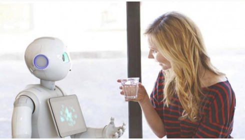


Natural Language Processing, Machine learning for Social Computing

Chloe Clavel

LTCI Telecom-ParisTech, IDS,S2a

Social Computing : scope and ML challenges

Some history : from Affective Computing to Social Computing

1997, MIT: Affective Computing

Now

From emotions to social signals and opinions

Also called **Social Computing** / **Emotion AI** /**artificial emotional intelligence**

Social Computing topic at Telecom-ParisTech

Two axis that share the same interdisciplinary issues

- Natural Language Processing, opinion mining, emotion analysis in social networks

Social Computing topic at Telecom-ParisTech

Two axis that share the same interdisciplinary issues

- Natural Language Processing, opinion mining, emotion analysis in social networks

- Human-agent interaction and social robots :

- analysis of the user's socio-emotional behavior,
- generation of agent's behaviour
- socio-emotional strategies

↪ Labex SMART, H2020 ICT Aria-Valuspa, H2020 ITN Pierre et Marie Curie Animatas, Chaire Machine Learning for Big Data, ANR JCJC MAOI, GRACE

Interdisciplinary contributors to the Social computing theme

Affective Computing

Giovanna Varni

Chloe Clavel

Computer Science

Jean-Louis Dessalles

Machine Learning

Florence d'Alché-Buc

Slim Essid

Laurence Likforman

Anne Sabourin

Social Sciences

Christian Licoppe

Nicolas Rollet

Post-doctorants/Phd students: Valentin Barriere, Atef Ben Youssef, Mohamed Ben Jannet, Emile Chapuis, Pierre Colombo, Tanvi Dinkar, Alexandre Garcia, Leo Hemamou, Hamid Jalalzai [6]

First Challenge: Big, Wild, Social Data

Supervised machine learning approaches require a labelled dataset (big dataset for deep learning approaches)

Difficulty to obtain big data labelled in socio-emotional phenomena

↪ opinions/emotions/personality are phenomena that are complex and costly to annotate

Example : opinion analysis in movie review

> *This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it cant hold up*[a]*.*
>
> [a]example taken from Pang et al

What is the opinion of the speaker about the movie? Negative?Positive? Which feature of the movie does the speaker like?

First Challenge: Big, Wild, Social Data

> *This film* **should be** *brilliant. It* **sounds like** *a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone* **is attempting to** *deliver a performance.* **However,** *it can't hold up.*

↪ more complex than a simple positive *vs.* negative word counts.

Difficulty to model complex linguistic phenomena

Find representations that allow modelling :

- conditional tense, discourse markers, metaphors – global warming *vs.* climate change [Ahmad et al., 2011], *etc.*)
- Target and source identification

User's Attitude features
source = user
target = Dali
polarity = neg

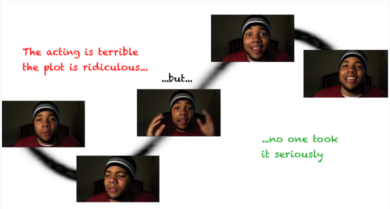
I don't like Dali

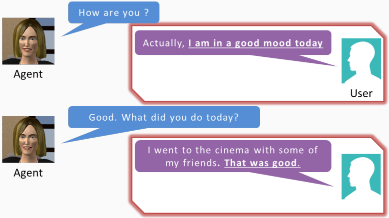
First challenge : Big, Wild, Social Data

Difficulty to model complex linguistic phenomena

Find representations that allow modelling :

- intra-speaker and inter-speaker dynamics





First Challenge: Big, Wild, Social Data

Data are "wild" [Schuller et al., 2016]

↪ in real applications, corpora contain spontaneous conversational data

- oral and interactional features :

  Disfluences combinées Vous regardez les 5 derniers **chiffres** des **chi** des numéros gravés, pas les chiffres qui défilent **hein**

- written features (ex : typos, chat features: `lol`, `A +`, `mouhahaha`)

Second challenge: explainability and transparency

- Perception of socio-emotional behavior is subjective $\Rightarrow$ Labels are subjective
- machine learning methods learn models of socio-emotional phenomena as they are expressed and labelled in the data
- $\Rightarrow$ machine learning algorithms should be transparent and their results interpretable

For human-agent/robot interactions

Transparency and explainablility are crucial so that we can understand agent/robot's strategies

What are our approaches?

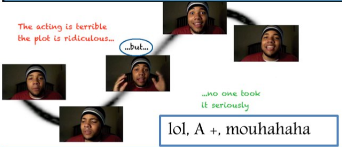
Symbolic AI vs. machine learning?

Our approach : both of them

Depending on the data that are available and on the maturity of the problem statement

What are our approaches?

Disfluences combinées Vous regardez les 5 derniers **chiffres** des **chi** des numéros gravés, pas les chiffres qui défilent **hein**



DEFI 1 :
Big, *Wild*, Social Data: collecte et modélisation linguistique des comportements socio-émotionnels en interaction

DEFI 2 :
Transparence et interprétabilité des modèles de comportements socio-émotionnels de l'agent/robot

Symbolic approaches : Reasoning models, Formal grammars
Machine learning : Structured output prediction, hierarchical and sequential models
Hybridation

Social Computing and symbolic AI

Social Computing and symbolic AI

Data issue

Problems that are still not correctly stated and for which we dont have a sufficient quantity of data with relevant labels require formal models built from :

- psycholinguistics
- small corpus observation

Explainability and Transparency issues

Social Computing and symbolic AI

Two examples of studies :

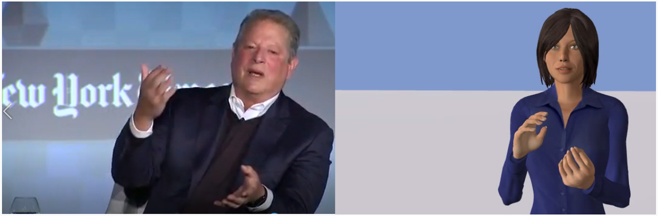
1. Reasoning models for agent's behavior generation [Ravenet et al., 2018]
2. Formal grammars for the detection of user's opinion in human-agent interaction [Langlet and Clavel, 2016]

Reasoning models for agent's behavior generation

Objective

Automatic production of the gestures corresponding to agent's utterances

Forms of the gestures according to the semantic meaning of what we are saying



Brian Ravenet, Chloé Clavel, **Catherine Pelachaud**, Automatic Nonverbal Behavior Generation from Image Schemas, AAMAS 2018

Formal models for agent's behavior generation

Problem statement and Data issue

The collection of labelled data for our problem seems to be difficult and the problem is not sufficiently formalized to define relevant annotation guide

- ex : Which type of semantic content for which gesture features?

Proposed methodology

Build models from

- psycholinguistics [Calbris], (Johnson 1987)
- small corpus observation : speeches of Obama and Al Gore

Reasoning models for agent's behavior generation

From psycho-linguistics: Image schema

Intermediate representation between gesture and verbal content

Image Schemas are recurring patterns of metaphorical reasoning (Johnson 1987)

Example : "My boss" corresponds to the **UP** Image Schema

Reasoning models for agent's behavior generation

Two steps of the reasoning model

1)Automatic extraction of **Image Schemas** from agent's verbal utterance

2)Define the physical properties of the gestures derived from Image Schemas

Future directions

How to learn to automatically map text to gestures?

Intermediate levels of representation?

text to Image Schemas ?

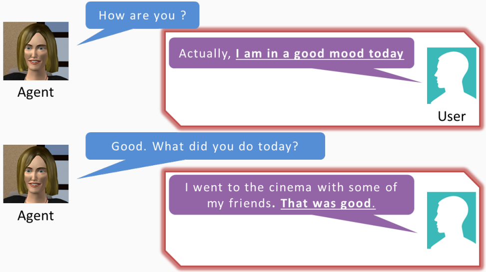
Image Schemas to gestural properties ?

gestural properties to gestures ?

Grammars for the detection of user's opinion in human-agent interaction

Objective

Detect verbal expressions of opinions as they occur in human-agent interaction



© C. Langlet's presentation at ACII 2015

Clavel, C.; **Callejas, Z.**, *Sentiment analysis: from opinion mining to human-agent interaction*, IEEE Transactions on Affective Computing,(2016)

Grammars for the detection of user's opinion in human-agent interaction

Related work

First sentiment analysis modules integrated in human-agent interactions [Smith et al., 2011, Pulman et al., 2010]

↪ use of a pretrained module [Moilanen and Pulman, 2007] not designed for human-agent interaction

Features of human-agent interaction

- User's expression of opinion depends on agent's utterance

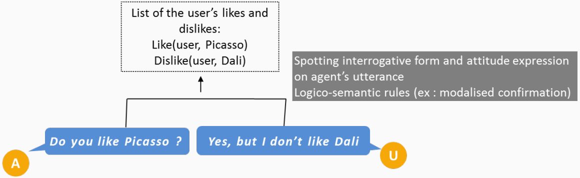
© C. Langlet's presentation at ACII 2015

Grammars for the detection of user's opinion in human-agent interaction

Grammars built from linguistic theories

- ↪ Opinions as defined by the appraisal theory [Martin and White, 2005] derived from functional systemic linguistics
- ↪ Interactive processes in conversation as defined in psycholinguistics [Clark and Schaefer, 1989], ex : **the adjacency pair** (agent utterance, user utterance)

Patterns and rules within **the adjacency pair**



C. Langlet and C. Clavel, *Improving social relationships in face-to-face human-agent interactions: when the agent wants to know users likes and dislikes* , in ACL 2015

Grammars for the detection of user's opinion in human-agent interaction

Evaluation method

Annotation of a corpus of human-agent interaction

Development of an annotation guide within a crowdsourcing platform for the annotation of opinions in interactions

Machine learning for Social Computing

Machine learning for Social Computing

Three examples of studies :

- Hidden conditional Random Fields for intra and inter-speaker Opinion dynamics modelling
- Accurate Opinion Prediction and Structured Output learning : for opinion target analysis
- Hierarchical Attention Model for social signal multimodal analysis

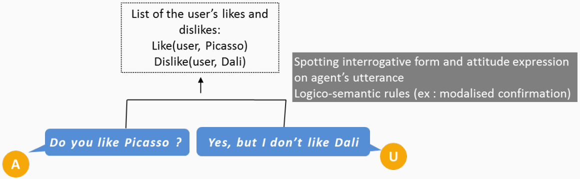
In order to answer the two challenges :

- Big Wild Social Data
- Explainability

Hidden Conditional random fields and Opinion dynamics in interactions

Challenge : Machine Learning for Interaction Modelling

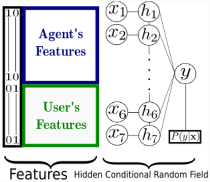
- Following the work on formal models for the detection of user's likes and dislikes,



- As the problem is now well-stated
- And annotation guide well defined
- ... ready to use supervised machine learning technics

Hidden Conditional random fields and Opinion dynamics in interactions

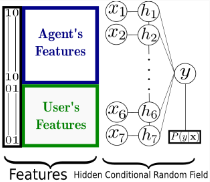
Challenge : Interaction Modelling



- Relying on the same idea that the user's expression of likes and dislikes depends on the agent's utterance ...
- Modeling the dynamics of the interaction using HCRF (Hidden Conditional Random Fields [Quattoni et al., 2007]) a latent state model interpretable and efficient with a small dataset.

V. Barriere, C. Clavel, **S. Essid**, Attitude classification in adjacency pairs of a human-agent interaction with hidden conditional random fields, to appear, ICASSP 2018

From knowledge-based to machine learning (HCRF)



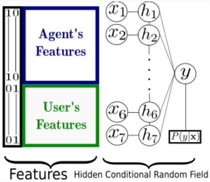
Learn weights θ_o, θ_s and θ_t that minimize the posterior probability which depends on the potential functions measuring the compatibility between label, a sequence of latent states and the observations

$$\Psi(y, \mathbf{h}, \mathbf{x}; \theta) = \underbrace{\sum_j \langle \phi(x_j) \,|\, \theta_o(h_j) \rangle}_{(1)} + \underbrace{\sum_j \theta_s(y, h_j)}_{(2)} + \underbrace{\sum_j \theta_t(y, h_j, h_{j+1})}_{(3)}$$

(1) compatibility between the observations x_j and the hidden state h_j
(2) compatibility between the hidden states and the global label y
(3) compatibility to go from one hidden state h_j to the next one h_{j+1}

V. Barriere, C. Clavel, **S. Essid**, Attitude classification in adjacency pairs of a human-agent interaction with hidden conditional random fields, to appear, ICASSP 2018

From knowledge-based to machine learning (HCRF)



Results : evaluation on a human-agent database

Inter-speaker dynamics modelling leads to better results

F-score : 0.80 *vs.* 0.74 (without considering agent's utterance)

Accurate Opinion prediction and structured output learning

Challenge : target and source identification

User's Attitude features
source = user
target = Dali
polarity = neg

I don't like Dali

- jointly predict opinions and their target is considered a **structured outupt prediction problem**

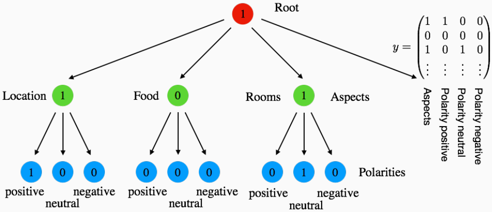
Alexandre Garcia, Chloé Clavel, **Slim Essid** , **Florence d'Alche-Buc**, Structured Output Learning with Abstention: Application to Accurate Opinion Prediction, ICML 2018

Accurate Opinion prediction and structured output learning

Structured output learning framework

- representation of the output by a directed graph

The room was ok, nothing special, still a perfect choice to quickly join the main places.



- Proposition of a new hierarchichal loss that allows learning this structure and to abstain on difficult nodes (h predicts the labels of each component of the structure and r is the abstention function)

$$\Delta_{Ha}(h(x), r(x), y) = \sum_{i=1}^{d} \underbrace{c_{A_i} 1_{\{f_i^{h,r}=a, f_{p(i)}^{h,r}=y_{p(i)}\}}}_{\text{abstention cost}} + \underbrace{c_{A_c i} 1_{\{f_i^{h,r}\neq y_i, f_{p(i)}^{h,r}=a\}}}_{\text{abstention regret}} + \underbrace{c_i 1_{\{f_i^{h,r}\neq y_i, f_{p(i)}^{h,r}=y_{p(i)}}}_{\text{misclassification c}}$$

Accurate Opinion prediction and structured output learning

Structured output learning framework

- Evaluation on TripAdvisor reviews annotated at the sentence level from [Marcheggiani et al. 2014]
- Robust framework for star rating prediction

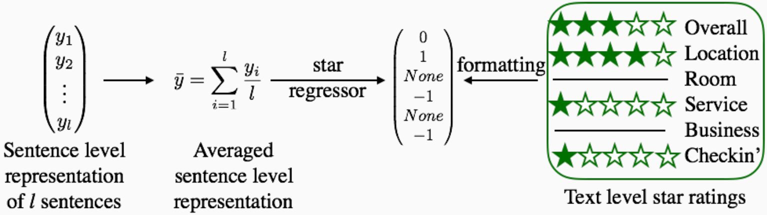


Figure 1: Star rating regression pipeline

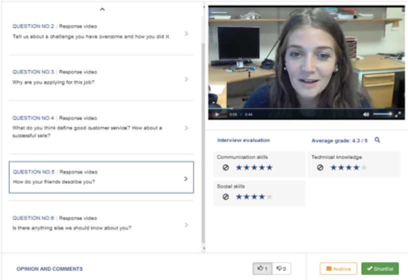
Alexandre Garcia, Chloé Clavel, **Slim Essid , Florence d'Alche-Buc**, Structured Output Learning with Abstention: Application to Accurate Opinion Prediction, ICML 2018

Hierarchichal Attention Model

Challenge : Big Wild Social data

Large scale database of labelled asynchronous video job interviews obtained in collaboration with industry (easyrecrue)

- 475 real commercial job positions
- 7095 real candidates
- 558 Hours of videos, 2845 videos
- Labels : recruiter assessments of hirability on the platform



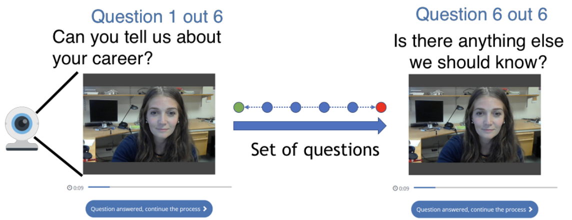
L. Hemamou, G. Felhi, V. Vandenbussche, J.-C. Martin, C. Clavel, HireNet: a Hierarchical Attention Model for the Automatic Analysis of Asynchronous Video Job Interviews. in AAAI 2019

Hierarchichal Attention Model

Challenge : Big Wild social Data

HireNet : Model structural and sequential information for social signal analysis

- Hierarchy, three levels : words, answers, interview
- Sequentiality is modelled via Gated Recurrent Units
- Attention mechanisms to add contextual cues (type of jobs, type of questions)



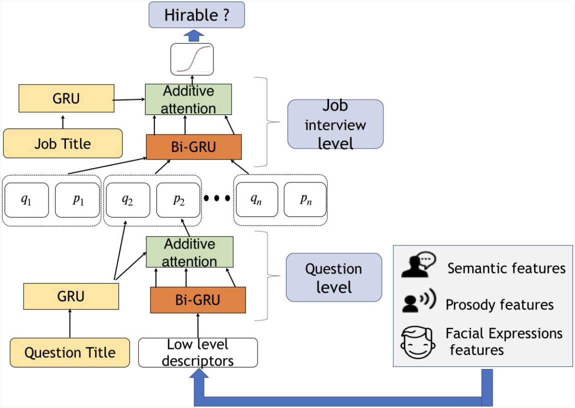
L. Hemamou, G. Felhi, V. Vandenbussche, J.-C. Martin, C. Clavel, HireNet: a Hierarchical Attention Model for the Automatic Analysis of Asynchronous Video Job Interviews. in AAAI 2019

Hierarchical Attention Model

Challenge : Big Wild social Data

HireNet

- Hierarchy, three levels : words, answers, interview
- Sequentiality is modelled via Gated Recurrent Units
- Attention mechanisms to add contextual cues (type of jobs, type of questions)



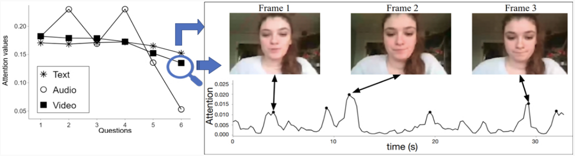
L. Hemamou, G. Felhi, V. Vandenbussche, J.-C. Martin, C. Clavel, HireNet: a Hierarchical Attention Model for the Automatic Analysis of Asynchronous Video Job Interviews. in AAAI 2019

Hierarchichal Attention Model

Challenge : Explainability and Transparency

Attention mechanism for interpretability

- to highlight the salient social cues and questions during the job interview,
- to provide the recruiter a visualization of the cues



L. Hemamou, G. Felhi, V. Vandenbussche, J.-C. Martin, C. Clavel, HireNet: a Hierarchical Attention Model for the Automatic Analysis of Asynchronous Video Job Interviews. in AAAI 2019

Hierarchichal Attention Model

Results

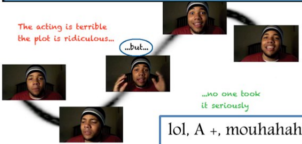
Model	Text (F1)	Audio (F1)	Video (F1)
Best baseline	0.499	0.532	0.507
Bi-GRU	0.561	0.566	0.528
With hierachical structure	0.617	0.598	0.528
With Attention	0.625	0.614	0.522
With Context	0.643	0.642	0.605

L. Hemamou, G. Felhi, V. Vandenbussche, J.-C. Martin, C. Clavel, HireNet: a Hierarchical Attention Model for the Automatic Analysis of Asynchronous Video Job Interviews. in AAAI 2019

Conclusion

In a nustshell

Disfluences combinées Vous regardez les 5 derniers **chiffres** des **chi** des numéros gravés, pas les chiffres qui défilent **hein**



lol, A +, mouhahaha

DEFI 1 :
Big, *Wild*, Social Data: collecte et modélisation linguistique des comportements socio-émotionnels en interaction

DEFI 2 :
Transparence et interprétabilité des modèles de comportements socio-émotionnels de l'agent/robot

Symbolic approaches : Reasoning models, Formal grammars [Ravenet et al., 2018] [Langlet and Clavel, 2016]
Machine learning : Structured output prediction, hierarchical and sequential models [Garcia et al., 2018], [Janssoone, 2015]
Hybridation [Barriere et al., 2017], [Dutrey, 2014]

Thank you

Just opening : Faculty position (Associate professor) in machine learning (and natural language processing)

Other references i

- Dessalles, J.-L. (2016). A Cognitive Approach to Relevant Argument Generation. In M. Baldoni, C. Baroglio, F. Bex, T. D. Bui, F. Grasso et al. (Eds.), Principles and Practice of Multi-Agent Systems, LNAI 9935, 3-15. Springer.
- ASSESSMENT OF NEW SPECTRAL FEATURES FOR EEG-BASED EMOTION RECOGNITION Conneau, Anne-Claire Et Essid, Slim, In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP),
- L. LIKFORMAN-SULEM, A. ESPOSITO, M. FAUNDEZ-ZANUY, S. CLEMENCON, G. CORDASCO EMOTHAW: A Novel Database for Emotional State Recognition from Handwriting and Drawing, IEEE THMS Transactions on Human-Machine Systems special issue on Drawing and Handwriting Processing for User-Centered Systems, E. Anquetil, G. Pirlo and R. Plamondon Eds (2017)
- Towards interactional symbiosis: epistemic balance and co-presence in a quantified self experiment Rollet, N., Jain, V., Licoppe, C., Devillers, L. (2017), in Gamberini L., Spagnolli A., Jacucci G., Blankertz B., Freeman J. (eds) Symbiotic Interaction. Symbiotic 2016. Lecture Notes in Computer Science, vol 9961. Springer, Cham.
- G. Varni, I. Hupont, C. Clavel, M. Chetouani, Computational Study of Primitive Emotional Contagion in Dyadic Interactions, IEEE Transactions of Affective Computing

Bibliography i

Ahmad, K., Budin, G., Wien, U., Devitt, A., Glucksberg, S., Heyer, G., Leipzig, U., Musacchio, M. T., Pazienza, M. T., Rogers, M., Vogel, C., and Wilks, Y. (2011).
Affective Computing and Sentiment Analysis: Emotion, Metaphor and Terminology, volume 45.
Springer Science+ Business Media.

Barriere, V., Clavel, C., and Essid, S. (2017).
Opinion dynamics modeling for movie review transcripts classification with hidden conditional random fields.
Proc. Interspeech 2017, pages 1457–1461.

Clark, H. H. and Schaefer, E. F. (1989).
Contributing to discourse.
Cognitive science, 13(2):259–294.

Dutrey, C. (2014).
Analyse et détection automatique de disfluences dans la parole spontanée conversationnelle.
PhD thesis, Université Paris Sud-Paris XI.

Bibliography ii

Garcia, A., Essid, S., Clavel, C., and d'Alché Buc, F. (2018).
Structured output learning with abstention: Application to accurate opinion prediction.
In *Proceedings of the eighteenth international conference on machine learning, ICML*.

Janssoone, T. (2015).
Temporal association rules for modelling multimodal social signals.
In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 575–579. ACM.

Langlet, C. and Clavel, C. (2016).
Grounding the detection of the user's likes and dislikes on the topic structure of human-agent interactions.
Knowledge-Based Systems, 106:116–124.

Martin, J. R. and White, P. R. (2005).
The Language of Evaluation. Appraisal in English.
Palgrave Macmillan Basingstoke and New York.

Bibliography iii

Moilanen, K. and Pulman, S. (2007).
Sentiment composition.
In *Proceedings of the Recent Advances in Natural Language Processing International Conference*, pages 378–382.

Pulman, S., Boye, J., Cavazza, M., Smith, C., and De La Cámara, R. (2010).
'how was your day?'.
In *Proceedings of the 2010 Workshop on Companionable Dialogue Systems*, pages 37–42. Association for Computational Linguistics.

Quattoni, A., Wang, S., Morency, L.-P., Collins, M., and Darrell, T. (2007).
Hidden conditional random fields.
IEEE transactions on pattern analysis and machine intelligence, 29(10).

Ravenet, B., Pelachaud, C., Clavel, C., and Marsella, S. (2018).
Automating the production of communicative gestures in embodied characters.
Frontiers in Psychology, 9:1144.

Bibliography iv

Schuller, B., Ganascia, J.-G., and Devillers, L. (2016).
Multimodal sentiment analysis in the wild: Ethical considerations on data collection, annotation, and exploitation.
In *Actes du Workshop on Ethics In Corpus Collection, Annotation & Application (ETHI-CA2), LREC, Portoroz, Slovénie*.

Smith, C., Crook, N., Dobnik, S., and Charlton, D. (2011).
Interaction strategies for an affective conversational agent.
In *Presence: Teleoperators and Virtual Environments*, volume 20, pages 395–411. MIT Press.