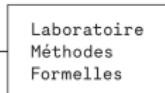


# Natural language processing for subjectivity analysis in personal narratives

Gustave Cortal

Thesis director: Alain Finkel  
Co-advisors: Patrick Paroubek and Lina Ye



école  
normale  
supérieure  
paris-saclay

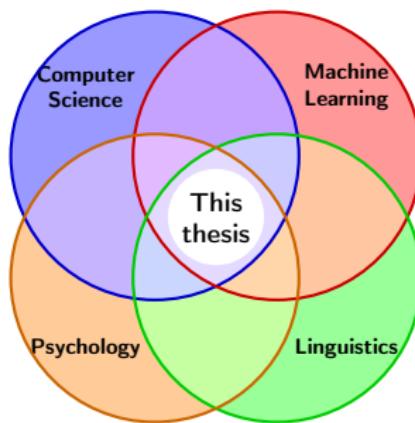
université  
PARIS-SACLAY

**LISN**  
LABORATOIRE INTERDISCIPLINAIRE  
DES SCIENCES DU NUMÉRIQUE



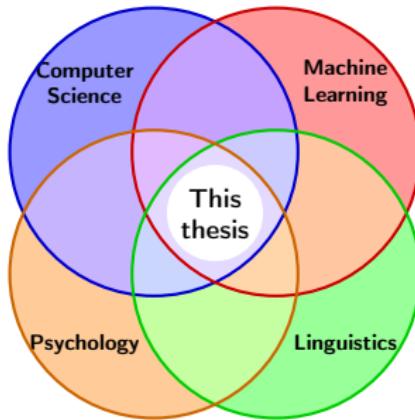
# Introduction

# Context



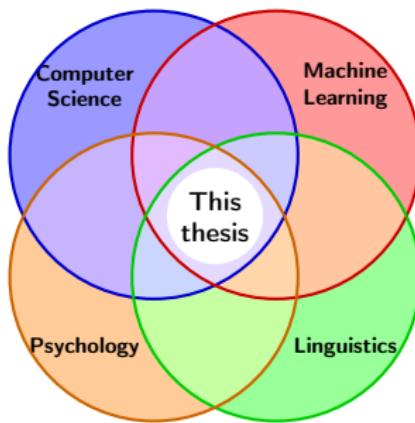
- ▶ Natural language processing for psychology is underexplored

# Context



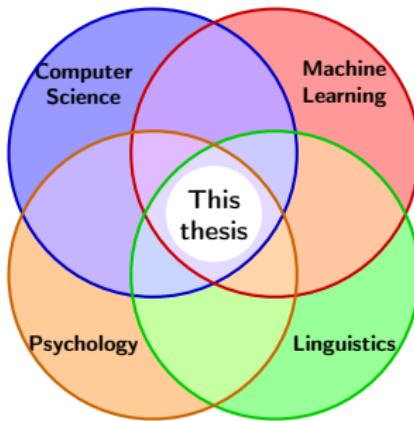
- ▶ Natural language processing for psychology is underexplored
- ▶ We build on an existing subfield: sentiment and emotion analysis

# Context



- ▶ Natural language processing for psychology is underexplored
- ▶ We build on an existing subfield: sentiment and emotion analysis
- ▶ We study subjectivity (e.g., first-person perspective, meaning-making processes, and experiential content)

# Context



- ▶ Natural language processing for psychology is underexplored
- ▶ We build on an existing subfield: sentiment and emotion analysis
- ▶ We study subjectivity (e.g., first-person perspective, meaning-making processes, and experiential content)
- ▶ We focus on personal narratives (emotional narratives, dream reports)

# Introduction

How to model subjective experience in personal narratives?

# Introduction

How to model subjective experience in personal narratives?

We first address the *content* by classifying elements of personal narratives, then the *form* through the concept of style

# Introduction

How to model subjective experience in personal narratives?

We first address the *content* by classifying elements of personal narratives, then the *form* through the concept of style

- ▶ Definition of objectives and scope using cognitive science

# Introduction

How to model subjective experience in personal narratives?

We first address the *content* by classifying elements of personal narratives, then the *form* through the concept of style

- ▶ Definition of objectives and scope using cognitive science
- ▶ Construction of an emotion dataset

# Introduction

How to model subjective experience in personal narratives?

We first address the *content* by classifying elements of personal narratives, then the *form* through the concept of style

- ▶ Definition of objectives and scope using cognitive science
- ▶ Construction of an emotion dataset
- ▶ Training of language models for emotion analysis

# Introduction

How to model subjective experience in personal narratives?

We first address the *content* by classifying elements of personal narratives, then the *form* through the concept of style

- ▶ Definition of objectives and scope using cognitive science
- ▶ Construction of an emotion dataset
- ▶ Training of language models for emotion analysis
- ▶ Formalization of style in personal narratives

# Introduction

How to model subjective experience in personal narratives?

We first address the *content* by classifying elements of personal narratives, then the *form* through the concept of style

- ▶ Definition of objectives and scope using cognitive science
- ▶ Construction of an emotion dataset
- ▶ Training of language models for emotion analysis
- ▶ Formalization of style in personal narratives

*2 international conferences (EMNLP, LREC-COLING); 3 international workshops; 2 national conferences and journals (TALN)*

## Definition of objectives using cognitive science

G. Cortal and C. Bonard. [Improving Language Models for Emotion Analysis: Insights from Cognitive Science](#). CMCL @ ACL 2024.

# Psychological theories and emotion annotation schemes

[add refs to klinger, etc.]

Psychological theories	In text, emotion is...	Example
Basic emotions theory	a <i>category</i>	"I love philosophy." → joy

# Psychological theories and emotion annotation schemes

[add refs to klinger, etc.]

Psychological theories	In text, emotion is...	Example
Basic emotions theory	a <i>category</i>	"I love philosophy." → joy
Constructivist theories	a continuous value with an <i>affective meaning</i>	"His voice soothes me." → valence (4/5), arousal (1/5)

# Psychological theories and emotion annotation schemes

[add refs to klinger, etc.]

Psychological theories	In text, emotion is...	Example
Basic emotions theory	a <i>category</i>	"I love philosophy." → joy
Constructivist theories	a continuous value with an <i>affective</i> meaning	"His voice soothes me." → valence (4/5), arousal (1/5)
Appraisal theory	a continuous value with a <i>cognitive</i> meaning	"I received a surprise gift." → sudden (4/5), control (0/5)

# Psychological theories and emotion annotation schemes

[add refs to klinger, etc.]

Psychological theories	In text, emotion is...	Example
Basic emotions theory	a <i>category</i>	"I love philosophy." → joy
Constructivist theories	a continuous value with an <i>affective</i> meaning	"His voice soothes me." → valence (4/5), arousal (1/5)
Appraisal theory	a continuous value with a <i>cognitive</i> meaning composed of <i>semantic roles</i>	"I received a surprise gift." → sudden (4/5), control (0/5)  "Louise (experiencer) was angry (cue) towards Paul (target), because he didn't inform her (cause)."

## Limitations in emotion analysis

- ▶ Different emotion theories lead to divergences in how to annotate them in the text [add refs]

## Limitations in emotion analysis

- ▶ Different emotion theories lead to divergences in how to annotate them in the text [add refs]
- ▶ Some linguistic and cognitive science theories are not considered [reformulate, not some]

## Limitations in emotion analysis

- ▶ Different emotion theories lead to divergences in how to annotate them in the text [add refs]
- ▶ Some linguistic and cognitive science theories are not considered [reformulate, not some]
- ▶ There is no benchmark that evaluates the richness of the emotional phenomenon [add refs]

## Linguistic and cognitive science theories

## Which verbal signs are used to infer expressed emotions?

Raphaël Micheli categorizes a range of linguistic markers into three *emotion expression modes* (Micheli, 2013). The emotion can be:

## Which verbal signs are used to infer expressed emotions?

Raphaël Micheli categorizes a range of linguistic markers into three *emotion expression modes* (Micheli, 2013). The emotion can be:

- ▶ *labeled* explicitly with an emotional term ("I am sad")

## Which verbal signs are used to infer expressed emotions?

Raphaël Micheli categorizes a range of linguistic markers into three *emotion expression modes* (Micheli, 2013). The emotion can be:

- ▶ *labeled* explicitly with an emotional term ("I am sad")
- ▶ *shown* with utterance features such as interjections and punctuations ("Ah! That's great!")

## Which verbal signs are used to infer expressed emotions?

Raphaël Micheli categorizes a range of linguistic markers into three *emotion expression modes* (Micheli, 2013). The emotion can be:

- ▶ *labeled* explicitly with an emotional term ("I am sad")
- ▶ *shown* with utterance features such as interjections and punctuations ("Ah! That's great!")
- ▶ *suggested* with the description of a situation which generally, in a given sociocultural context, leads to an emotion ("She gave me a gift")

## Which verbal signs are used to infer expressed emotions?

Raphaël Micheli categorizes a range of linguistic markers into three *emotion expression modes* (Micheli, 2013). The emotion can be:

- ▶ *labeled* explicitly with an emotional term ("I am sad")
- ▶ *shown* with utterance features such as interjections and punctuations ("Ah! That's great!")
- ▶ *suggested* with the description of a situation which generally, in a given sociocultural context, leads to an emotion ("She gave me a gift")

## Which verbal signs are used to infer expressed emotions?

Raphaël Micheli categorizes a range of linguistic markers into three *emotion expression modes* (Micheli, 2013). The emotion can be:

- ▶ *labeled* explicitly with an emotional term ("I am sad")
  - ▶ *shown* with utterance features such as interjections and punctuations ("Ah! That's great!")
  - ▶ *suggested* with the description of a situation which generally, in a given sociocultural context, leads to an emotion ("She gave me a gift")
- Different emotion expression modes are more or less difficult to interpret [add refs psycholinguistic, psychiatry, refs aline etienne]

## Which verbal signs are used to infer expressed emotions?

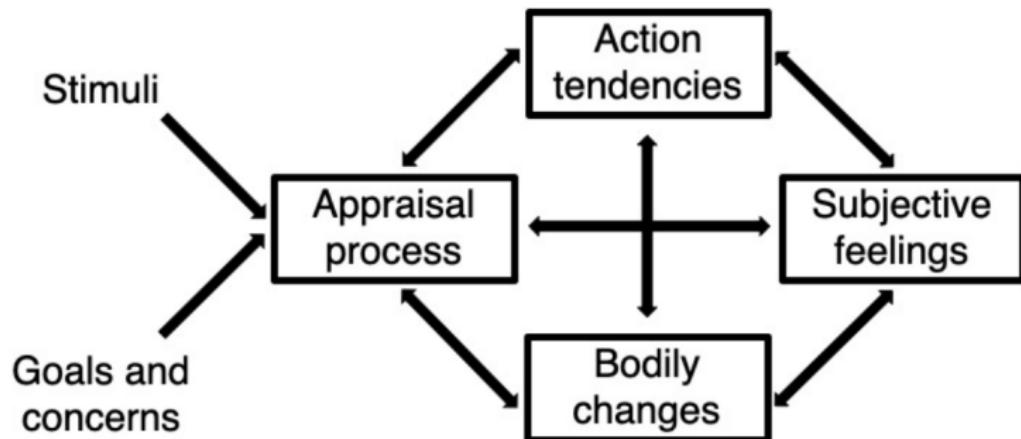
Raphaël Micheli categorizes a range of linguistic markers into three *emotion expression modes* (Micheli, 2013). The emotion can be:

- ▶ *labeled* explicitly with an emotional term ("I am sad")
- ▶ *shown* with utterance features such as interjections and punctuations ("Ah! That's great!")
- ▶ *suggested* with the description of a situation which generally, in a given sociocultural context, leads to an emotion ("She gave me a gift")

→ Different emotion expression modes are more or less difficult to interpret [add refs psycholinguistic, psychiatry, refs aline etienne]

→ [add refs aline etienne]

# How to integrate psychological theories of emotion?



**Figure: Integrated framework for emotion theories.** Emotional episodes are synchronized changes in four components (Scherer, 2022).

→ There is no dataset of personal narratives structured according to emotion components [add refs]

## Construction of an emotion dataset

G. Cortal, A. Finkel, P. Paroubek, L. Ye. [Emotion Recognition based on Psychological Components in Guided Narratives for Emotion Regulation.](#)  
*SIGHUM @ EACL 2023.*

# French emotional narratives based on components

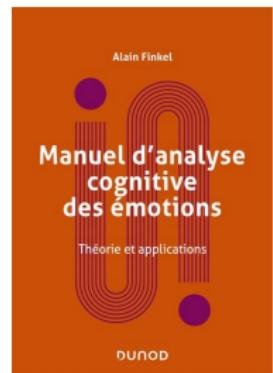
**Goal:** Create a dataset of narratives structured according to emotion components, following the *Cognitive Analysis of Emotion* ([finkel2022](#))

# French emotional narratives based on components

**Goal:** Create a dataset of narratives structured according to emotion components, following the *Cognitive Analysis of Emotion* ([finkel2022](#))

The questionnaire:

- ▶ explores emotions with behavioral (*behavior*), physiological (*feeling*), and cognitive (*thinking* and *territory*) components



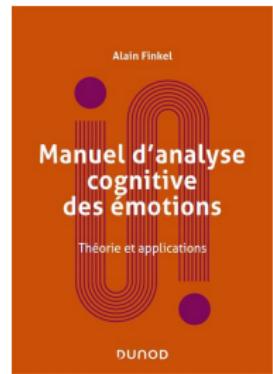
[finkel2022](#), april  
[finkel2022](#)

# French emotional narratives based on components

**Goal:** Create a dataset of narratives structured according to emotion components, following the *Cognitive Analysis of Emotion* ([finkel2022](#))

The questionnaire:

- ▶ explores emotions with behavioral (*behavior*), physiological (*feeling*), and cognitive (*thinking* and *territory*) components
- ▶ uses emotion components to reorganize the narrative of experienced events



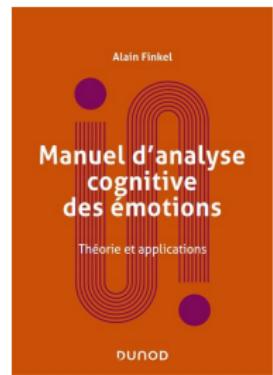
[finkel2022](#), april  
[finkel2022](#)

# French emotional narratives based on components

**Goal:** Create a dataset of narratives structured according to emotion components, following the *Cognitive Analysis of Emotion* ([finkel2022](#))

The questionnaire:

- ▶ explores emotions with behavioral (*behavior*), physiological (*feeling*), and cognitive (*thinking* and *territory*) components
- ▶ uses emotion components to reorganize the narrative of experienced events
- ▶ helps individuals better regulate their emotions



[finkel2022](#), april  
[finkel2022](#)

## French emotional narratives based on components

**Contribution:** 1000 narratives from emotion regulation sessions, providing a more comprehensive understanding of emotional events

# French emotional narratives based on components

**Contribution:** 1000 narratives from emotion regulation sessions, providing a more comprehensive understanding of emotional events

Component	Answer
Behavior	I'm giving a lecture on a Friday morning at 8:30. A student goes out and comes back a few moments later with a coffee in his hand.
Feeling	My heart is beating fast, and I freeze, waiting to know how to act.
Thinking	I think this student is disrupting my class.
Territory	The student attacks my ability to be respected in class.

Chosen emotion: anger (possible choices: anger, fear, joy, sadness)

# Training language models for emotion analysis

**G. Cortal**, A. Finkel, P. Paroubek, L. Ye. [Emotion Recognition based on Psychological Components in Guided Narratives for Emotion Regulation](#). *SIGHUM @ EACL 2023*

**G. Cortal**. [Sequence-to-Sequence Language Models for Character and Emotion Detection in Dream Narratives](#). *LREC-COLING 2024*

## Discrete emotion detection based on components

Component	Logistic Regression			CamemBERT		
	Precision	Recall	$F_1$	Precision	Recall	$F_1$
All	71.2 (2.6)	69.1 (2.2)	67.8 (2.3)	<b>85.1</b>	<b>84.8</b>	<b>84.7</b>

## Discrete emotion detection based on components

Component	Logistic Regression			CamemBERT		
	Precision	Recall	$F_1$	Precision	Recall	$F_1$
All	71.2 (2.6)	69.1 (2.2)	67.8 (2.3)	<b>85.1</b>	<b>84.8</b>	<b>84.7</b>
Without behavior	77.4 (2.3)	75.8 (2.4)	74.5 (2.6)	80.3	79.8	79.7
Without feeling	64.3 (1.9)	61.5 (1.2)	61.3 (2.2)	81.6	79.8	79.9
Without thinking	70.9 (1.8)	69.1 (2.0)	68.3 (2.2)	79.6	78.5	78.7
Without territory	64.3 (4.1)	64.5 (2.4)	62.3 (2.8)	78.7	78.5	78.6

# Discrete emotion detection based on components

Component	Logistic Regression			CamemBERT		
	Precision	Recall	$F_1$	Precision	Recall	$F_1$
All	71.2 (2.6)	69.1 (2.2)	67.8 (2.3)	<b>85.1</b>	<b>84.8</b>	<b>84.7</b>
Without behavior	77.4 (2.3)	75.8 (2.4)	74.5 (2.6)	80.3	79.8	79.7
Without feeling	64.3 (1.9)	61.5 (1.2)	61.3 (2.2)	81.6	79.8	79.9
Without thinking	70.9 (1.8)	69.1 (2.0)	68.3 (2.2)	79.6	78.5	78.7
Without territory	64.3 (4.1)	64.5 (2.4)	62.3 (2.8)	78.7	78.5	78.6
Only behavior	52.1 (3.5)	54.6 (2.9)	51.7 (2.9)	68.4	67.1	66.6
Only feeling	69.6 (1.5)	68.9 (2.1)	68.4 (2.0)	67.8	68.4	67.7
Only thinking	50.1 (3.4)	53.8 (2.3)	50.6 (2.7)	70.5	70.1	70.1
Only territory	68.2 (1.8)	66.8 (2.2)	66.6 (2.3)	71.4	68.4	68.9

## Discrete emotion detection based on components

Component	Logistic Regression			CamemBERT		
	Precision	Recall	$F_1$	Precision	Recall	$F_1$
All	71.2 (2.6)	69.1 (2.2)	67.8 (2.3)	<b>85.1</b>	<b>84.8</b>	<b>84.7</b>
Without behavior	77.4 (2.3)	75.8 (2.4)	74.5 (2.6)	80.3	79.8	79.7
Without feeling	64.3 (1.9)	61.5 (1.2)	61.3 (2.2)	81.6	79.8	79.9
Without thinking	70.9 (1.8)	69.1 (2.0)	68.3 (2.2)	79.6	78.5	78.7
Without territory	64.3 (4.1)	64.5 (2.4)	62.3 (2.8)	78.7	78.5	78.6
Only behavior	52.1 (3.5)	54.6 (2.9)	51.7 (2.9)	68.4	67.1	66.6
Only feeling	69.6 (1.5)	68.9 (2.1)	68.4 (2.0)	67.8	68.4	67.7
Only thinking	50.1 (3.4)	53.8 (2.3)	50.6 (2.7)	70.5	70.1	70.1
Only territory	68.2 (1.8)	66.8 (2.2)	66.6 (2.3)	71.4	68.4	68.9

→ Each component improves prediction performance, the best results are achieved by jointly considering all components

## Discrete emotion detection based on components

Component	Logistic Regression			CamemBERT		
	Precision	Recall	$F_1$	Precision	Recall	$F_1$
All	71.2 (2.6)	69.1 (2.2)	67.8 (2.3)	<b>85.1</b>	<b>84.8</b>	<b>84.7</b>
Without behavior	77.4 (2.3)	75.8 (2.4)	74.5 (2.6)	80.3	79.8	79.7
Without feeling	64.3 (1.9)	61.5 (1.2)	61.3 (2.2)	81.6	79.8	79.9
Without thinking	70.9 (1.8)	69.1 (2.0)	68.3 (2.2)	79.6	78.5	78.7
Without territory	64.3 (4.1)	64.5 (2.4)	62.3 (2.8)	78.7	78.5	78.6
Only behavior	52.1 (3.5)	54.6 (2.9)	51.7 (2.9)	68.4	67.1	66.6
Only feeling	69.6 (1.5)	68.9 (2.1)	68.4 (2.0)	67.8	68.4	67.7
Only thinking	50.1 (3.4)	53.8 (2.3)	50.6 (2.7)	70.5	70.1	70.1
Only territory	68.2 (1.8)	66.8 (2.2)	66.6 (2.3)	71.4	68.4	68.9

- Each component improves prediction performance, the best results are achieved by jointly considering all components
- Some components benefit from contextual understanding and world knowledge (e.g., behavior and thinking)

## Quantitative analysis of dream narratives

[add illustration dreambank, hdvc scheme]

Need other datasets with narrative structure, emotional content, and available for research

## Quantitative analysis of dream narratives

[add illustration dreambank, hdvc scheme]

Need other datasets with narrative structure, emotional content, and available for research

Quantitative dream analysis most relies on dream databases (e.g., DreamBank composed of 27,000 dreams) and annotation schemes (e.g., Hall and Van de Castle system) (Domhoff and Schneider, 2008)

## Quantitative analysis of dream narratives

[add illustration dreambank, hdvc scheme]

Need other datasets with narrative structure, emotional content, and available for research

Quantitative dream analysis most relies on dream databases (e.g., DreamBank composed of 27,000 dreams) and annotation schemes (e.g., Hall and Van de Castle system) (Domhoff and Schneider, 2008)

The annotation process is complex and costly

# Quantitative analysis of dream narratives

[add illustration dreambank, hdvc scheme]

Need other datasets with narrative structure, emotional content, and available for research

Quantitative dream analysis most relies on dream databases (e.g., DreamBank composed of 27,000 dreams) and annotation schemes (e.g., Hall and Van de Castle system) (Domhoff and Schneider, 2008)

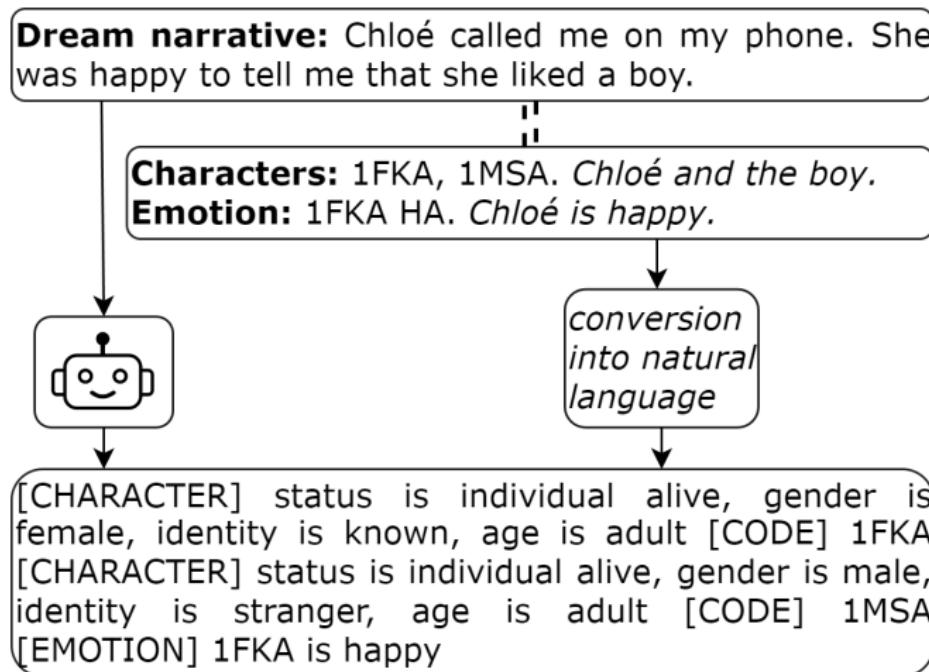
The annotation process is complex and costly

How to automate the annotation process?

## Hall and Van de Castle annotation scheme

# Character and emotion detection in dream narratives

[add seq2seq architecture illustration, add list of characters and emotions]



## Results

Baseline is LaMini-Flan-T5 finetuned on 1823 dream narratives

Model	Status	Gender	Identity	Age	Character	Emotion
Baseline	82.87	78.02	76.17	86.21	64.74	75.13

**Table: Character and emotion detection.** \*\* ( $p < 0.01$ ), \* ( $p < 0.05$ ).

## Results

Baseline is LaMini-Flan-T5 finetuned on 1823 dream narratives

Model	Status	Gender	Identity	Age	Character	Emotion
Baseline	82.87	78.02	76.17	86.21	64.74	75.13
No <sub>semantics</sub>	71.37	56.54*	61.0	90.51	41.79*	75.79
No <sub>names</sub>	80.66*	74.32**	74.2	83.95*	60.93**	73.04*

Table: Character and emotion detection. \*\* ( $p < 0.01$ ), \* ( $p < 0.05$ ).

## Results

Baseline is LaMini-Flan-T5 finetuned on 1823 dream narratives

Model	Status	Gender	Identity	Age	Character	Emotion
Baseline	82.87	78.02	76.17	86.21	64.74	75.13
No <sub>semantics</sub>	71.37	56.54*	61.0	90.51	41.79*	75.79
No <sub>names</sub>	80.66*	74.32**	74.2	83.95*	60.93**	73.04*
Size <sub>small</sub>	78.35**	72.13**	70.25**	81.66**	56.79**	70.15**
Size <sub>large</sub>	84.51*	80.3**	78.63**	87.29	67.63**	74.71

Table: Character and emotion detection. \*\* ( $p < 0.01$ ), \* ( $p < 0.05$ ).

## Results

Baseline is LaMini-Flan-T5 finetuned on 1823 dream narratives

Model	Status	Gender	Identity	Age	Character	Emotion
Baseline	82.87	78.02	76.17	86.21	64.74	75.13
No <sub>semantics</sub>	71.37	56.54*	61.0	90.51	41.79*	75.79
No <sub>names</sub>	80.66*	74.32**	74.2	83.95*	60.93**	73.04*
Size <sub>small</sub>	78.35**	72.13**	70.25**	81.66**	56.79**	70.15**
Size <sub>large</sub>	84.51*	80.3**	78.63**	87.29	67.63**	74.71
First <sub>group</sub>	82.33	77.71	74.86	85.61	63.71	71.94
First <sub>individual</sub>	80.59**	76.14	74.22*	83.87**	62.67	67.32
First <sub>emotion</sub>	83.92	78.74	77.06	87.63	64.97	72.03

Table: Character and emotion detection. \*\* ( $p < 0.01$ ), \* ( $p < 0.05$ ).

# Results

Baseline is LaMini-Flan-T5 finetuned on 1823 dream narratives

Model	Status	Gender	Identity	Age	Character	Emotion
Baseline	82.87	78.02	76.17	86.21	64.74	75.13
No <sub>semantics</sub>	71.37	56.54*	61.0	90.51	41.79*	75.79
No <sub>names</sub>	80.66*	74.32**	74.2	83.95*	60.93**	73.04*
Size <sub>small</sub>	78.35**	72.13**	70.25**	81.66**	56.79**	70.15**
Size <sub>large</sub>	84.51*	80.3**	78.63**	87.29	67.63**	74.71
First <sub>group</sub>	82.33	77.71	74.86	85.61	63.71	71.94
First <sub>individual</sub>	80.59**	76.14	74.22*	83.87**	62.67	67.32
First <sub>emotion</sub>	83.92	78.74	77.06	87.63	64.97	72.03
Conversion <sub>comma</sub>	84.02**	79.84**	77.67**	87.08*	66.69**	73.68
Conversion <sub>marker</sub>	82.39	78.45	76.53	86.09	65.44	74.36

Table: Character and emotion detection. \*\* ( $p < 0.01$ ), \* ( $p < 0.05$ ).

# Results

Baseline is LaMini-Flan-T5 finetuned on 1823 dream narratives

Model	Status	Gender	Identity	Age	Character	Emotion
Baseline	82.87	78.02	76.17	86.21	64.74	75.13
No <sub>semantics</sub>	71.37	56.54*	61.0	90.51	41.79*	75.79
No <sub>names</sub>	80.66*	74.32**	74.2	83.95*	60.93**	73.04*
Size <sub>small</sub>	78.35**	72.13**	70.25**	81.66**	56.79**	70.15**
Size <sub>large</sub>	84.51*	80.3**	78.63**	87.29	67.63**	74.71
First <sub>group</sub>	82.33	77.71	74.86	85.61	63.71	71.94
First <sub>individual</sub>	80.59**	76.14	74.22*	83.87**	62.67	67.32
First <sub>emotion</sub>	83.92	78.74	77.06	87.63	64.97	72.03
Conversion <sub>comma</sub>	84.02**	79.84**	77.67**	87.08*	66.69**	73.68
Conversion <sub>marker</sub>	82.39	78.45	76.53	86.09	65.44	74.36

Table: Character and emotion detection. \*\* ( $p < 0.01$ ), \* ( $p < 0.05$ ).

→ Our models can address this task, while there is still room for improvement

## Case study on the war veteran

Group	Category	% Vet	% Total	Δ
Identity	known*	24.9	51.6	-26.7
	prominent	1.9	2.5	-0.6
	occupational*	22.4	8.0	14.4
	ethnic*	4.1	0.9	3.1
	unknown*	46.8	37.0	9.8

## Case study on the war veteran

Group	Category	% Vet	% Total	Δ
Identity	known*	24.9	51.6	-26.7
	prominent	1.9	2.5	-0.6
	occupational*	22.4	8.0	14.4
	ethnic*	4.1	0.9	3.1
	unknown*	46.8	37.0	9.8
Gender	male*	56.2	43.0	13.1
	female*	24.1	33.1	-9.0
	joint	10.9	12.2	-1.3
	undefined	7.9	8.7	-0.9

**Table:** Identity and gender proportions for the veteran (n=566 narratives) versus other dreamers. Δ shows the difference in percentage points. \* ( $p < 0.05$ ).

## Case study on the war veteran

Group	Category	% Vet	% Total	Δ
Identity	known*	24.9	51.6	-26.7
	prominent	1.9	2.5	-0.6
	occupational*	22.4	8.0	14.4
	ethnic*	4.1	0.9	3.1
	unknown*	46.8	37.0	9.8
Gender	male*	56.2	43.0	13.1
	female*	24.1	33.1	-9.0
	joint	10.9	12.2	-1.3
	undefined	7.9	8.7	-0.9

**Table: Identity and gender proportions for the veteran (n=566 narratives)**  
versus other dreamers. Δ shows the difference in percentage points. \* ( $p < 0.05$ ).

→ The veteran dreams more about *occupational*, *ethnic*, and *unknown* identities compared to other dreamers

## Formalization of style in personal narratives

G. Cortal and A. Finkel. Formalizing Style in Personal Narratives. EMNLP 2025.

## Expand the Cognitive Model of Isabelle Tellier

Tellier was a PhD student of Alain (30 years ago)

- ▶ Tellier's cognitive model proposes formally capturing redundancies in decision-making texts, characterizing a *cognitive style*; using algebraic languages and finite automata.
- ▶ For implementation, Tellier performs a partial automatic analysis using categorial grammars.

We propose:

- ▶ to fully implement the cognitive model using language models
- ▶ to linguistically justify the relevant traits to capture and organize
- ▶ to extend the model to account for the attention mechanisms involved in expressing subjective experience

# How is subjective experience communicated in narratives?

## How is subjective experience communicated in narratives?

We use narratives to express our representations of reality and make sense of the world (Bruner, 1990) [add more recent refs]

## How is subjective experience communicated in narratives?

We use narratives to express our representations of reality and make sense of the world (Bruner, 1990) [add more recent refs]

[better transition]

In everyday usage, style refers to a distinctive manner of expression

## How is subjective experience communicated in narratives?

We use narratives to express our representations of reality and make sense of the world (Bruner, 1990) [add more recent refs]

[better transition]

In everyday usage, style refers to a distinctive manner of expression

Style has been studied in stylometry and stylistics [add refs, maybe cognitive linguistics]

## How is subjective experience communicated in narratives?

We use narratives to express our representations of reality and make sense of the world (Bruner, 1990) [add more recent refs]

[better transition]

In everyday usage, style refers to a distinctive manner of expression

Style has been studied in stylometry and stylistics [add refs, maybe cognitive linguistics]

We use style as a proxy to study how subjective experience is linguistically communicated

## How is subjective experience communicated in narratives?

We use narratives to express our representations of reality and make sense of the world (Bruner, 1990) [add more recent refs]

[better transition]

In everyday usage, style refers to a distinctive manner of expression

Style has been studied in stylometry and stylistics [add refs, maybe cognitive linguistics]

We use style as a proxy to study how subjective experience is linguistically communicated

[for second slide, give a formal definition]

We narrow the general definition of style: *a distinctive manner of communicating subjective experience in narratives*

## How to give an operational definition of style?

## How to give an operational definition of style?

**Hypothesis:** An individual uses some redundant choices of features that characterize its style

# How to give an operational definition of style?

**Hypothesis:** An individual uses some redundant choices of features that characterize its style

## **Contributions:**

1. A sequence-based framework defining style as patterns in sequences of linguistic choices

# How to give an operational definition of style?

**Hypothesis:** An individual uses some redundant choices of features that characterize its style

## **Contributions:**

1. A sequence-based framework defining style as patterns in sequences of linguistic choices
2. A methodology for identifying patterns using sequence analysis

# How to give an operational definition of style?

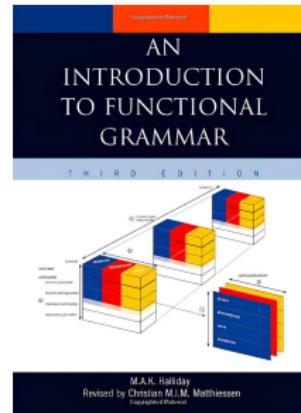
**Hypothesis:** An individual uses some redundant choices of features that characterize its style

## **Contributions:**

1. A sequence-based framework defining style as patterns in sequences of linguistic choices
2. A methodology for identifying patterns using sequence analysis
3. A case study on dream narratives

# What linguistic features encode subjective experience?

We ground our framework in *systemic functional linguistics* (Halliday et al., 2014)



**Figure:** Halliday et al. (2014).  
+57,000 citations.

# What linguistic features encode subjective experience?

We ground our framework in *systemic functional linguistics* (Halliday et al., 2014)

Meaning emerges through choices in systems of linguistic features to achieve communicative goals

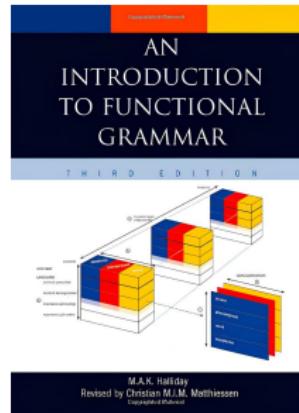


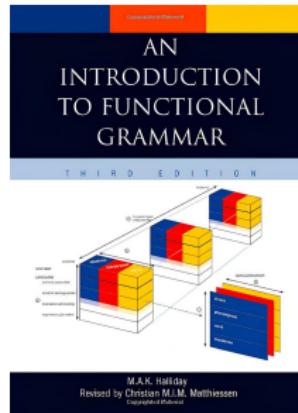
Figure: Halliday et al. (2014).  
+57,000 citations.

# What linguistic features encode subjective experience?

We ground our framework in *systemic functional linguistics* (Halliday et al., 2014)

Meaning emerges through choices in systems of linguistic features to achieve communicative goals

Language achieves three functions: interpersonal (language builds social relationships), textual (information is organized to create coherent messages), and *ideational* (language represents experience)



**Figure:** Halliday et al. (2014).  
+57,000 citations.

# What linguistic features encode subjective experience?

Processes	Examples
Action: actions and events in the physical world.	He <sub>Actor</sub> takes <sub>Action</sub> the valuable <sub>Affected</sub> I <sub>Actor</sub> give <sub>Action</sub> her <sub>Recipient</sub> a chance <sub>Range</sub>
Mental: internal experiences such as thoughts, perceptions, and feelings.	The moon <sub>Senser</sub> sees <sub>Mental</sub> the earth <sub>Phenomenon</sub> He <sub>Senser</sub> disliked <sub>Mental</sub> Gilbert's writing <sub>Phenomenon</sub>
Verbal: acts of communication.	David <sub>Sayer</sub> said <sub>Verbal</sub> "the corrupt, [...]" <sub>Verbiage</sub>
State: states of being, having, or existence.	John <sub>Carrier</sub> is <sub>State</sub> an interesting teacher <sub>Attribute</sub> Chloé <sub>Possessor</sub> has <sub>State</sub> a cat <sub>Possessed</sub>

**Table:** According to *systemic functional linguistics*, language represents experience through **processes** and **participants**.

# Pipeline for our sequence-based framework

[remove bullet points, make it more visual]

Clause	Process (symbol)	Participants
I wake in a dark room	Action (a)	Actor
I feel a cold wind	Mental (m)	Senser, Phenomenon
I tell myself to move	Verbal (v)	Sayer, Recipient

**Sequence:** *amv* | **Substrings:** {*am*, *mv*}

## Results on the war veteran

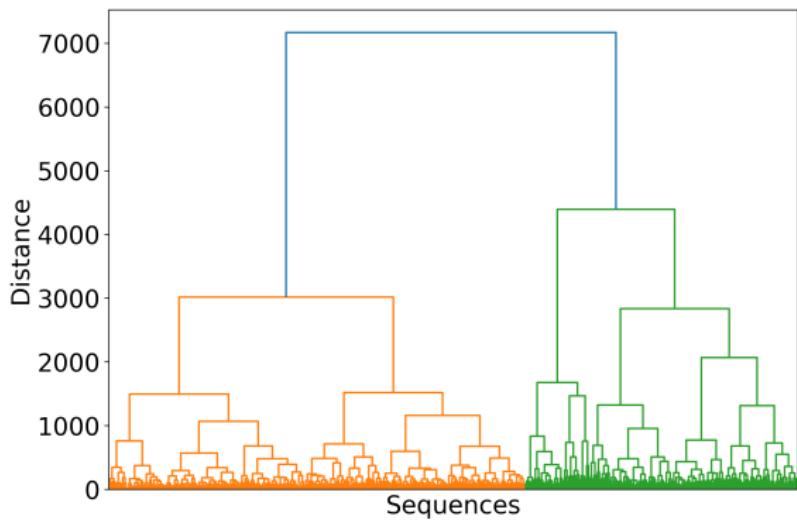


Figure: Dendrogram with Ward linkage and cosine similarity

## Results on the war veteran

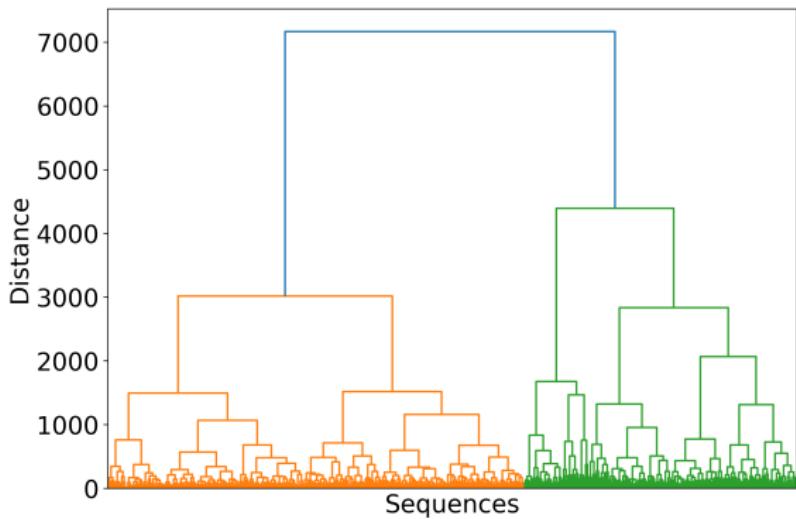


Figure: Dendrogram with Ward linkage and cosine similarity

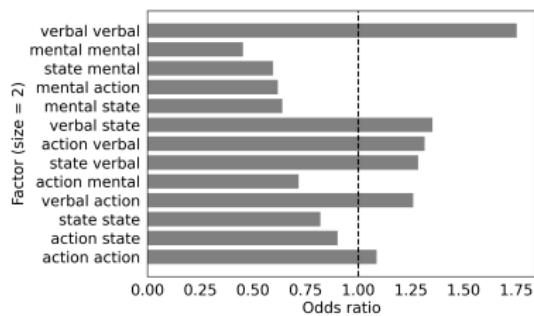
**Representative sequences:** *savamasasaaaamaaaasavvvaaaaaaavssaaaaa* and  
*ssssavaavssvsavvvvsmasasaasasaamaamvmsss* with  
 $a = \text{action}$ ,  $m = \text{mental}$ ,  $s = \text{state}$ ,  $v = \text{verbal}$

## Results on the war veteran

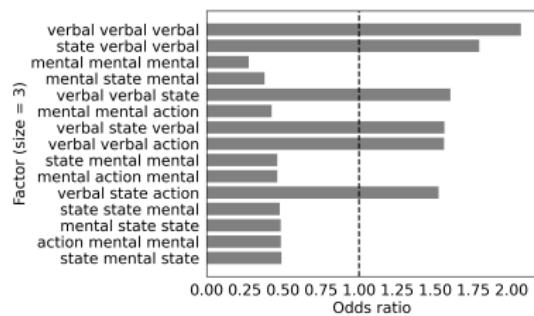
We compare the proportion of sequences containing a given substring

# Results on the war veteran

We compare the proportion of sequences containing a given substring



(a) Size 2.

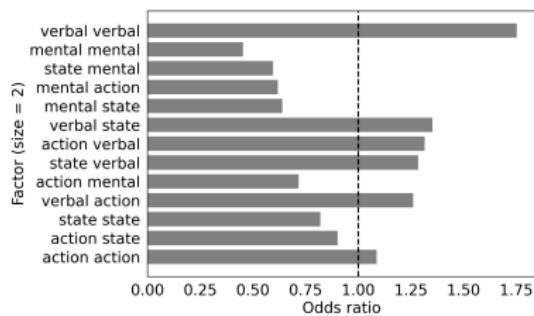


(b) Size 3.

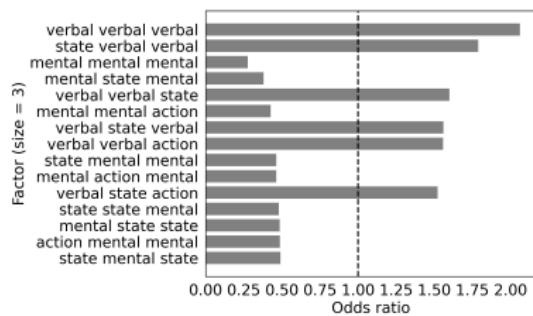
Figure: Top substring odds ratio between the veteran and the norm

# Results on the war veteran

We compare the proportion of sequences containing a given substring



(a) Size 2.



(b) Size 3.

Figure: Top substring odds ratio between the veteran and the norm

→ The veteran favors verbal processes over mental ones

## How can this framework be extended?

- ▶ Authorship profiling

## How can this framework be extended?

- ▶ Authorship profiling
- ▶ Style-conditioned narrative generation

## How can this framework be extended?

- ▶ Authorship profiling
- ▶ Style-conditioned narrative generation
- ▶ Applying methods from complexity science and formal language theory

## Conclusion and perspectives

# Conclusion

How to model subjective experience in personal narratives?

# Conclusion

How to model subjective experience in personal narratives?

- ▶ Definition of objectives and scope using cognitive science

# Conclusion

How to model subjective experience in personal narratives?

- ▶ Definition of objectives and scope using cognitive science
- ▶ Construction of an emotion dataset

# Conclusion

How to model subjective experience in personal narratives?

- ▶ Definition of objectives and scope using cognitive science
- ▶ Construction of an emotion dataset
- ▶ Training of language models for emotion analysis

# Conclusion

How to model subjective experience in personal narratives?

- ▶ Definition of objectives and scope using cognitive science
- ▶ Construction of an emotion dataset
- ▶ Training of language models for emotion analysis
- ▶ Formalization of style in personal narratives

# Conclusion

How to model subjective experience in personal narratives?

- ▶ Definition of objectives and scope using cognitive science
- ▶ Construction of an emotion dataset
- ▶ Training of language models for emotion analysis
- ▶ Formalization of style in personal narratives

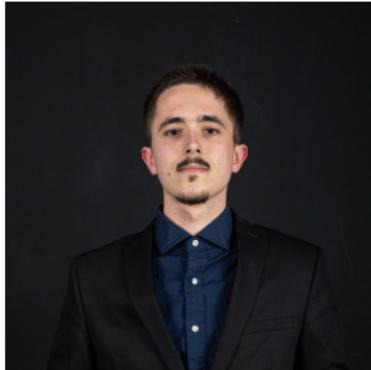
*My research models are publicly hosted on Hugging Face and were trained using the Jean Zay supercomputer*

# Impact

## Impact: Ongoing PhD thesis related to my works



(a) A. Haddou on cognitive distortions  
(2025, ENS Paris-Saclay).



(b) R. Faure on style analysis  
(2025, ENS Paris-Saclay).



(c) N. Richet on multimodal emotion  
(2024, ETS Montréal).

## Impact: NLP for psychiatry (industry)

6-month PhD internship at Callyope on *NLP for quantifying memory, future thinking, and the self in mental health narratives*



# Automatic thematic analysis in mental health narratives using language models

**G. Cortal**, S. Guessoum, X. Cao, R. Riad. *Fine-grained mental health topic modeling in different cohorts using large language models* (preprint). 2025.

## Introduction

- ▶ Qualitative analysis of speech content is central to clinical practice

# Introduction

- ▶ Qualitative analysis of speech content is central to clinical practice
- ▶ Thematic analysis studies how people construct meaning

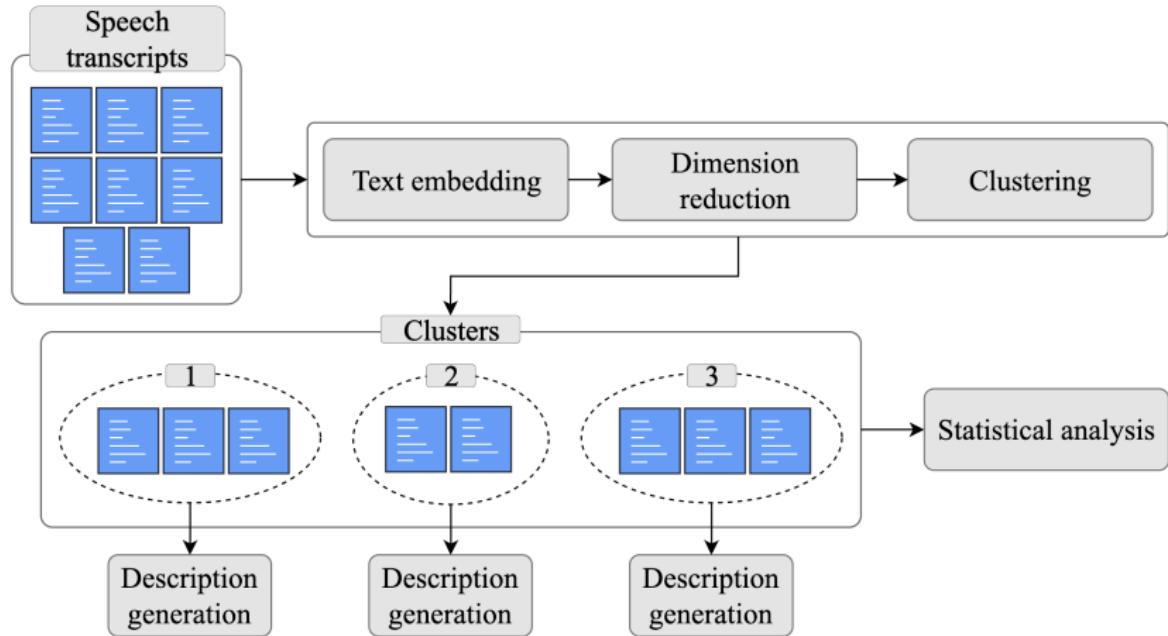
# Introduction

- ▶ Qualitative analysis of speech content is central to clinical practice
- ▶ Thematic analysis studies how people construct meaning
- ▶ Thematic analysis is time-consuming, often constrained to small, monolingual corpora

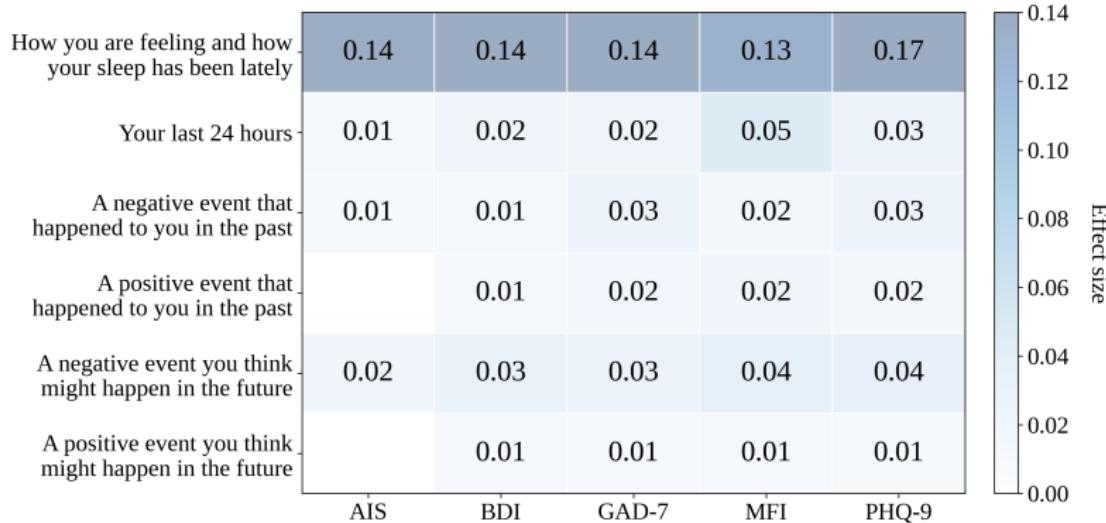
# Introduction

- ▶ Qualitative analysis of speech content is central to clinical practice
- ▶ Thematic analysis studies how people construct meaning
- ▶ Thematic analysis is time-consuming, often constrained to small, monolingual corpora
- ▶ Computational approaches offers time savings, can analyze a larger amount of data

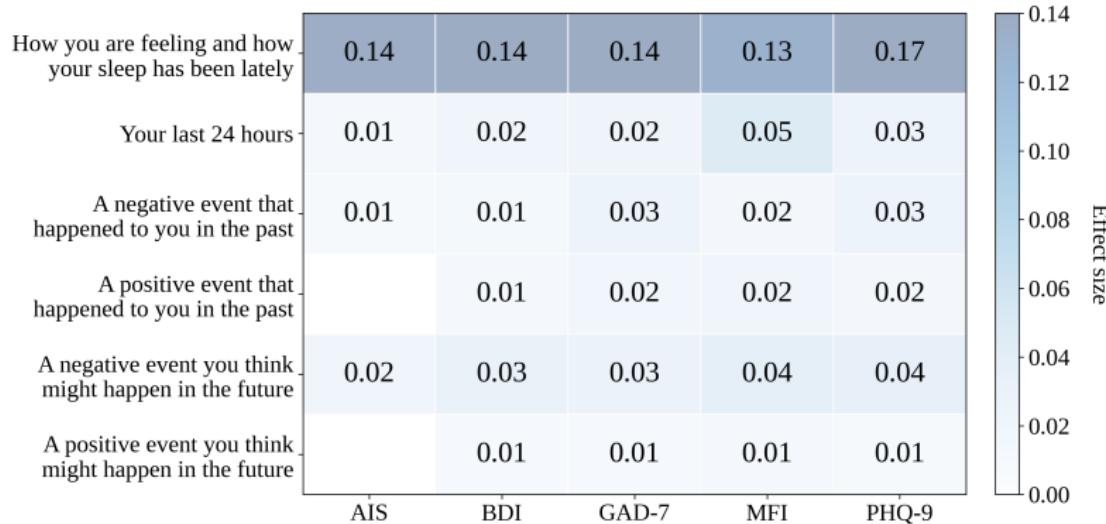
# Semantic clustering and description generation



# Effect size across questions and clinical scores

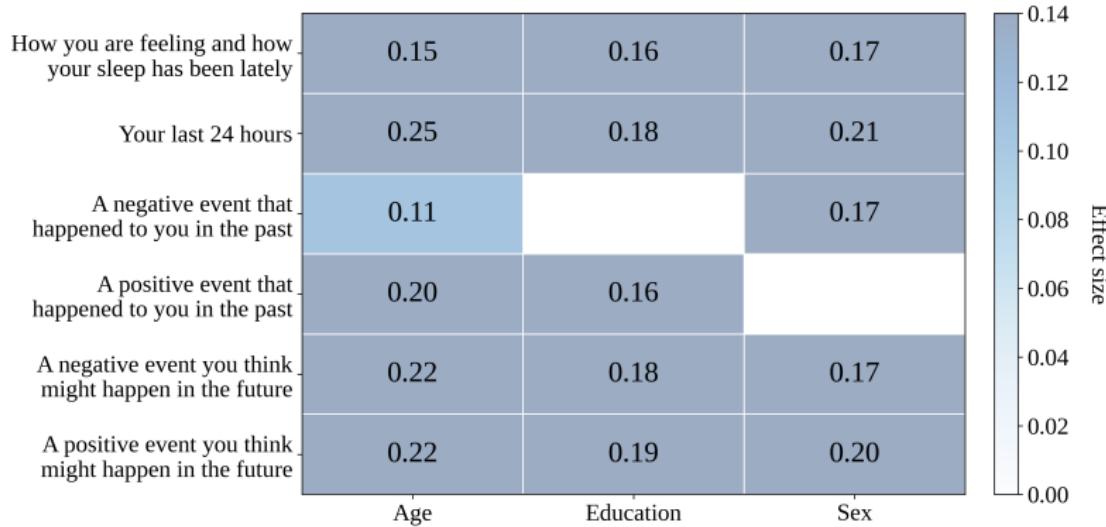


## Effect size across questions and clinical scores

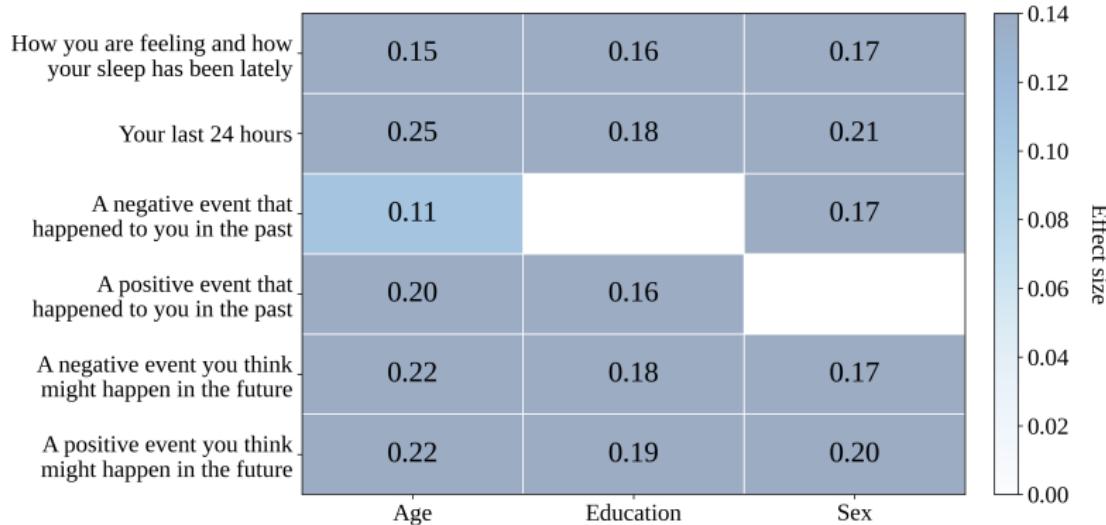


→ Certain questions better discriminate clinical scores

# Effect size across questions and sociodemographics



# Effect size across questions and sociodemographics



→ Nearly all questions discriminate sociodemographics

# Perspectives

## Perspectives

- ▶ Emotion analysis for mental health (empathic support, cognitive distortions, theory of mind)

## Perspectives

- ▶ Emotion analysis for mental health (empathic support, cognitive distortions, theory of mind)
- ▶ Psychology of language models (sycophancy, thought operations)

## Perspectives

- ▶ Emotion analysis for mental health (empathic support, cognitive distortions, theory of mind)
- ▶ Psychology of language models (sycophancy, thought operations)
- ▶ Post-training for psychology (preferences and reasoning data)

# Conclusion

How to model subjective experience in personal narratives?

- ▶ Definition of objectives and scope using cognitive science
- ▶ Construction of an emotion dataset
- ▶ Training of language models for emotion analysis
- ▶ Formalization of style in personal narratives

# Selected research papers

**Constant Bonard and Gustave Cortal (2024).** "Improving Language Models for Emotion Analysis: Insights from Cognitive Science". In: *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*. Ed. by Tatsuki Kurabayashi et al. Bangkok, Thailand: Association for Computational Linguistics, pp. 264–277. DOI: [10.18653/v1/2024.cmcl-1.23](https://doi.org/10.18653/v1/2024.cmcl-1.23)

**Gustave Cortal, Alain Finkel, et al. (2023).** "Emotion Recognition Based on Psychological Components in Guided Narratives for Emotion Regulation". In: *Proceedings of the 7th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Ed. by Stefania Degaetano-Ortlieb et al. Dubrovnik, Croatia: Association for Computational Linguistics, pp. 72–81. DOI: [10.18653/v1/2023.latechclf1-1.8](https://doi.org/10.18653/v1/2023.latechclf1-1.8)

**Gustave Cortal (2024).** "Sequence-to-Sequence Language Models for Character and Emotion Detection in Dream Narratives". In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Ed. by Nicoletta Calzolari et al. Torino, Italia: ELRA and ICCL, pp. 14717–14728

**Gustave Cortal and Alain Finkel (2025).** "Formalizing Style in Personal Narratives". In: *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*. Ed. by Christos Christodoulopoulos et al. Suzhou, China: Association for Computational Linguistics, pp. 7322–7337. ISBN: 979-8-89176-332-6

# Appendix

# Component classification in emotional narratives

Model	Precision	Recall	$F_1$
Logistic Regression	84.9 (0.3)	84.3 (0.3)	84.4 (0.3)
CamemBERT	<b>93.2</b>	<b>93.0</b>	<b>93.1</b>

Table: Scores ( $\pm$  std) for emotion component classification.

## Component classification in emotional narratives

Model	Precision	Recall	$F_1$
Logistic Regression	84.9 (0.3)	84.3 (0.3)	84.4 (0.3)
CamemBERT	<b>93.2</b>	<b>93.0</b>	<b>93.1</b>

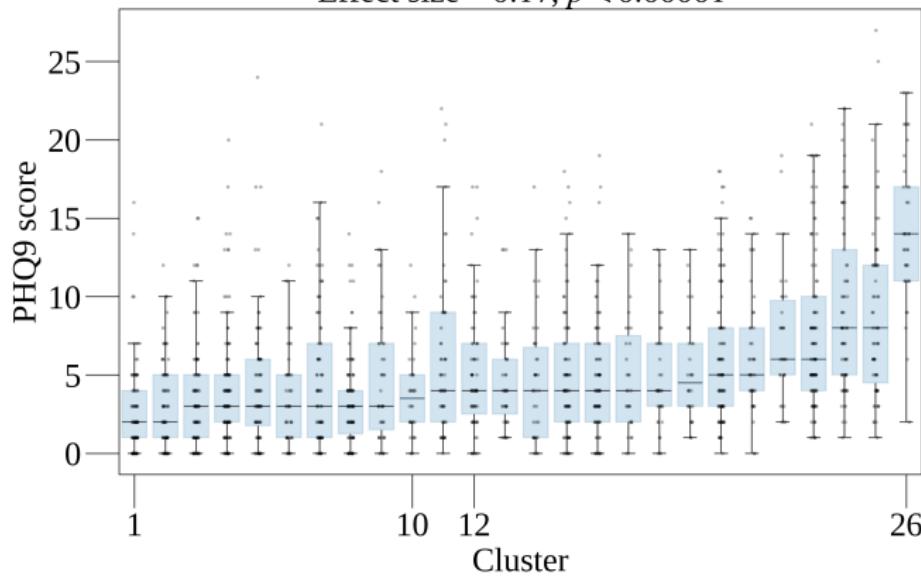
Table: Scores ( $\pm$  std) for emotion component classification.

→ Models can be used to automatically classify unstructured narratives

# Distribution of depression scores across clusters

*How you are feeling and how your sleep has been lately*

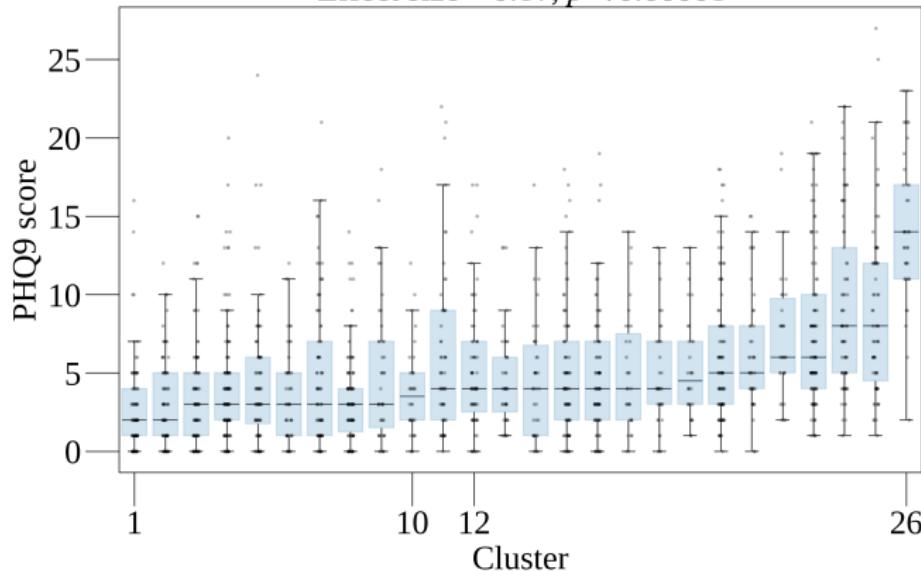
Effect size = 0.17,  $p < 0.00001$



# Distribution of depression scores across clusters

*How you are feeling and how your sleep has been lately*

Effect size = 0.17,  $p < 0.00001$



→ Depression scores vary significantly: cluster 26 highest ( $13.4 \pm 5.4$ ), cluster 1 lowest ( $2.6 \pm 2.2$ )

# Generated cluster descriptions

**Cluster 1 description:** The individuals express consistent satisfaction with their current well-being, emphasizing good sleep quality, restful or pleasant nights, and a general sense of relaxation, even when noting variations in sleep duration or occasional fatigue. (age=39±19, n=92)

**Cluster 10 description:** The individuals express frequent nighttime urinary interruptions disrupting sleep, often attributed to age-related conditions like prostate issues or overactive bladder, alongside mixed reports of physical well-being, mental resilience, and lifestyle factors such as retirement or exercise influencing their overall health and sleep patterns. (age=69±15, n=34)

**Cluster 12 description:** The individuals express stress related to academic exams, significant life decisions, and workloads, alongside sleep disturbances caused by lifestyle changes, increased responsibilities, or environmental adjustments, while some also highlight temporary relief from pressures through personal achievements or upcoming positive events. (age=24±9, n=67)

**Cluster 26 description:** The individuals express sleep disturbances characterized by insomnia, frequent awakenings, and restless sleep, alongside pervasive anxiety, emotional instability, and self-esteem issues, which collectively contribute to persistent fatigue, impaired daily functioning, and a diminished sense of well-being. (age=25±9, n=37)

# Generated cluster descriptions

**Cluster 1 description:** The individuals express consistent satisfaction with their current well-being, emphasizing good sleep quality, restful or pleasant nights, and a general sense of relaxation, even when noting variations in sleep duration or occasional fatigue. (age=39±19, n=92)

**Cluster 10 description:** The individuals express frequent nighttime urinary interruptions disrupting sleep, often attributed to age-related conditions like prostate issues or overactive bladder, alongside mixed reports of physical well-being, mental resilience, and lifestyle factors such as retirement or exercise influencing their overall health and sleep patterns. (age=69±15, n=34)

**Cluster 12 description:** The individuals express stress related to academic exams, significant life decisions, and workloads, alongside sleep disturbances caused by lifestyle changes, increased responsibilities, or environmental adjustments, while some also highlight temporary relief from pressures through personal achievements or upcoming positive events. (age=24±9, n=67)

**Cluster 26 description:** The individuals express sleep disturbances characterized by insomnia, frequent awakenings, and restless sleep, alongside pervasive anxiety, emotional instability, and self-esteem issues, which collectively contribute to persistent fatigue, impaired daily functioning, and a diminished sense of well-being. (age=25±9, n=37)

→ Clustering captures symptom severity and age-related circumstances

# Demographics

	General Population n=1809	Androids n=116	MODMA n=52	VOCES n=90
<b>Demographics</b>				
<b>Language</b>	French	Italian	Chinese	Spanish
<b>Age</b>	***	n.s.	n.s.	***
Mean (SD)	37.8 (18.2)	37.4 (12.0)	31.3 (9.2)	38.6 (14.9)
Range	18–91	19–71	18–52	21–76
<b>Sex, n (%)</b>	n.s.	n.s.	n.s.	n.s.
Female	1187 (66.2)	84 (72.4)	16 (30.8)	39 (43.3)
Male	595 (33.2)	32 (27.6)	36 (69.2)	48 (53.3)
Other	11 (0.6)	0 (0.0)	0 (0.0)	3 (3.3)
<b>Education, n (%)</b>	n.s.	n.s.	n.s.	n.s.
No diploma	52 (2.9)	11 (9.5)	7 (13.5)	-
Secondary	291 (16.2)	37 (31.9)	8 (15.4)	-
Higher short	213 (11.9)	52 (44.8)	0 (0.0)	-
Higher long	1236 (69.0)	16 (13.8)	37 (71.2)	-

## Clinical evaluation

	General Population n=1809	Androids n=116	MODMA n=52	VOCES n=90
<b>C-SSRS</b>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Suicidal risk, n (%)	-	-	-	60 (66.7)
No suicidal risk, n (%)	-	-	-	30 (33.3)
<b>MADRS / MDD</b>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>
Depression, n (%)	-	64 (55.2)	23 (44.2)	-
No depression, n (%)	-	52 (44.8)	29 (55.8)	-
<b>PHQ-9</b>	<i>n.s.</i>	<i>n.s.</i>	***	***
Mean (SD)	5.2 (4.6)	-	9.4 (8.5)	10.5 (6.8)
Range	0–27	-	0–25	0.0–26.0

## References

## References |

-  Bonard, Constant and Gustave Cortal (2024). "Improving Language Models for Emotion Analysis: Insights from Cognitive Science". In: *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*. Ed. by Tatsuki Kurabayashi et al. Bangkok, Thailand: Association for Computational Linguistics, pp. 264–277. DOI: [10.18653/v1/2024.cmcl-1.23](https://doi.org/10.18653/v1/2024.cmcl-1.23).
-  Bruner, Jerome (1990). *Acts of Meaning*. Acts of Meaning. Cambridge, MA, US: Harvard University Press. ISBN: 978-0-674-00360-6.
-  Cortal, Gustave (2024). "Sequence-to-Sequence Language Models for Character and Emotion Detection in Dream Narratives". In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Ed. by Nicoletta Calzolari et al. Torino, Italia: ELRA and ICCL, pp. 14717–14728.

## References II

-  Cortal, Gustave and Alain Finkel (2025). "Formalizing Style in Personal Narratives". In: *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*. Ed. by Christos Christodoulopoulos et al. Suzhou, China: Association for Computational Linguistics, pp. 7322–7337. ISBN: 979-8-89176-332-6.
-  Cortal, Gustave, Alain Finkel, et al. (2023). "Emotion Recognition Based on Psychological Components in Guided Narratives for Emotion Regulation". In: *Proceedings of the 7th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Ed. by Stefania Degaetano-Ortlieb et al. Dubrovnik, Croatia: Association for Computational Linguistics, pp. 72–81. DOI: 10.18653/v1/2023.latechclf1-1.8.
-  Domhoff, G. William and Adam Schneider (2008). "Studying Dream Content Using the Archive and Search Engine on DreamBank.Net". In: *Consciousness and Cognition* 17.4, pp. 1238–1247. ISSN: 10538100. DOI: 10.1016/j.concog.2008.06.010.

## References III

-  Halliday, M.A.K. et al. (2014). *An Introduction to Functional Grammar*. 0th ed. Routledge. ISBN: 978-1-4441-1908-4. DOI: 10.4324/9780203783771.
-  Micheli, Raphaël (2013). "Esquisse d'une typologie des différents modes de sémiotisation verbale de l'émotion". In: *Semen. Revue de sémio-linguistique des textes et discours* 35. ISSN: 0761-2990. DOI: 10.4000/semen.9795.
-  Scherer, Klaus R. (2022). "Theory Convergence in Emotion Science Is Timely and Realistic". In: *Cognition and Emotion* 36.2, pp. 154–170. ISSN: 0269-9931. DOI: 10.1080/02699931.2021.1973378.