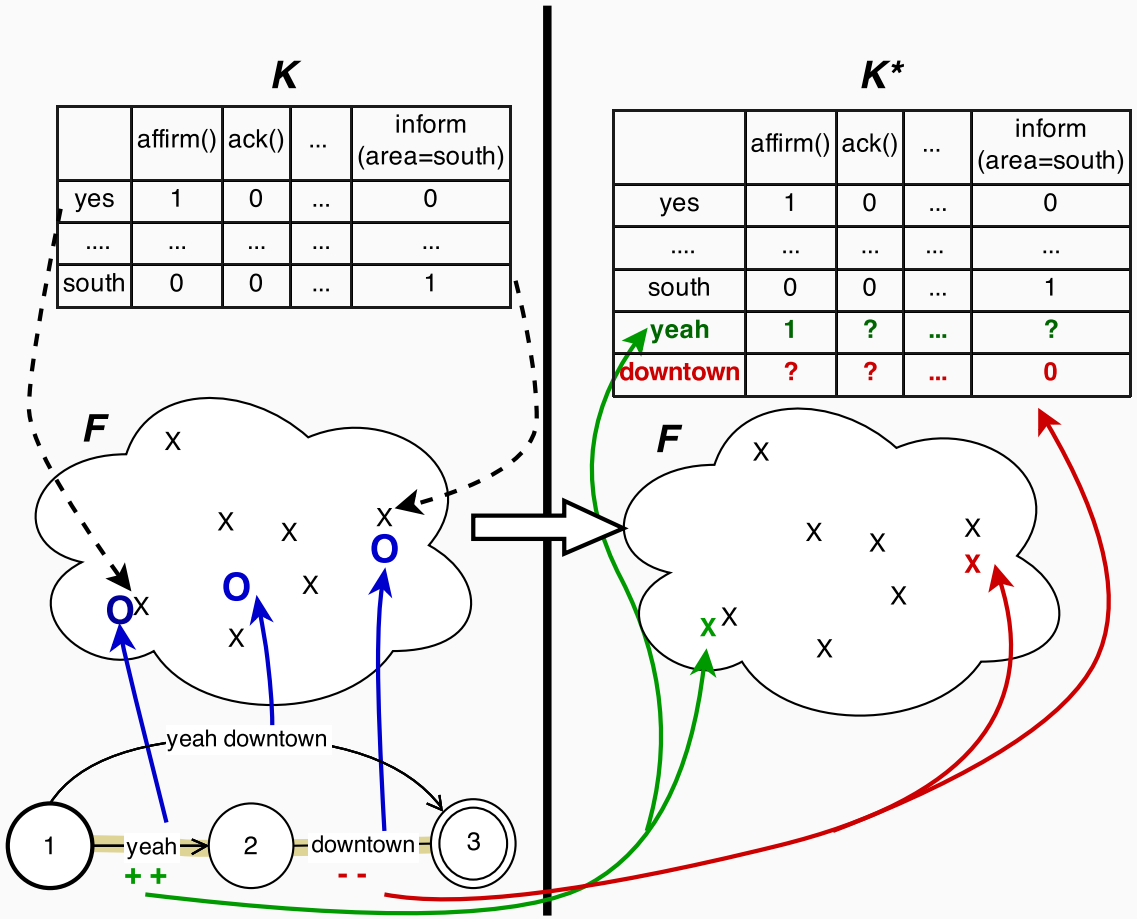


Illustration d'un décodage sémantique basé sur une technique d'apprentissage sans données de référence

Adaptation en ligne

- Mise à jour (en ligne) des valeurs d'affectation dans K en fonction des retours utilisateurs suite à l'interrogation du module de compréhension
- Scénario avec **supervision limitée** à un ensemble de retours binaires (validation/rejet) sur les étiquettes sémantiques produites.
  - sans correction manuelle des étiquettes de la part des utilisateurs
  - facilement intégrable sur une plate-forme de dialogue existante en utilisant des retours simples de l'utilisateur ("oui", "ok", "non", "faux"...)
- Ratio coût/amélioration contrôlé en déterminant une **politique de demande** de retours aux utilisateurs.
  - (définition d'une stratégie optimale, objet de travaux en cours)

Processus d'adaptation en ligne



	affirm()	ack()	...	inform (area=south)
yes	1	0	...	0
....	...	...	...	...
south	0	0	...	1

	affirm()	ack()	...	inform (area=south)
yes	1	0	...	0
....	...	...	...	...
south	0	0	...	1
yeah	1	?	...	?
downtown	?	?	...	0

Illustration d'un décodage sémantique basé sur une technique d'apprentissage sans données de référence avec adaptation en ligne

Description de données

- Défi de recherche dédié à la détection du but de l'utilisateur tout au long d'un dialogue oral
- Corpus DSTC2 et DSTC3
  - DSTC2 recherche d'informations sur des restaurants
  - DSTC3 couvre recherche d'informations touristiques plus générale, nouveaux types d'établissement (pubs, coffee shops) mais aussi de nouveaux champs et valeurs
  - pas orientés étiquetage sémantique des énoncés
- Données de test
  - 1k dialogues (10k énoncés utilisateurs) pour DSTC2 et 2k (20k) pour DSTC3
- Deux modes d'évaluation :
  - transcriptions manuelles et
  - n-meilleures transcriptions automatiques

Évaluation de l'approche proposée

Dans nos expériences,

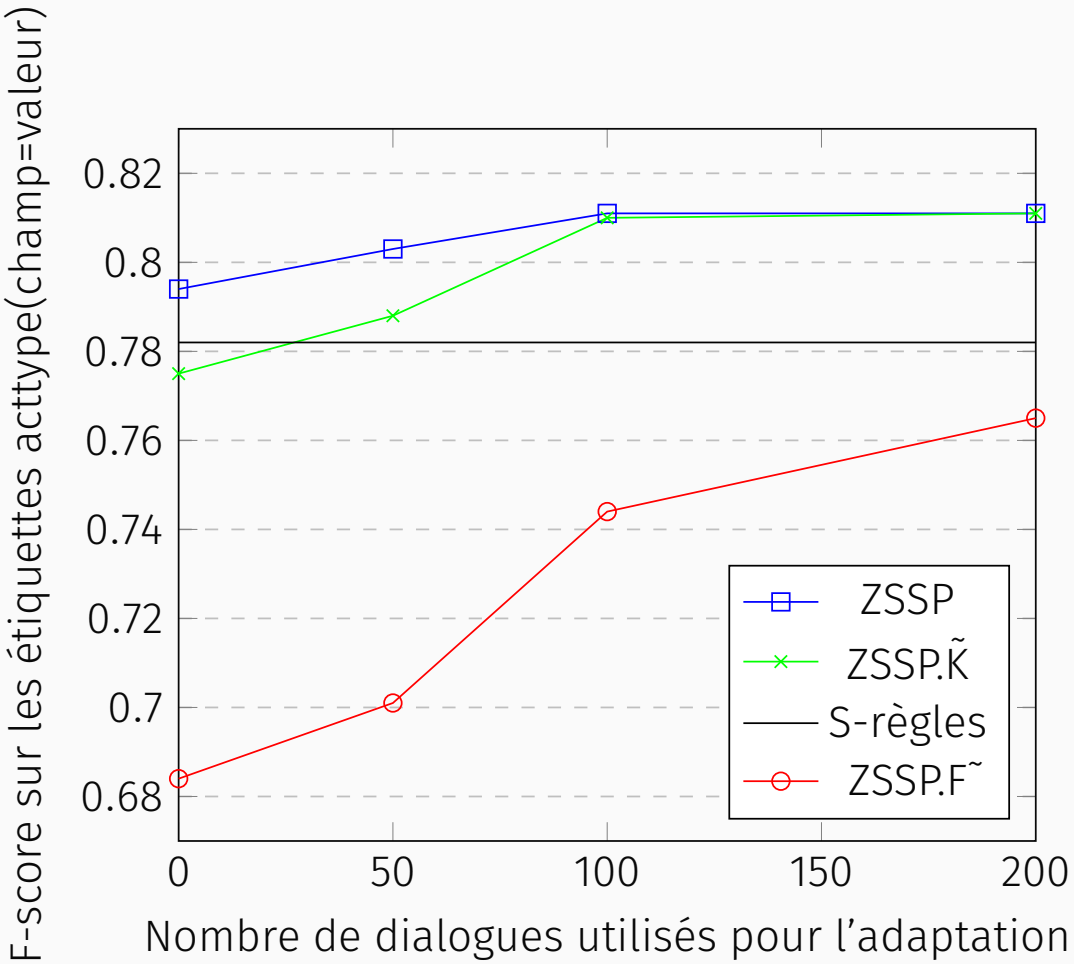
- **Modèle word2vec** (Mikolov et al., 2013) pour apprendre la représentation vectorielle *F* des mots avec 300 dimensions
  - *Skip-gram* (fenêtre de 10 mots) sur une grande quantité de données libres (>4e9 mots en contexte), anglais et large couverture thématique
- **Base de connaissances** basée sur la description ontologique du domaine fournie dans le challenge (liste des champs/valeurs et BD) et un ensemble d'informations de dialogue générique
  - étiquettes : 633 DSTC2, 855 DSTC3 // entrées : 4160 DSTC2, 6555 DSTC3
  - 53 entrées "manuelles" pour acttypes non présents dans la BD (e.g. "say again" $\leftrightarrow$ `repeat`)
- Deux **systèmes état de l'art** pour comparaison : règles expertes, utilisé dans le défi DSTC ("S-règles"), et système SLU1 de (Williams et al, 2014), appris sur les données d'apprentissage du DSTC2 ("S-appris")

Evaluation analyseur sémantique

Tâche	Modèle	Entrée	F-score	P	R
DSTC2	S-règles	n-meilleures	0,782	0,900	0,691
	S-appris	n-meilleures	0,802	0,846	0,762
	ZSSP	manuelle	0,919	0,898	0,942
		n-meilleures	0,794	0,796	0,792
DSTC3	S-règles	n-meilleures	0,824	0,852	0,797
	ZSSP	manuelle	0,899	0,873	0,928
		n-meilleures	0,826	0,806	0,849

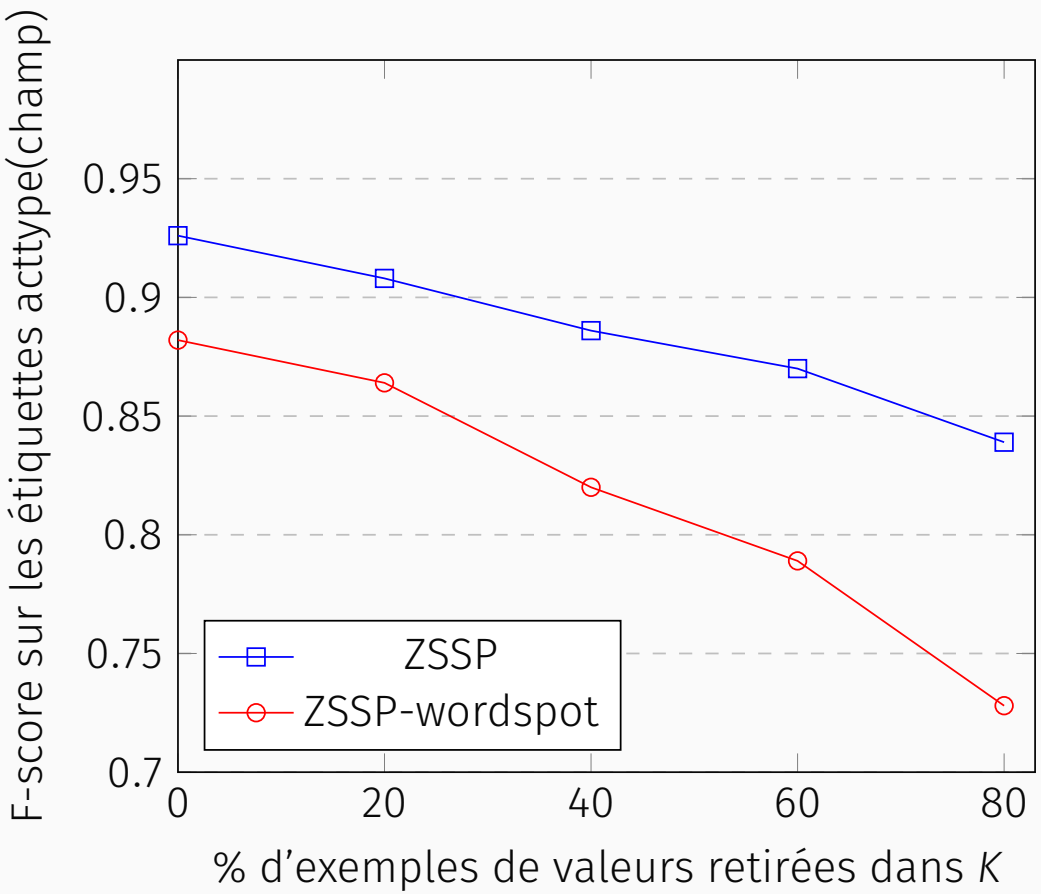
Évaluation des performances de l'analyseur sémantique basé sur l'apprentissage sans données de référence en termes de F-score, **P**récision et **R**appel.

Evaluation adaptation en ligne



Performances de diverses configurations de la méthode ZSSP en termes de F-score en fonction du nombre de dialogues utilisées pour l'adaptation.

Evaluation robustesse et généralisation



Capacité de généralisation de l'approche ZSSP sur la corpus de test DSTC2 exprimée en termes de F-score sur la détection d 'actes de dialogue génériques (i.e. acttype(champ)) en fonction du pourcentage d'exemples de valeurs retirées dans *K*.

Conclusions

Approche d'apprentissage sans données de référence pour la compréhension de la parole

1. Représentation sémantique riche apprise sur des données généralistes
2. Description ontologique minimale décrivant la tâche visée

Avantages

1. Approche, très peu coûteuse et performances comparables à des méthodes statistiques apprises sur de grandes quantités de données annotées ou à un système à bases de règles expertes
2. Meilleure tolérance à des valeurs de concept manquantes
3. Processus d'adaptation simple et ajustable en ligne
4. Supervision légère (utilisateur confirme/réfute les hypothèses du système, pas de corrections explicites des erreurs)

Perspectives

Travaux en cours

- Comparaison avec d'autres techniques d'apprentissage active et généralisation de cette technique par l'adaptation d'une vision plus probabiliste et dynamique
- Évaluation dans le contexte d'interactions complètes

Merci de votre attention

Questions ?