

## SEC : contexte émotionnel phrasistique intégré pour la reconnaissance émotionnelle efficiente dans la conversation

Barbara Gendron-Audebert<sup>1,2</sup> et Gaël Guibon<sup>1</sup>  
{prénom.nom}@loria.fr

(1) LORIA, Université de Lorraine, CNRS      (2) Université du Luxembourg

# *Metric learning pour la reconnaissance d'émotions en contexte*

**Support**  
(données d'entraînement)



**Méta-représentation**



**Requête**  
(jeu de test)



# Objectifs

## Détection et identification des émotions dans le contenu généré par les utilisateurs



Cadre de l'étude :

Lien vers l'article

- Dialogues sous forme de conversations dyadiques
- Reconnaissance d'Émotions en Conversation (ERC)

# Objectifs

## Détection et identification des émotions dans le contenu généré par les utilisateurs



Cadre de l'étude :

Lien vers l'article

- Dialogues sous forme de conversations dyadiques
- Reconnaissance d'Émotions en Conversation (ERC)

Questions de recherche :

- **RQ1** : Comment utiliser l'information provenant du contexte conversationnel pour guider la détection d'émotions en conversation ?
- **RQ2** : Est-ce que la prise en compte du contexte conversationnel permet d'améliorer la détection d'émotions en conversation dans le cas dyadique ?

# Travaux connexes

- Tâche traditionnellement évaluée en microF1, **de plus en plus en macroF1**

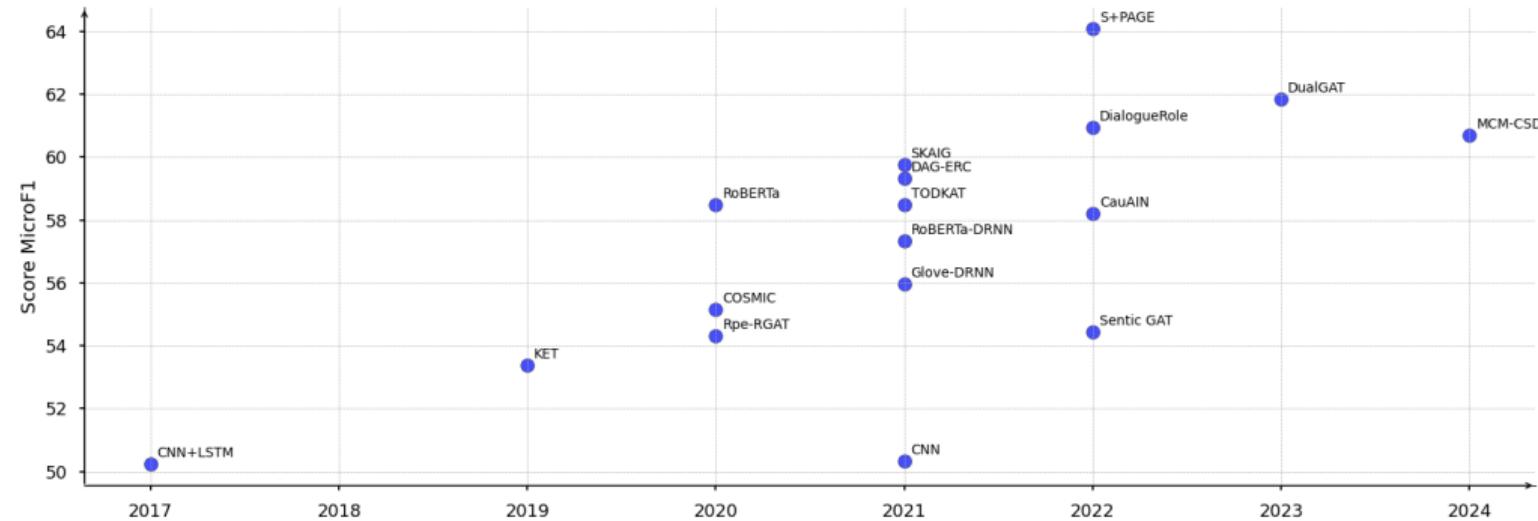
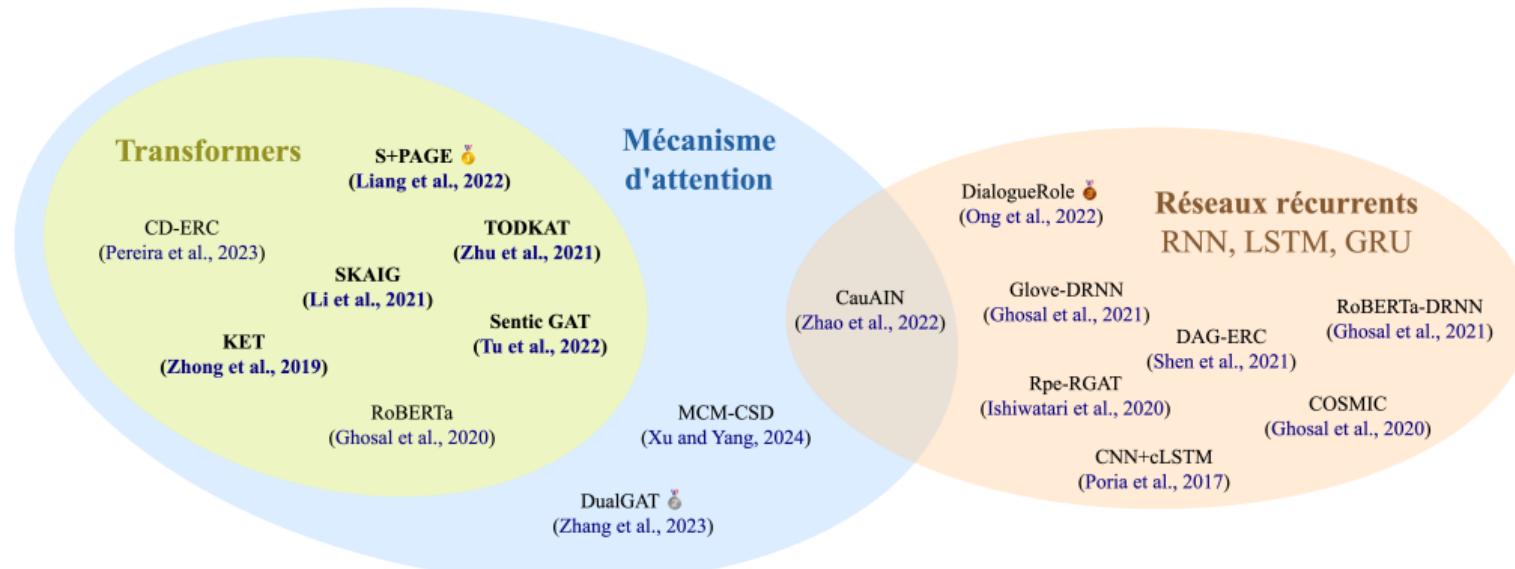


Figure 2 – Modèles état-de-l'art en ERC suivant le microF1 sur les données textuelles de DailyDialog (6 étiquettes émotionnelles)

# Travaux connexes

Figure 3 – Vue d'ensemble des architectures de modèle utilisées par les modèles état-de-l'art en ERC.  
Les modèles **en gras** intègrent des graphes de connaissances.



# Apports du *deep learning* et du *metric learning*

Les modèles neuronaux obtiennent des résultats état-de-l'art en ERC. [Poria et al., 2019, Pereira et al., 2022]

# Apports du *deep learning* et du *metric learning*

Les modèles neuronaux obtiennent des résultats état-de-l'art en ERC. [Poria et al., 2019, Pereira et al., 2022]

Le *deep learning* par **apprentissage contrastif** permet :

- Un cadre de classification souple permettant l'extraction de relations entre les labels et la prédiction sur des labels inconnus
- Une adaptation intrinsèque à l'apprentissage *few-shot* pour apprendre des émotions peu représentées
- Un variété d'architectures : réseaux de correspondances [Vinyals et al., 2016], siamois [Koch et al., 2015], prototypiques [Snell et al., 2017, Guibon et al., 2021]

# Réseaux siamois

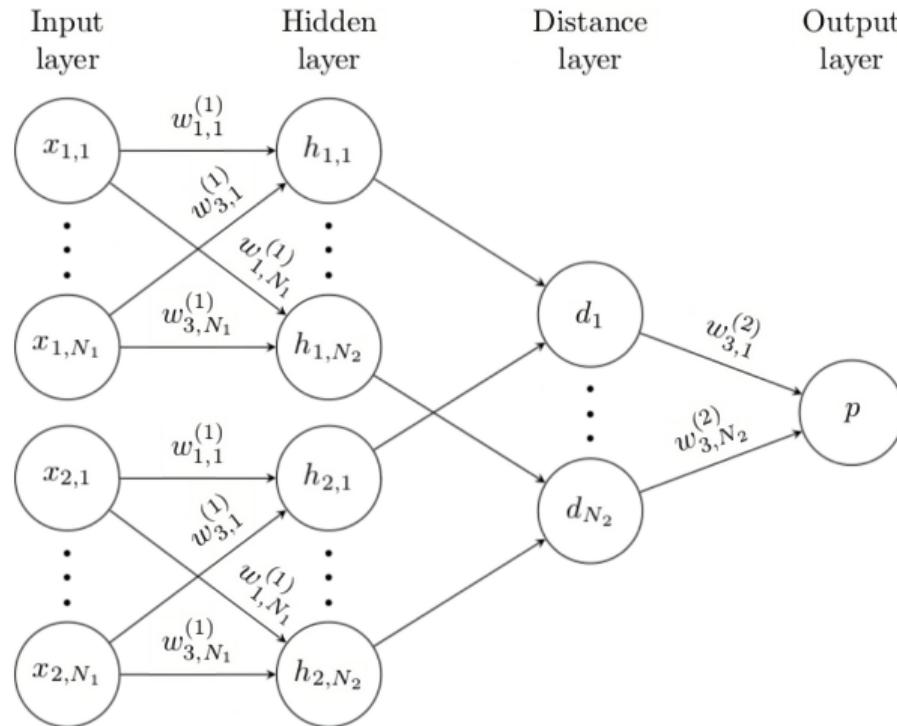


Figure 4 – Architecture des réseaux siamois [Koch et al., 2015]

# Réseaux siamois

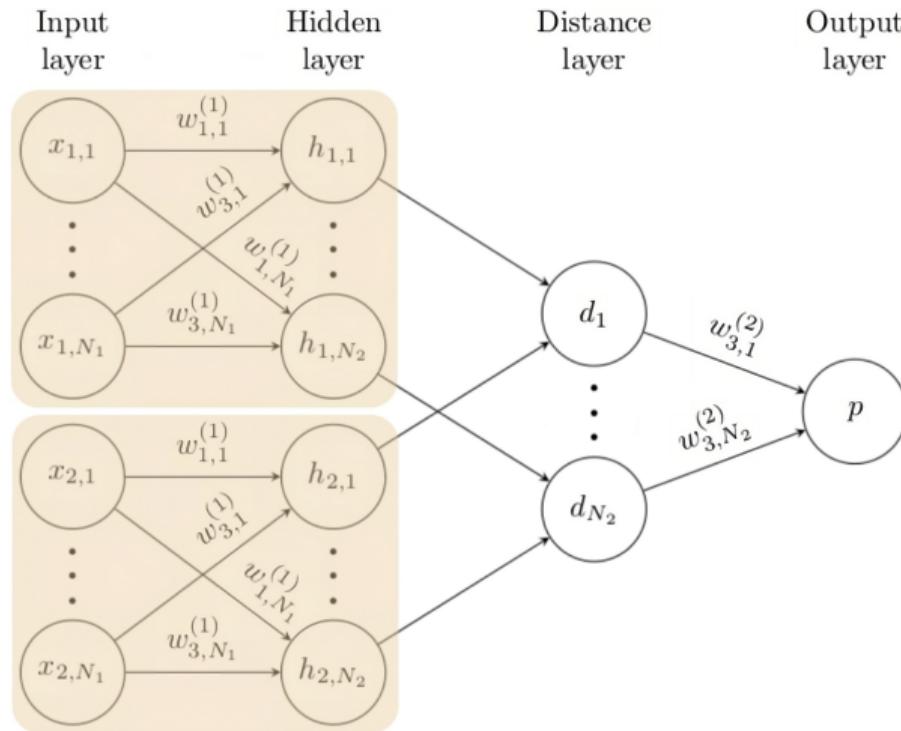


Figure 4 – Architecture des réseaux siamois [Koch et al., 2015]

# Réseaux siamois

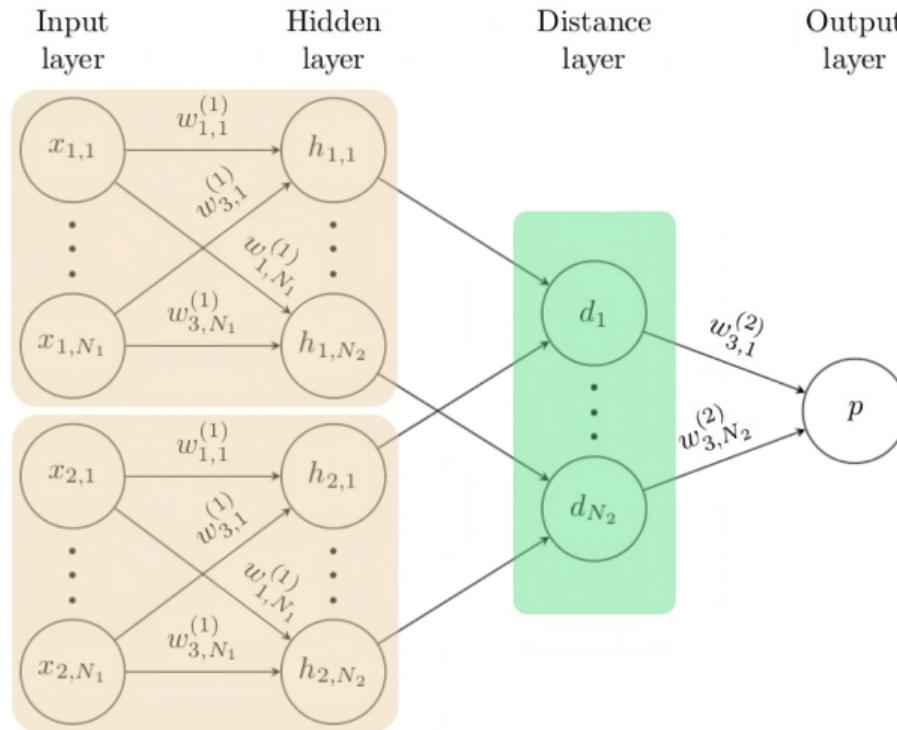


Figure 4 – Architecture des réseaux siamois [Koch et al., 2015]

# *Triplet loss* : fonction de coût par triplets

Un triplet de propos :

- Ancre ( $A$ )
- Positif ( $P$ )
- Négatif ( $N$ )

$A$  et  $P$  sont de la même classe,  $N$  est d'une autre.

Objectif de la *triplet loss* :

- Minimiser  $d(A, P)$
- Maximiser  $d(A, N)$

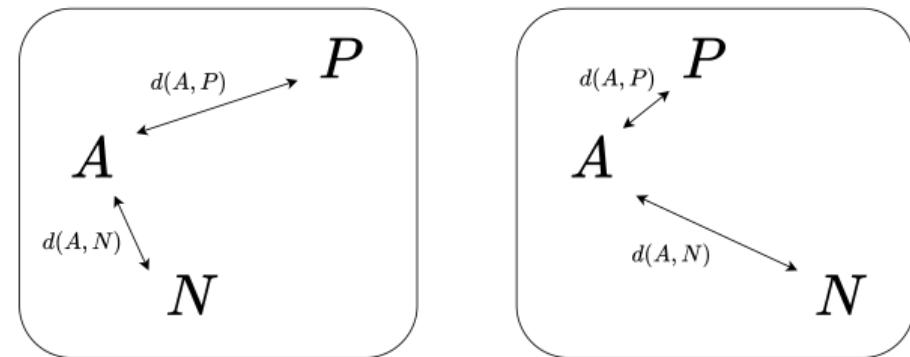
# *Triplet loss* : fonction de coût par triplets

Un triplet de propos :

- Ancre ( $A$ )
- Positif ( $P$ )
- Négatif ( $N$ )

$A$  et  $P$  sont de la même classe,  $N$  est d'une autre.

$$\mathcal{L}(a, p, n) = \max \{d(a, p) - d(a, n) + \text{marge}, 0\}$$



Objectif de la *triplet loss* :

- Minimiser  $d(A, P)$
- Maximiser  $d(A, N)$

$$\begin{aligned} d(A, P) - d(A, N) &> 0 \\ \mathcal{L}(A, P, N) &> 0 \end{aligned}$$

$$\begin{aligned} d(A, P) - d(A, N) &< 0 \\ \mathcal{L}(A, P, N) &= 0 \end{aligned}$$

Figure 5 – Illustration du principe de la *triplet loss*

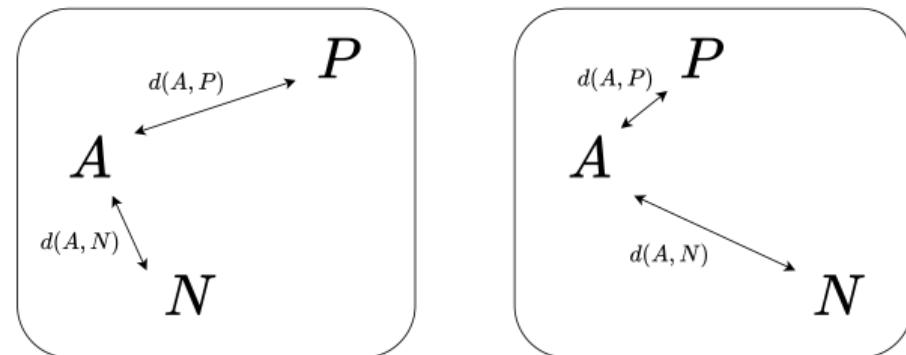
# *Triplet loss* : fonction de coût par triplets

Un triplet de propos :

- Ancre ( $A$ )
- Positif ( $P$ )
- Négatif ( $N$ )

$A$  et  $P$  sont de la même classe,  $N$  est d'une autre.

$$\mathcal{L}(a, p, n) = \max \{d(a, p) - d(a, n) + \text{marge}, 0\}$$



Objectif de la *triplet loss* :

- Minimiser  $d(A, P)$
- Maximiser  $d(A, N)$

$$\begin{aligned}d(A, P) - d(A, N) &> 0 \\ \mathcal{L}(A, P, N) &> 0\end{aligned}$$

$$\begin{aligned}d(A, P) - d(A, N) &< 0 \\ \mathcal{L}(A, P, N) &= 0\end{aligned}$$

Figure 5 – Illustration du principe de la *triplet loss*

# *Triplet loss* : fonction de coût par triplets

Un triplet de propos :

- Ancre ( $A$ )
- Positif ( $P$ )
- Négatif ( $N$ )

$A$  et  $P$  sont de la même classe,  $N$  est d'une autre.

Objectif de la *triplet loss* :

- Minimiser  $d(A, P)$
- Maximiser  $d(A, N)$

$$\mathcal{L}(a, p, n) = \max \{d(a, p) - d(a, n) + \text{marge}, 0\}$$

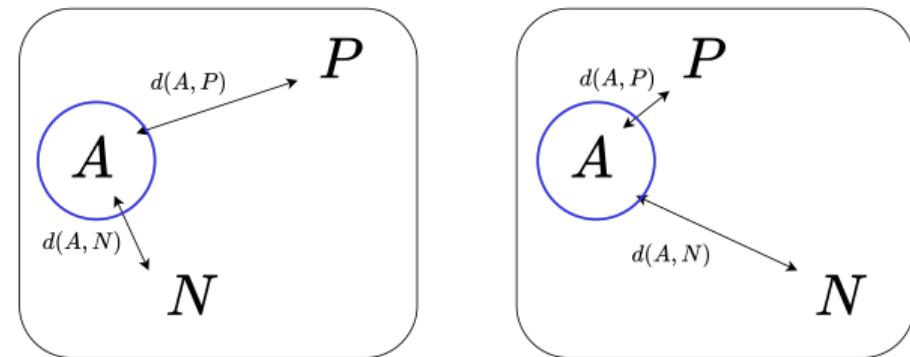


Figure 5 – Illustration du principe de la *triplet loss*

# Données

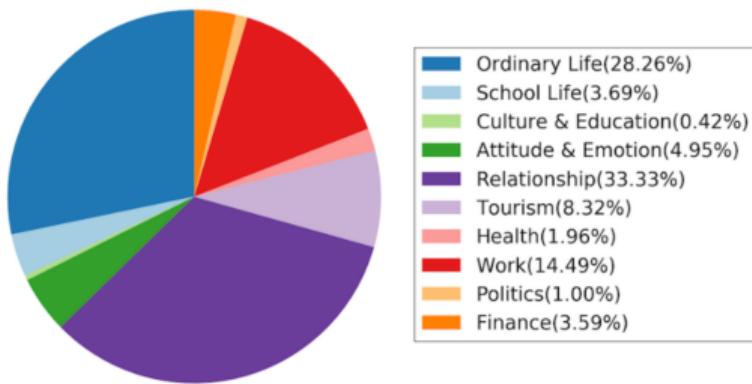
Jeu de données DailyDialog [Li et al., 2017]

- 13 118 dialogues dyadiques en anglais sur des sujets de la vie quotidienne
- Annotation **au niveau du tour de parole** : happiness, anger, disgust, fear, surprise, sadness et no emotion

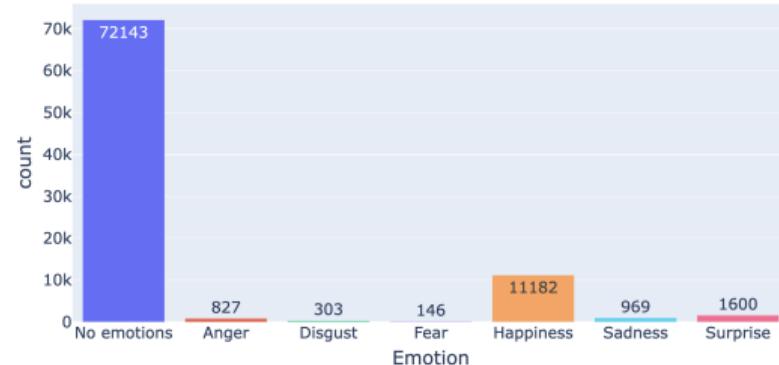
# Données

Jeu de données DailyDialog [Li et al., 2017]

- 13 118 dialogues dyadiques en anglais sur des sujets de la vie quotidienne
- Annotation **au niveau du tour de parole** : happiness, anger, disgust, fear, surprise, sadness et no emotion



(a) Répartition des sujets des dialogues



(b) Distribution des émotions dans les données d'entraînement

# Procédure d'entraînement - prédictions d'émotions avec contexte

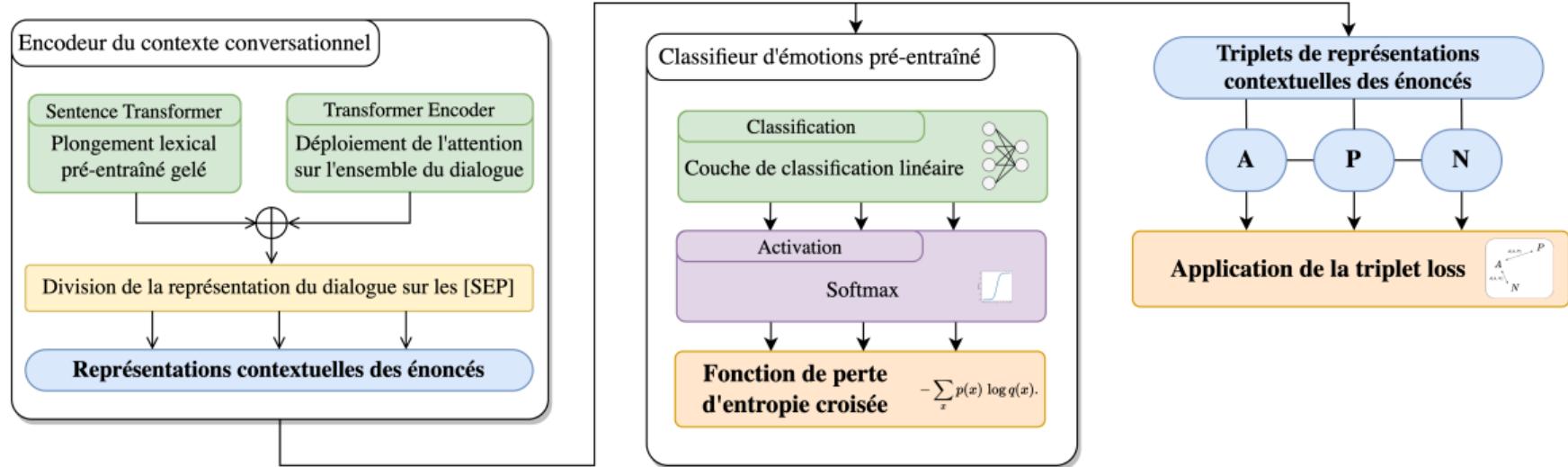


Figure 7 – Prédiction d'émotions sur des représentations de propos contextuelles.

# Métriques d'évaluation

- MicroF1 : choix historiquement privilégié par la littérature
- MacroF1 : métrique plus exigeante qui favorise une **reconnaissance émotionnelle polyvalente**
- MCC (*Matthews Correlation Coefficient*) : métrique plus contraignante utilisée pour des classes fortement déséquilibrées

# Métriques d'évaluation

- MicroF1 : choix historiquement privilégié par la littérature
- MacroF1 : métrique plus exigeante qui favorise une **reconnaissance émotionnelle polyvalente**
- MCC (*Matthews Correlation Coefficient*) : métrique plus contraignante utilisée pour des classes fortement déséquilibrées

Le MCC est défini comme suit [Matthews, 1975] :

$$\text{MCC} = \frac{TP/N - S \times P}{\sqrt{PS(1 - S)(1 - P)}} \quad (1)$$

Où  $TP$  est le nombre de vrais positifs,  $N$  la taille des données,  $P$  la précision et  $S$  le rappel (pour *sensitivity*).

# Résultats quantitatifs

Nom du modèle	macroF1*	microF1*	MCC
Modèles état-de-l'art en ERC			
CNN+cLSTM (Poria <i>et al.</i> , 2017)	–	50.24	–
KET (Zhong <i>et al.</i> , 2019)	–	53.37	–
COSMIC (Ghosal <i>et al.</i> , 2020)	51.05	58.48	–
RoBERTa (Ghosal <i>et al.</i> , 2020)	48.20	55.16	–
Rpe-RGAT (Ishiwatari <i>et al.</i> , 2020)	–	54.31	–
Glove-DRNN (Ghosal <i>et al.</i> , 2021)	41.80	55.95	–
roBERTa-DRNN (Ghosal <i>et al.</i> , 2021)	49.65	57.32	–
CNN (Ghosal <i>et al.</i> , 2021)	36.87	50.32	–
DAG-ERC (Shen <i>et al.</i> , 2021)	–	59.33	–
TODKAT (Zhu <i>et al.</i> , 2021)	<u>52.56</u>	58.47	–
SKAIG (Li <i>et al.</i> , 2021)	51.95	59.75	–
Sentic GAT (Tu <i>et al.</i> , 2022)	–	54.45	–
CauAIN (Zhao <i>et al.</i> , 2022)	–	58.21	–
DialogueRole (Ong <i>et al.</i> , 2022)	–	60.95	–
S+PAGE (Liang <i>et al.</i> , 2022)	–	<b>64.07</b>	–
DualGAT (Zhang <i>et al.</i> , 2023)	–	<u>61.84</u>	–
CD-ERC (Pereira <i>et al.</i> , 2023)	51.23	–	–
MCM-CSD (Xu & Yang, 2024)	–	60.70	–
Notre approche			
SentEmoContext	<b>57.71</b>	57.75	<b>0.49</b>

Figure 8 – Résultats en ERC sur DailyDialog

# Évaluation qualitative

## • Impact de l'information contextuelle

===== DIALOG #125 =====

- May I know where you are going ?
- Yes . I want to go to Beijing Hotel .
- I'm sorry . You are going in the wrong direction .
- Oh no ! What shall I do ?
- Don't worry . You can get off at the next stop and go across the street through the overpass . The bus stop is right there .
- Thank you very much .
- My pleasure .

Utterance: Oh no ! What shall I do ?

Emotion: no emo. Prediction: no emo.

===== DIALOG #354 =====

- Hit ' em high , hit ' em low . Class of ' 93 – let's go !
- Hi there , everyone . We hope you're having a good night !
- Wasn't that football game great ! I just knew we'd win !
- The night is young , folks . Get some food and mingle with those faces from yesterday .
- Later we'll let you know who the King and Queen of the Reunion will be .
- But for now , the band is playing the songs from our senior year . Get out on that dance floor !

Utterance: Wasn't that football game great ! I just knew we'd win !

Emotion: happiness Prediction: no emo.

Figure 9 – Deux exemples de prédictions sur des dialogues où le contexte semble être pris en compte (à gauche) et ne pas être utilisé (à droite)

# Évaluation qualitative

## ● Pertinence des prédictions

===== DIALOG #461 =====

- I'm a little nervous .
- Don't worry . You'll be fine . First of all , put on your seat belt . Adjust the mirrors .
- You don't think I'll need the seat belt , do you ?
- Of course not . But it's a good habit to put it on every time you drive .
- Just in case , right ?
- Right . Hold the steering wheel with your hands at ten o'clock and two o'clock .

Utterance: Just in case , right ?

Emotion: no emo. Prediction: fear X

===== DIALOG #601 =====

- Do you have anything to do after this ?
- No , I don't .
- Shall we drop in somewhere for a couple of drinks ?
- That sounds like a good idea .
- I know a very interesting place .
- Oh , do you ? Good .

Utterance: I know a very interesting place .

Emotion: happiness Prediction: no emo. X

Figure 10 – Un exemple de dialogue où une prédition incorrecte paraît légitime

# Comparaison avec les LLMs

Nom du modèle	macroF1*	microF1*	MCC
LLMs			
Llama2-7b (Touvron <i>et al.</i> , 2023)	09.70	24.92	0.08
Llama2-13b (Touvron <i>et al.</i> , 2023)	22.26	43.37	0.15
Falcon-7b (Penedo <i>et al.</i> , 2023)	07.54	42.75	0.01
Notre approche			
SentEmoContext	<b>57.71</b>	57.75	<b>0.49</b>

Figure 11 – Résultats avec les LLMs et comparaison avec SentEmoContext

	Transformers		LLMs			Notre approche
Modèle	MiniLM	MPNet	Llama2-7b	Llama2-13b	Falcon-7b	SentEmoContext
Tokens	1bn+	1bn+	2T	2T	1.5T	4M
Taille	80 MB	420 MB	13 GB	25 GB	15 GB	604,8 MB
Paramètres	22M	110M	7B	13B	7B	157M

Figure 12 – Aperçu de la taille des modèles et comparaison avec SentEmoContext

# Conclusion

**RQ1 :** Comment utiliser l'information provenant du contexte conversationnel pour guider la détection d'émotions en conversation ?

- Déployer de l'attention à l'échelle du dialogue
- Modification des représentations des propos

# Conclusion

**RQ1 :** Comment utiliser l'information provenant du contexte conversationnel pour guider la détection d'émotions en conversation ?

- Déployer de l'attention à l'échelle du dialogue
- Modification des représentations des propos

**RQ2 :** Est-ce que la prise en compte du contexte conversationnel permet d'améliorer la détection d'émotions en conversation dans le cas dyadique ?

- Enjeu de la stabilité du modèle
- Déployer l'attention engendre du bruit et propage les émotions

# Perspectives

- Prédictions sur des étiquettes émotionnelles inconnues du modèle
- Enrichissement des données d'entraînement pour améliorer la stabilité du modèle (exemple : [EmpatheticDialogues](#) [Rashkin et al., 2019])

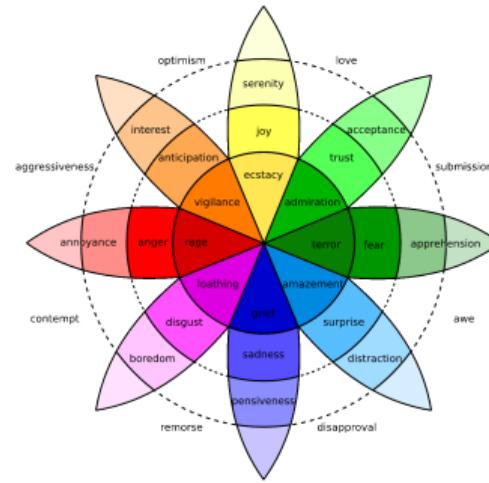


Figure 13 – Roue des émotions de Plutchik [Plutchik, 2001]

# Merci pour votre attention !



Naya



Tim



Chamallow



Feydo



Tsuki



Figure 15 – Code : <https://github.com/B-Gendron/sentEmoContext>

## Références I

- GUIBON, G. et al. *Few-Shot Emotion Recognition in Conversation with Sequential Prototypical Networks*. EMNLP. 2021.
- KOCH, G. et al. *Siamese Neural Networks for One-shot Image Recognition*. 2015.
- LI, Y. et al. *DailyDialog : A Manually Labelled Multi-turn Dialogue Dataset*. IJCNLP. Taipei, Taiwan, 2017.
- MATTHEWS, B. W. « Comparison of the predicted and observed secondary structure of T4 phage lysozyme. ». In : *Biochimica et biophysica acta* (1975).
- PEREIRA, P. et al. *Deep Emotion Recognition in Textual Conversations : A Survey*. 2022.
- PLUTCHIK, Robert. *The Nature of Emotions*. American Scientist. 2001.
- PORIA, S. et al. *Emotion Recognition in Conversation : Research Challenges, Datasets, and Recent Advances*. IEEE Access. 2019.
- RASHKIN, H. et al. *Towards Empathetic Open-domain Conversation Models : A New Benchmark and Dataset*. ACL. 2019.

## Références II

- | SNELL, J. et al. *Prototypical Networks for Few-Shot Learning*. NeurIPS. 2017.
- | SUNG, Flood et al. *Learning to Compare : Relation Network for Few-Shot Learning*.  
| IEEE/CVF. 2018.
- | VINYALS, O. et al. *Matching Networks for One Shot Learning*. NeurIPS. 2016.

## Animaux en page 2

- félin :  
<https://www.hachette.fr/livre/decouvre-le-monde-felins-9782017074717>
- canidés : <https://www.livres-medicaux.com/sante-tout-public/20544-canides-du-monde-loups-chiens-sauvages-renards-chacals-coyotes-et-a.html>
- loup arctique : <https://www.josephfiler.com/photo/canada-arctic-wolf-0348/>
- tigre : <https://evasion-online.com/tag/tigre>
- tigreau : <http://www.trouver-tout.fr/bebe-tigre-blanc/>
- caracal : <https://fr.wikipedia.org/wiki/Caracal>
- lionne :  
<https://www.futura-sciences.com/fonds-ecran/lion-lionne-lionceaux-1625/>
- renard : <http://eliotkitty.centerblog.net/27563-Magnifique-renard>

# About Classification Metrics

Given :

- $P$  the quantity of positive predictions
- $N$  the quantity of negative predictions
- $TP$ ,  $TN$ ,  $FP$  and  $FN$  the True Positives, True Negatives, False Positives and False Negatives

$$\text{Accuracy} = \frac{TP + TN}{P + N} \quad \text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad F_1 = \frac{2TP}{2TP + FP + FN} \quad (3)$$

## About Matthews Correlation Coefficient (MCC) (Cramér, 1946)

Given :

$$N = TN + TP + FN + FP , \quad S = \frac{TP + FN}{N} \quad \text{and} \quad P = \frac{TP + FP}{N} \quad (4)$$

MCC has been defined in (Matthews, 1975) as :

$$\text{MCC} = \frac{TP/N - S \times P}{\sqrt{PS(1-S)(1-P)}} \quad (5)$$

# Distribution of emotions within the dialog



Figure 16 – Cumulative number of expressed emotions *w.r.t.* the utterance index

# Prompt examples

Here is a dialog :

- Hello , Miao Li , Where are you going ?
- Hello , I am going to the store to buy some fruit .
- Oh , Would you do me a favor ?
- Yes ?
- Please mail this letter for me on your way to the store .
- Sure . Do you want it to be registered ?
- Yes , I think so . There are some pictures in it . It would be a great pity if they were lost .
- Yes , I will be glad to mail your letter .
- Thanks .
- you are welcome .

Regarding its conversational context, give me the appropriate emotion to describe this utterance : "Yes , I think so . There are some pictures in it . It would be a great pity if they were lost .", using only one of the following labels: happiness, sadness, anger, surprise, fear, disgust, no emotion. Predicted label :

(a) Requête utilisée pour Llama2

Here is a dialog :

- Hello , Miao Li , Where are you going ?
- Hello , I am going to the store to buy some fruit .
- Oh , Would you do me a favor ?
- Yes ?
- Please mail this letter for me on your way to the store .
- Sure . Do you want it to be registered ?
- Yes , I think so . There are some pictures in it . It would be a great pity if they were lost .
- Yes , I will be glad to mail your letter .
- Thanks .
- you are welcome .

Regarding its conversational context, return the appropriate emotion for the last utterance among: sadness, happiness, anger, surprise, fear and disgust. If none of them properly correspond, return 'no emotion'.

(b) Requête utilisée pour Falcon

FIGURE 2 – Requêtes pour Llama2 et Falcon

# Some unexplainable predictions

===== DIALOG #82 =====

- Excuse me . Check please .
- OK , how was everything ?
- Very nice . Thank you .
- Would you like this to-go ?
- Yes , can you put it in a plastic bag ?
- Sure , no problem . Here you are . That'll be 25 dollars .
- Do you take credit cards ?
- Yes , we accept Visa and MasterCard .
- OK , here you are .
- Thanks . I'll be right back .
- OK .
- Here's your receipt .
- Thank you .
- You're welcome . Please come again .

Utterance: OK .

Emotion: no emo. Prediction: happiness X

(a)

===== DIALOG #286 =====

- I need to get my high speed internet installed .
- You'll need to make an appointment .
- Could I do that right now , please ?
- What day would you like us to do the installation ?
- Is Friday good ?
- We're only available at 3
- You can't come any earlier than that ?
- I'm sorry . That's the only available time .
- Are you available this Saturday ?
- Yes . Anytime on Saturday will be fine .
- How does 11
- We can do it . See you then .

Utterance: How does 11

Emotion: no emo. Prediction: sadness X

(b)

Figure 4.11: Two examples of unexplainable predictions.

# Preprocessing Pipeline

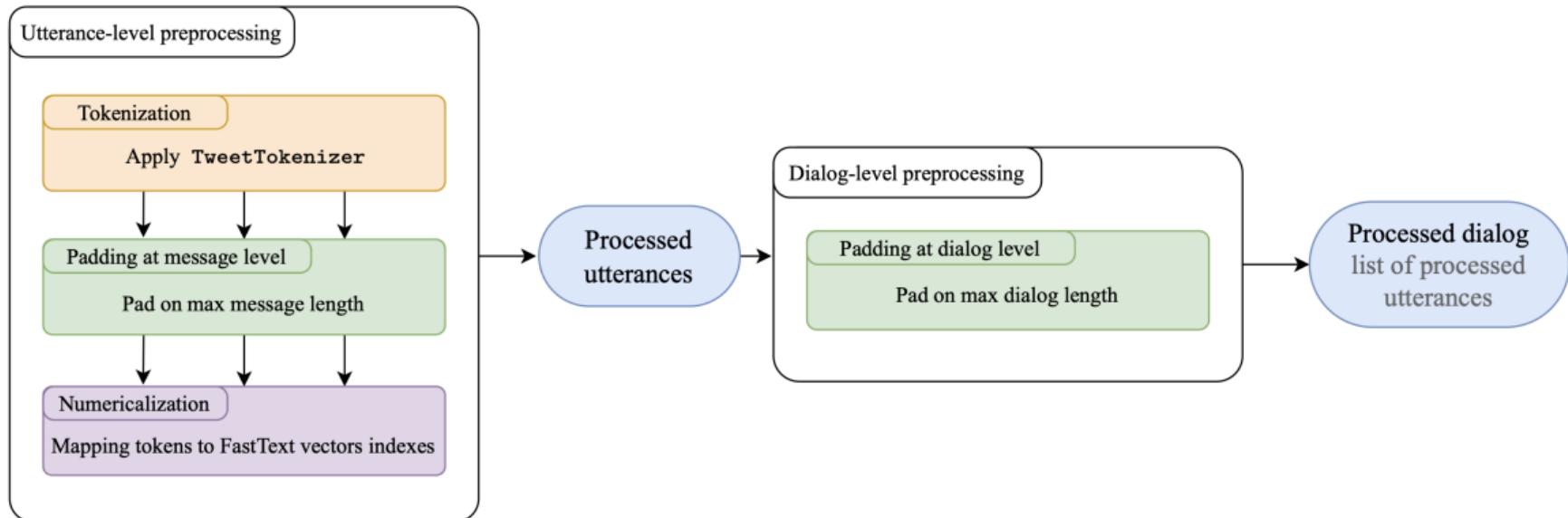


Figure 17 – Preprocessing steps for isolated utterance representations

# Preprocessing Pipeline

## Original dialog

— Say, Jim, how about going for a few beers after dinner ?  
— You know that is tempting but is really not good for our fitness.

## Formatted dialog with [SEP] tokens

Say , Jim , how about going for a few beers after dinner ? [SEP] You know that is tempting  
but is really not good for our fitness .

## Word-piece tokenization

```
['[CLS]', 'say', ',', 'jim', ',', 'how', 'about', 'going', 'for', 'a', 'few',
'beers', 'after', 'dinner', '?', '[SEP]', 'you', 'know', 'that', 'is', 'tempting',
'but', 'is', 'really', 'not', 'good', 'for', 'our', 'fitness', '.', '[PAD]',
..., '[PAD]']
```

## Numericalization

```
[101, 2360, 1010, 3958, 1010, 2129, 2055, 2183, 2005, 1037, 2261, 18007,
2044, 4596, 1029, 102, 2017, 2113, 2008, 2003, 23421, 2021, 2003, 2428, 2025,
2204, 2005, 2256, 10516, 1012, 0, ..., 0]
```

# Prediction Distributions - Isolated Utterances



No emotion



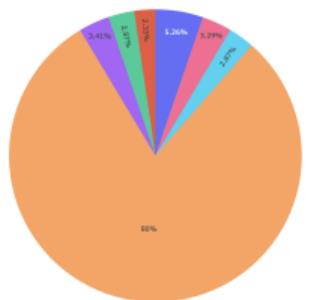
Anger



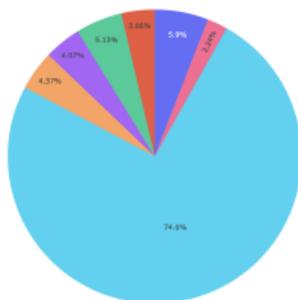
Disgust



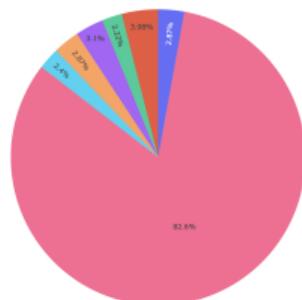
Fear



Happiness



Sadness

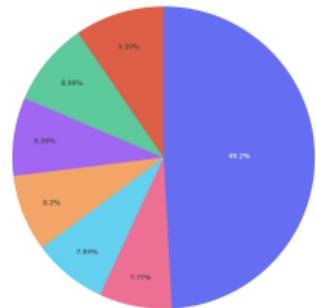


Surprise

Emotion

- No emotions
- Anger
- Disgust
- Fear
- Happiness
- Sadness
- Surprise

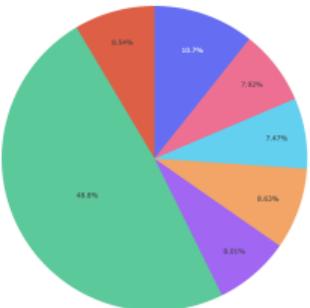
# Prediction Distributions - Contextual Utterances



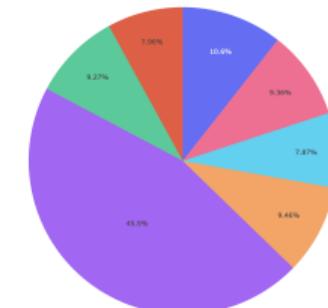
No emotion



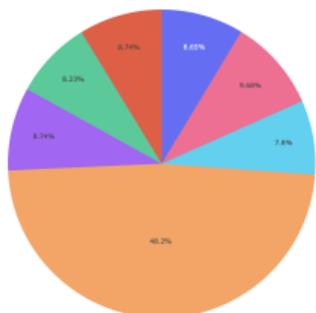
Anger



Disgust



Fear



Happiness



Sadness



Surprise

Emotion

- No emotions
- Anger
- Disgust
- Fear
- Happiness
- Sadness
- Surprise

## Performances - Isolated Utterances

Layers	Accuracy↑	Loss↓	Precision↑	Recall↑	wF1 score↑
LSTM-based models					
3	68.0	<b>0.706</b>	0.682	0.680	0.680
4	67.9	0.734	0.678	0.679	0.677
5	<b>70.0</b>	0.711	<b>0.702</b>	<b>0.700</b>	<b>0.700</b>
6	65.4	0.756	0.658	0.654	0.655
7	66.2	0.793	0.664	0.662	0.661
MLP-based models					
3	<b>68.4</b>	<b>0.747</b>	<b>0.685</b>	<b>0.684</b>	<b>0.684</b>
4	62.9	0.834	0.629	0.629	0.628
5	64.7	0.801	0.650	0.647	0.647

Table 4.1: Main results using Siamese Networks on static utterances representations.  
Best values are in **bold** and arrows indicates if greater or lower is better.