De la modélisation au traitement automatique des données linguistiques

Infos pratiques

- > ECTS : 3,0
- > Nombre d'heures : 24,0
- > Période de l'année : Enseignement neuvième semestre
- > Méthodes d'enseignement : En présence
- > Forme d'enseignement : Cours magistral
- > Composante : Philo, Info-Comm, Langages, Littératures & Arts du spectacle
- > Code ELP : 4L9TL03P

Présentation

Une des problématiques principales du TAL est d'arriver à ce que les machines comprennent le langage humain en s'appuyant sur des indices présents directement dans les données linguistiques. Depuis quelques années, les travaux en TAL portent de plus en plus vers le repérage de l'information « implicite », déduite en quelque sorte du corpus. Il s'agit tout d'abord de la fouille d'opinions visant la détection des avis, des émotions ou des sentiments. D'autres travaux se concentrent sur le repérage des intentions émises dans les discussions orales ou sur Internet, dans les avis sur les lieux visités, etc. Toutes ces informations qu'on cherche à détecter doivent être d'abord modélisées. Cette modélisation guidera ensuite les technologies utilisées : méthodes symboliques, apprentissage supervisé de surface ou apprentissage profond.

Objectifs

Savoir modéliser un phénomène linguistique et mettre en place une chaîne de son traitement automatique.

Évaluation

M3C en 2 sessions

- Régime standard session 1 – avec évaluation continue (au moins 2 notes, partiel compris) :

ou

- Régime standard session 1 – avec évaluation terminale (1 seule note) :

Un projet par groupe

- Régime dérogatoire session 1 :

Un projet par groupe

- Session 2 dite de rattrapage :

Un projet par groupe

Compétences visées

savoir modéliser un phénomène linguistique,
procéder à l'annotation manuelle et automatique
connaître des outils d'annotation manuelle et automatique

être capable d'adapter le traitement automatique selon la nature du corpus traité

Bibliographie

- Barbedette A., Eshkol-Taravella I. (2020), « Prédire automatiquement les intentions du locuteur dans des questions issues du discours oral spontané », TALN2020
- Bunt, H., Petukhova, V., Traum, D., and Alexandersson, J. (2017). Dialogue Act Annotation with the ISO 24617-2 Standard, pages 109–135.
- Eshkol-Taravella I., Kang H. J. (2019). « Observation de l'expérience client dans les restaurants », *TALN2019*, 1-5 juillet 2019, Toulouse, France.
- Flamein H., Eshkol-Taravella I. (2020), « De la parole à la carte : repérage, analyse et visualisation automatique de la perception d'une ville », CMLF2020

Flamein H., Eshkol-Taravella I. (2020). « Noms de lieux dans le corpus de français parlé : Une approche symbolique pour un traitement automatisé », *Le français moderne* 2020, n.1

Grabar N., Eshkol-Taravella I. (2016). Prédiction automatique de fonctions pragmatiques dans les reformulations. *TALN2016*, Paris, France.

Grice, H. (1975). Logic and conversation. Syntax and Semantics 3 : Speech Acts. New York : Academic Press, 41-58. Reprinted in Grice, pages 22-40.

Jakobson, R. (1963). Linguistique et poétique. Essais de Linguistique Générale, pages 209-248.

Kang H. J., Eshkol-Taravella I. (2020), « Les avis sur les restaurants à l'épreuve de l'apprentissage automatique », TALN2020

Karoui, J., Benamara Zitoune, F., Moriceau, V., Aussenac-Gilles, N., and Hadrich Belguith, L. (2015). Détection automatique de l'ironie dans les tweets en français. In 22eme Conference sur le Traitement Automatique des Langues Naturelles (TALN 2015), pages 1-6, Caen, France.

Contact(s)

> Iris Eshkol taravella

Responsable pédagogique
ieshkolt@parisnanterre.fr