



Interactions multimodales humain-machine

Analyse de sentiments et modèles de comportements
construits par *apprentissage automatique sur corpus*

Magalie Ochs, Sébastien Fournier, et Patrice Bellot

Equipe R2I, Laboratoire LIS, Aix Marseille Université

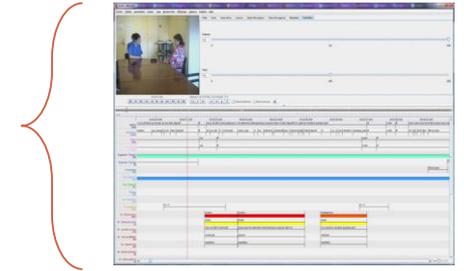
Rôle du LIS

Objectifs

Comportement socio-émotionnels



Signaux multimodaux



WP1 : Corpus

1. Annotation automatique

- Modèle d'annotation automatique des sentiments (modèles existants développés dans l'équipe)

2. Signaux multimodaux utilisateurs -> comportements persuasifs

- Apprentissage automatique/data-mining sur corpus

WP2 : Modélisation du comportement non-verbal

1. Lorsque l'agent *écoute* : Modèle de feedbacks

- Modèle de calcul automatique des feedbacks de l'agent durant l'interaction (verbal et non-verbal)

2. Lorsque l'agent *parle* : Modèle de comportements non-verbaux

- Quels gestes, expressions faciales, regards, accompagnent le message verbal pour un comportement persuasif ?
- Modèle de raisonnement automatique sur le comportement non-verbal à adopter (Dynamic bayésien network)

3. Synchronisation verbal et non-verbal :

- Coupler le modèle verbal avec le modèle de comportement non-verbal
- Intégrer le modèle de comportement non-verbal avec les stratégies haut-niveau (LIMSIS)

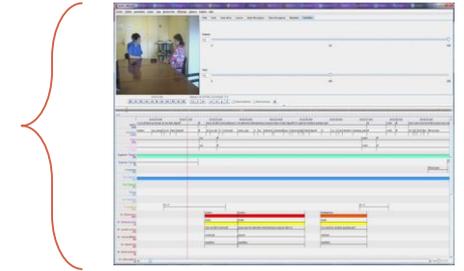
Perception : Analyse automatique de corpus

Objectifs

Comportement socio-émotionnels



Signaux multimodaux



Questions de recherche

1. Collecte et annotation de corpus *multimodaux*

- Interactions *spontanées* vs *actées* (acteurs)
- Annotations :
 - Annotation automatique vs. Manuelle
 - Subjectivité : interprétation des signaux *en* comportements socio-émotionnels
- **Problématiques** : peu de corpus, difficile à acquérir et à annoter

2. Hypothèse de la *réplicabilité*

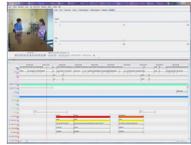
- Comportements sociaux d'un personnage virtuel exprimés de la même manière qu'un humain ?
- Capacités expressives des systèmes interactifs humanoïdes \neq humains
- **Problématiques** : *réplicabilité a priori* : modèle d'expressions construit sur des corpus humain-humain

3. Apprentissage multimodal sur corpus

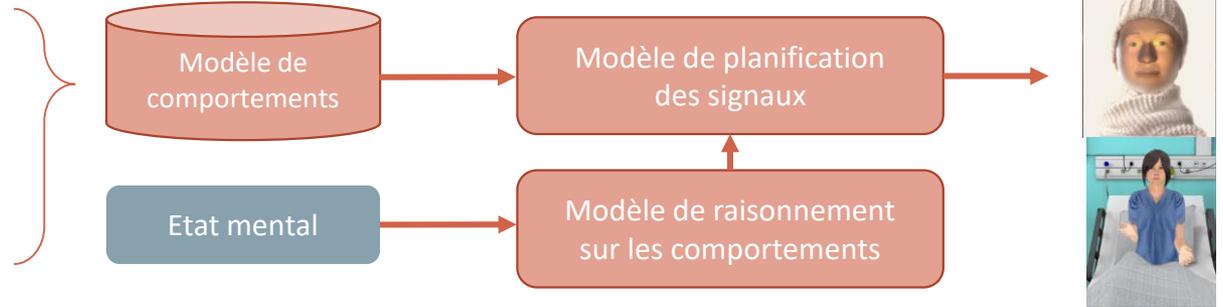
- Taille de données, grande dimension
- Interprétabilité des modèles
- **Problématiques** : apprentissage sur des données brutes, grande masse de données

Modélisation : Planification des signaux et comportements socio-émotionnels

Objectifs



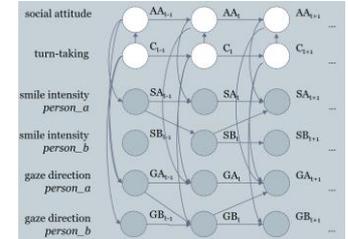
Signaux multimodaux
↕
Comportements socio-émotionnels



Questions de recherche

Comment exprimer un comportement socio-émotionnel ?

- Modélisation de la relation signaux-comportements : *modèle graphique*
- Modèle de planification des signaux : *signaux non-additifs*
- **Problématiques** : apprentissage d'un modèle explicite *multimodal*



Comportements socio-émotionnels

Cohérence
Crédibilité
Variabilité



Interaction humain-machine

Objectifs généraux du projet

❑ Former les médecins avec un patient virtuel

➤ Interaction « naturelle »

- ✓ Situation réelle
- ✓ Interaction en langage naturel
- ✓ Multimodalité

➤ *Mesures comportementales* de performance

➤ Différent degrés d'*immersion*



Collecte, annotation et transcription de corpus humain-humain

✓ **Corpus**

- 2 heures de corpus issues des centres de simulation (13 vidéos – différents médecins, différents acteurs)
- Acteurs-patients : « patient standardisé »

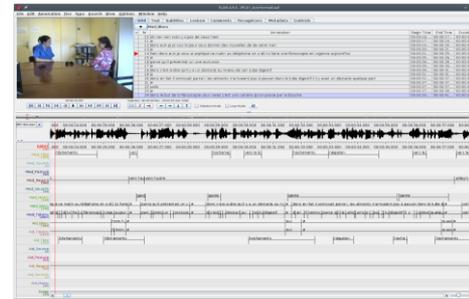
✓ **Annotation automatique**

- **Catégories morphosyntaxiques** : adjectifs, verbes, noms (*outil Marsatag, LPL*)

✓ **Annotation manuelle**

- **Transcription manuelle** du corpus
- 3 annotateurs (étudiants en linguistique payés)
- Chaque annotateur : 1/3 du corpus (sessions d'annotation encadré)
- Définition d'un *schéma de codage* (guide d'annotation)

- Annotation des indices non-verbaux : mvts de tête, changement de posture, direction de regard, mvts de sourcils, mvts des mains, sourires -> *bon accord inter-annotateur (5% du corpus)*
- Ressource : corpus d'interaction médecins-patients en contexte de 2 heures



CHU
ANGERS
CENTRE HOSPITALIER
UNIVERSITAIRE

INSTITUT
PAOLI-CALMETTES

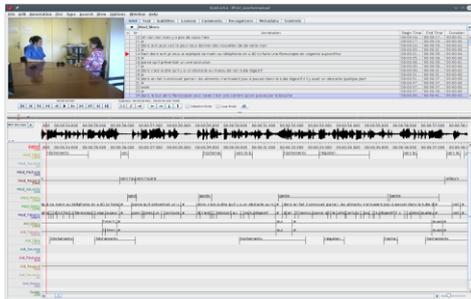
PAROLE ET
LANGAGE

LIS
LABORATOIRE
D'INFORMATIQUE
& SYSTEMES

Recherche #1 : Extraction automatique de séquences pour un modèle de prédiction de feedback

Projet ANR Acorformed

Objectifs : *Corpus annotés* → *Apprentissage automatique*
séquences de signaux multimodaux du médecin → *feedbacks du patient*



Extraction

Corpus de 8368 séquences multimodales terminées par un feedback

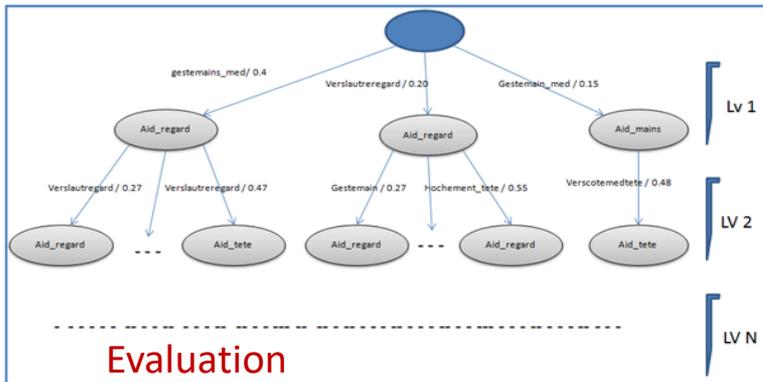
doctor_gaze_interlocutor; doctor_head_nod; patient_head_side; doctor_head_nod; doctor_noun; patient_hand_mvt; etc.

Relations causales
 $X, Z, W \rightarrow Y$

47 règles (ERMiner Algorithm)

Modélisation

doctor_head_nod ⇒ *patient_head_nod*
doctor_medical_vocabulary ⇒ *patient_head_nod*

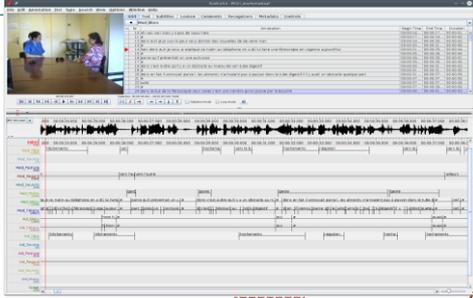


10-cross-validation

Bonne prédiction de la catégorie du feedback

Prediction of Feedback (yes/no)				
Precision (%)		Recall (%)		
97.1		63.3		
Category	Category (nead, gaze, ...)		Attribut (nead uut, etc)	
Category	Prec. (%)	Recall (%)	Prec. (%)	Recall (%)
Head	89.0	74.4	0.71	0.54
Hand	83.4	85.6	0.30	0.10
Gaze	94.6	76.7	0.10	0.05

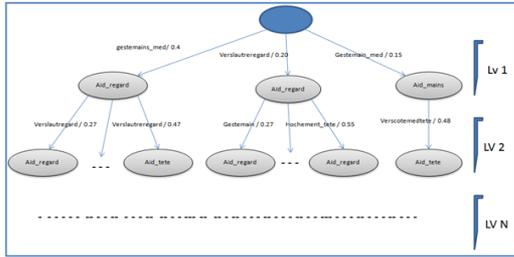
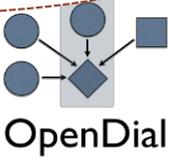
Modèles computationnels de comportements verbaux et non-verbaux d'un patient virtuel



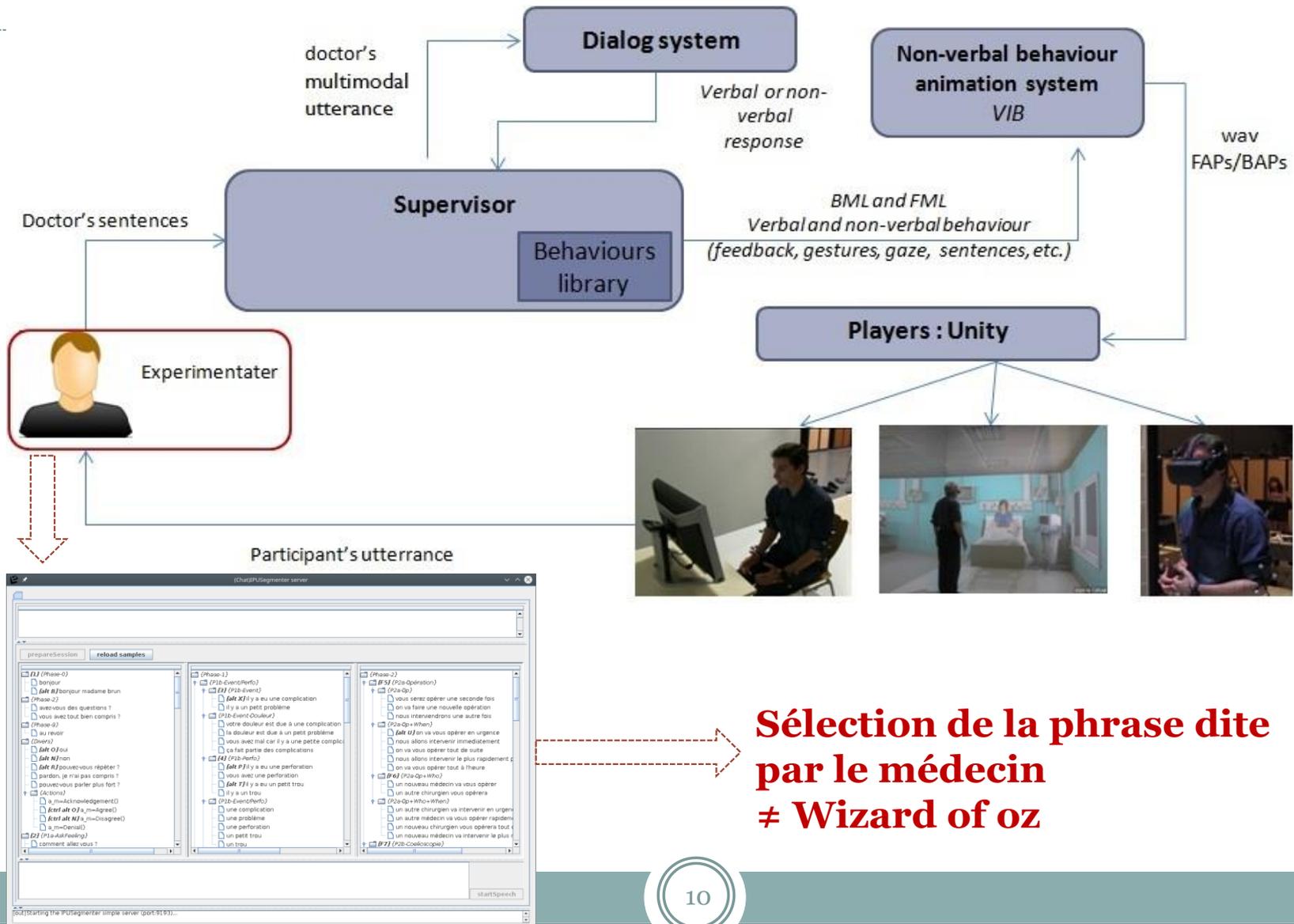
Bibliothèque de comportement
VIB existant + 16 gestes
spécifiques au scénario
d'annonce

Modèle de dialogue fondé sur
« common ground »

Modèle de feedbacks



Systeme semi-autonome



Sélection de la phrase dite par le médecin ≠ Wizard of oz

IHM, Sciences cognitives: une étude comparative

Objectifs : Comparer l'expérience d'interaction d'utilisateurs (naïfs et experts) avec le patient virtuel sur **trois dispositifs** (PC, Casque, Cave)

Pré-questionnaire
(démographique, empathie fictive,
expériences jeux/RV/Annonce)

26 naïfs+10 experts

Utilisateurs

Tâche
Annonce d'une perforation digestive



Salle de réalité virtuelle



PC simple



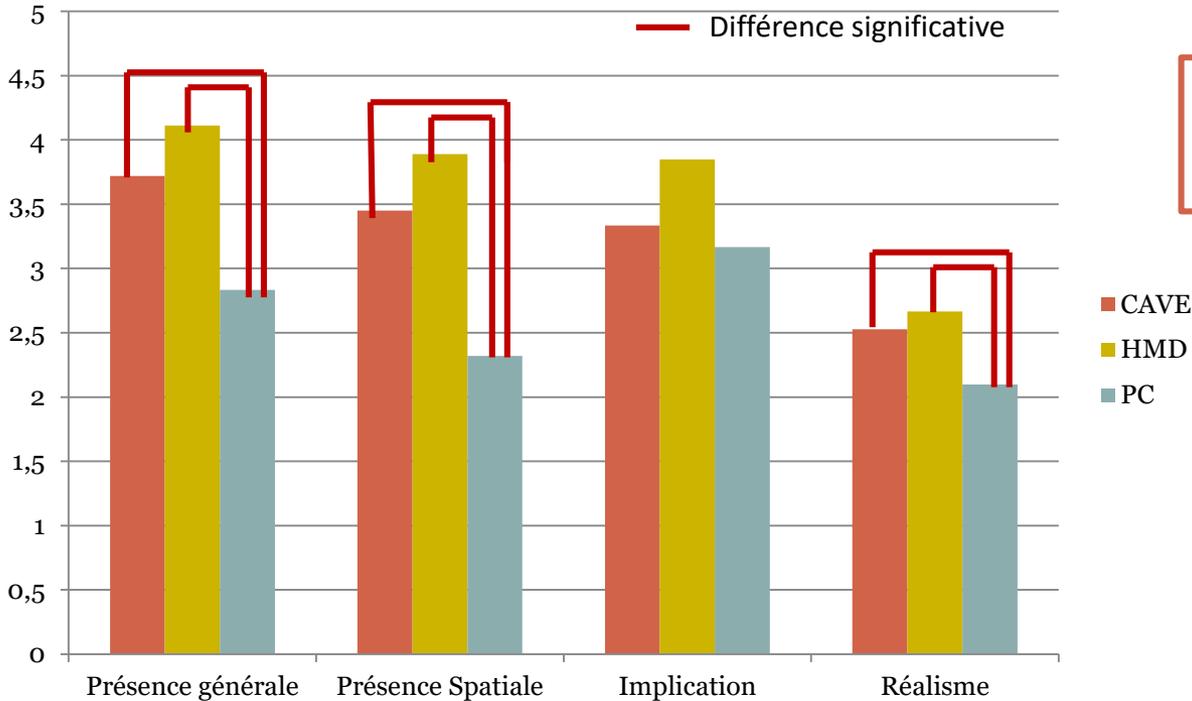
Casque de réalité virtuelle

Chaque
utilisateur
interagit
avec les
trois
dispositifs

Post-questionnaires

- Questionnaire de présence : IPQ (Shubert, 2003)
- Questionnaire de présence sociale (Bailenson et al., 2005)
- Questionnaire sur la perception des performances d'annonce

IHM, machine learning: une étude sur le sentiment de présence



Différences significatives PC –
Cave/casque
Pas de différence CAVE – Casque

- **La présence spatiale : [Casque \cong CAVE] > PC**
- La sensation physique d'être présent dans l'environnement virtuel
- **L'implication (involvement) dans le dispositif : Casque \cong CAVE \cong PC**
- Le sentiment d'être absorbé par l'environnement virtuel
- **Le réalisme : [Casque \cong CAVE] > PC**
- L'environnement virtuel semblait-il réel en comparaison de l'environnement réel ?

Recherche #3 : Analyse automatique de corpus

Projet ANR Acorformed

❑ Objectifs

Corpus annotés → Apprentissage automatique → signaux multimodaux ↔ engagement de l'utilisateur

❑ Apprentissage sur corpus multimodaux d'interactions h-m



- ✓ 86 interactions
- ✓ 3 dispositifs avec différents niveaux d'immersion
- ✓ Coordonnées spatio-temporelles de la tête, bras et mains
- ✓ Texte retranscrit (reco. vocale)
- ✓ Questionnaires post-expérience sentiment de présence et coprésence



Peu de données + grande dimensionnalité

Recherche #3 : Analyse automatique de corpus

Projet ANR Acorformed



- Signal de la parole
- Texte retranscrit
- Coordonnées spatio-temporelles tête, bras, main

Classification

Niveau de présence

Niveau de co-présence

Pré-traitement des données

- Découpage de l'interaction
- Définition de paramètres haut niveau

✓ Réduction de la dimensionnalité

✓ Paramètres *interprétables*

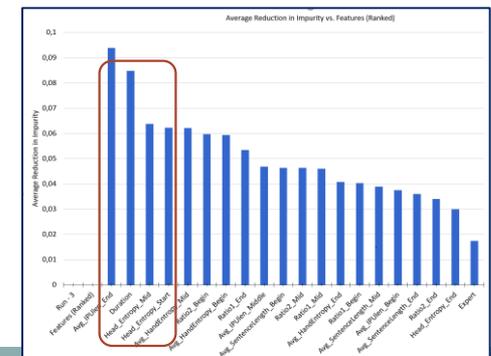
$$\frac{nb_adj + nb_adv}{\sum tokens}$$

Id	Expert	Head_Entropy Start	Head_Entropy Mid	Head_Entropy End	Hand_Entropy Start...	SentenceLength Start...	IPU-duration Start...	Linguistic_Complexity Start...	Lexical_richness Start...	Duration Interaction	Presence Class	Co-presence Class
N12F	0	1,620065371	2,772323473	2,041698294	1,392627466	5,777777778	2,297611111	0,145454545	0,163636364	176,69	low	medium
N12F	0	1,607020413	3,375024905	1,685875934	1,430638595	9,166666667	2,9125	0,148648648	0,202702703	223,46875	high	low
N12F	0	1,894656971	3,235628836	1,760955103	1,582670533	8	2,771177778	0,155555556	0,222222222	228,4375	high	medium
N1A	1	1,20840471	2,840888985	1,755899322	1,623034453	10	2,13755	0,170731707	0,219512195	136,46875	low	high
N1A	0	1,810642566	2,842002331	0,70063174	1,121666454	6,333333333	2,86985	0,159090909	0,181818182	116,5625	high	high
Macro	1	1,602270546	2,922271701	1,925000917	1,375025177	6,9	1,270	0,132407552	0,177650574	171	low	medium

Random Forest (k-cross validation)

Presence	F-score	Precision	Recall
Low	0,87	0,84	0,9
Medium	0,78	0,88	0,0,81
High	0,61	0,77	0,6
Macro-average	0,79	0,79	0,77

Co-presence	F-score	Precision	Recall
Low	0,94	0,92	1
Medium	0,66	0,63	0,71
High	0,69	0,8	0,75
Macro-average	0,78	0,79	0,76



Recherche : Analyse de sentiment

Entreprise Caléa

Thèse de Gaël Guibon

Objectifs

- Prédiction automatique d'*Emojis*

Approche

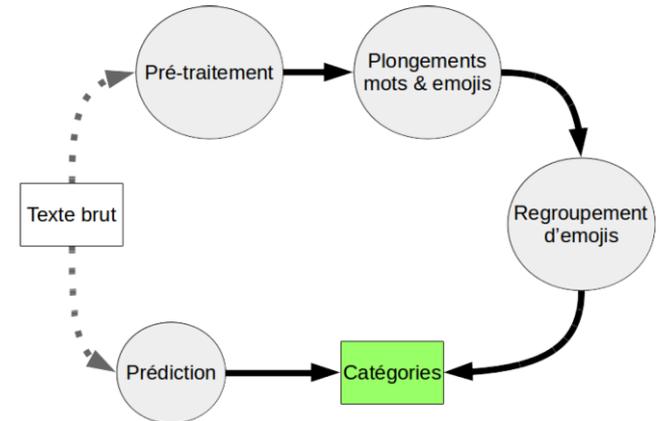
- Emojis = Mots → Sémantique distributionnelle (*word embeddings*)

Corpus d'1 million de tweets avec Emojis

Plongements lexicaux d'emojis
(*Word2Vec*)

Partitionnement
(*spectral clustering*)

face emoji embedding space



Annotation automatique des opinions

<https://lab.hypotheses.org/1384>



The screenshot shows the OpenEdition website interface. At the top, there is a navigation bar with links for 'OpenEdition Books', 'OpenEdition Journals', 'Calenda', 'Hypothèses', 'Lettre & alertes', and 'OpenEdition Freemium'. Below this, there are links for 'A propos', 'Crédits', 'Prototypes', and 'Bibliographie'. The main header features the OpenEdition LAB logo with the tagline 'Recherche - Développement - Innovation'. A secondary navigation bar includes 'Bilbo', 'Inter-Textes', 'AgoraWeb', 'Évaluation ouverte', and 'Usages et appropriations'. The article title 'Annotation des opinions dans un corpus de comptes-rendus de lecture' is prominently displayed, along with the author 'PAR ÉLODIE FAATH' and publication date 'PUBLIÉ 05/06/2015'. A search bar is visible on the right side of the page.

- présentation de l'ouvrage
- problématique de l'ouvrage
- contexte scientifique
- méthode scientifique
- argumentation de l'auteur
- organisation de l'ouvrage
- jugement sur l'ouvrage

- target ou référence explicite : il s'agit d'un mot ou groupe de mots ou expression présent dans le texte ;
- category ou propriété du compte-rendu : il s'agit de propriétés propres à un objet sur lesquelles des opinions sont exprimées ;
- polarity : valeur attribuée à chaque catégorie (neutre, positive ou négative) ;
- polarityTerms ou mot porteur d'opinion : mots ou expressions qui évoquent le sentiment.

```
<sentence id="6">
<text>Mais la méthode avec laquelle il est présenté comme seule hypothèse recevable
pose problème.</text>
<Opinions>
<Opinion target="méthode" polarityTerms="pose problème" category="methodology"
polarity="very negative" occurrence="1"/>
</Opinions>
</sentence>
```

Fig. 2 Annotation des opinions d'une phrase issue du corpus d'apprentissage

□ Analyse des bases cérébrales de l'interaction humain-machine

- Analyse EEG et IRMf lors de l'interaction entre un humain et un personnage virtuel ou un robot
- Comparaison des activités neurophysiologiques lors d'interaction humain-humain et humain-machine
- Identifier des mesures objectives neurophysiologiques d'une interaction humain-machine

Quelle est l'activité cérébrale qui peut refléter un comportement socio-émotionnel crédible d'un personnage virtuel ou un robot ?





Interactions multimodales humain-machine

Analyse de sentiments et modèles de comportements
construits par *apprentissage automatique sur corpus*

Magalie Ochs, Sébastien Fournier et Patrice Bellot

Equipe R2I, Laboratoire LIS, Aix Marseille Université