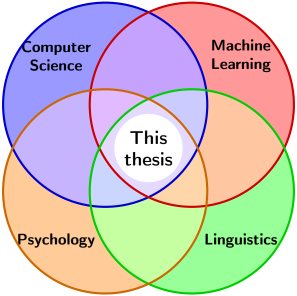
Natural language processing for subjectivity analysis in personal narratives

Gustave Cortal

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

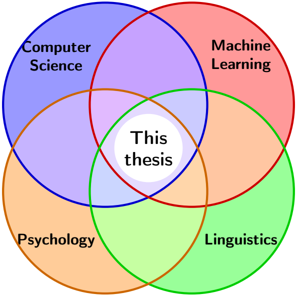
Introduction

Context



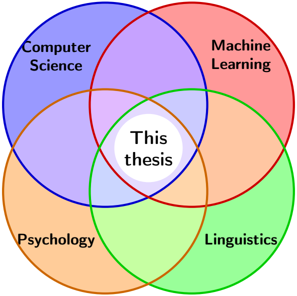
- Natural language processing for psychology is underexplored

Context



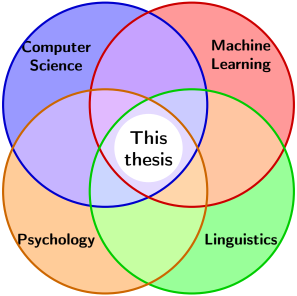
- Natural language processing for psychology is underexplored
- We build on an existing subfield: sentiment and emotion analysis

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

Psychological theories	**In text, emotion is...**	**Example**
Basic emotions theory	a *category*	"I love philosophy." $\rightarrow$ `joy`
Darwin (1872), Tomkins (1962), Ekman (1999), and Plutchik (2001) Demszky et al. (2020) and Bostan and Klinger (2018)		

Psychological theories and emotion annotation schemes

Psychological theories	**In text, emotion is...**	**Example**
Basic emotions theory	a *category*	"I love philosophy." → `joy`
Darwin (1872), Tomkins (1962), Ekman (1999), and Plutchik (2001) Demszky et al. (2020) and Bostan and Klinger (2018)		
Constructivist theories	a continuous value with an *affective* meaning	"His voice soothes me." → `valence` (4/5), `arousal` (1/5)
Schachter and Singer (1962) and Russell and Barrett (1999) Buechel and Hahn (2017)		

Psychological theories and emotion annotation schemes

Psychological theories	**In text, emotion is...**	**Example**
Basic emotions theory	a *category*	"I love philosophy." → `joy`
Darwin (1872), Tomkins (1962), Ekman (1999), and Plutchik (2001) Demszky et al. (2020) and Bostan and Klinger (2018)		
Constructivist theories	a continuous value with an *affective* meaning	"His voice soothes me." → `valence` (4/5), `arousal` (1/5)
Schachter and Singer (1962) and Russell and Barrett (1999) Buechel and Hahn (2017)		
Appraisal theory	a continuous value with a *cognitive* meaning	"I received a surprise gift." → `sudden` (4/5), `control` (0/5)
Arnold (1960) and Lazarus (1991) Hofmann et al. (2020), Troiano, Oberländer, and Klinger (2022), and Zhan, Ong, and J. J. Li (2023)		

Psychological theories and emotion annotation schemes

Psychological theories	**In text, emotion is...**	**Example**
	composed of *semantic roles*	"Louise (`experiencer`) was angry (`cue`) towards Paul (`target`), because he didn't inform her (`cause`)."

Psychological theories and emotion annotation schemes

Psychological theories	**In text, emotion is...**	**Example**
	composed of *semantic roles*	"Louise (`experiencer`) was angry (`cue`) towards Paul (`target`), because he didn't inform her (`cause`)."

Similar to aspect-based sentiment analysis (Zhang, X. Li, et al., 2022): "The battery life is *amazing* (+), but its camera quality is *disappointing* (−)."

Limitations in emotion analysis

- Though the three theories reviewed are usually considered rivals, their integration is possible and desirable (Scherer, 2022a)

Limitations in emotion analysis

- Though the three theories reviewed are usually considered rivals, their integration is possible and desirable (Scherer, 2022a)
- Emotion verbalization is underexplored
(Micheli, 2013b; Etienne, Battistelli, and Lecorvé, 2022)

Limitations in emotion analysis

- Though the three theories reviewed are usually considered rivals, their integration is possible and desirable (Scherer, 2022a)
- Emotion verbalization is underexplored
  (Micheli, 2013b; Etienne, Battistelli, and Lecorvé, 2022)
- Benchmarks evaluate certain aspects of emotional understanding but do not consider its full complexity
  (Campagnano, Conia, and Navigli, 2022; Zhang, Deng, et al., 2023; Paech, 2024)

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, 2013a). 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, 2013a). 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, 2013a). 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, 2013a). 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, 2013a). 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, 2013a). 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

(Davidson, Luo, and Burden, 2001; Nathalie Blanc, 2010; Creissen and N. Blanc, 2017; Foppolo and Ma

Which verbal signs are used to infer expressed emotions?

Raphaël Micheli categorizes a range of linguistic markers into three *emotion expression modes* (Micheli, 2013a). 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
(Davidson, Luo, and Burden, 2001; Nathalie Blanc, 2010; Creissen and N. Blanc, 2017; Foppolo and Ma

→ [add refs] (Etienne, Battistelli, and Lecorvé, 2022; Dragos et al., 2022)

How to integrate psychological theories of emotion?

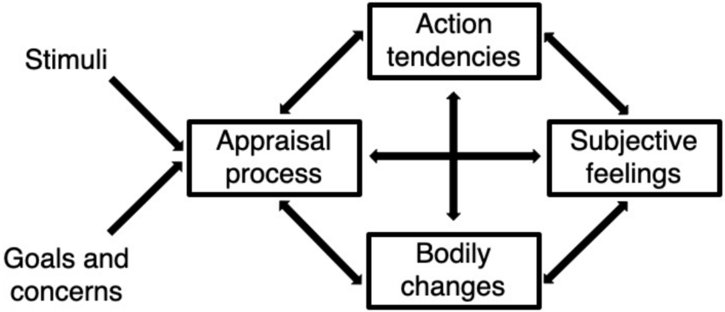


Figure: **The integrated framework for emotion theories** (Scherer, 2022b). Rectangles represents the components constituting an emotional episode, and arrows represent causation.

→ We use this framework to construct an emotion dataset

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

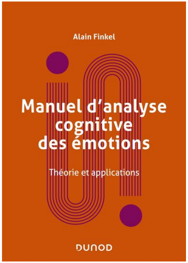
Goal: Create a dataset of narratives structured according to emotion components, following a questionnaire from *Cognitive Analysis of Emotion* (Finkel, 2022)

French emotional narratives based on components

Goal: Create a dataset of narratives structured according to emotion components, following a questionnaire from *Cognitive Analysis of Emotion* (Finkel, 2022)

The questionnaire:

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



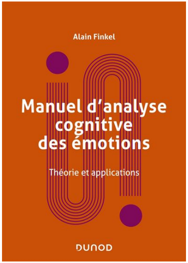
Finkel (2022)

French emotional narratives based on components

Goal: Create a dataset of narratives structured according to emotion components, following a questionnaire from *Cognitive Analysis of Emotion* (Finkel, 2022)

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



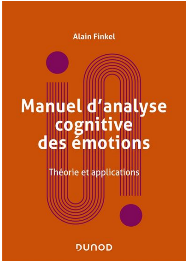
Finkel (2022)

French emotional narratives based on components

Goal: Create a dataset of narratives structured according to emotion components, following a questionnaire from *Cognitive Analysis of Emotion* (Finkel, 2022)

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



Finkel (2022)

French emotional narratives based on components

Limitation: No annotation schemes consider all emotion components

French emotional narratives based on components

Limitation: No annotation schemes consider all emotion components

E. Kim and Klinger (2019) study emotion communication in fiction via subjective sensations, postures, and facial expressions

French emotional narratives based on components

Limitation: No annotation schemes consider all emotion components

E. Kim and Klinger (2019) study emotion communication in fiction via subjective sensations, postures, and facial expressions

Casel, Heindl, and Klinger (2021) associate text spans with Scherer's emotion components

French emotional narratives based on components

Limitation: No annotation schemes consider all emotion components

E. Kim and Klinger (2019) study emotion communication in fiction via subjective sensations, postures, and facial expressions

Casel, Heindl, and Klinger (2021) associate text spans with Scherer's emotion components

$\rightarrow$ We structure emotional narratives according to behaviors, thoughts, physical feelings, and appraisals

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

	Logistic Regression			CamemBERT		
Component	Precision	Recall	F_1	Precision	Recall	F_1
All	71.2 (2.6)	69.1 (2.2)	67.8 (2.3)	**85.1**	**84.8**	**84.7**

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)	**85.1**	**84.8**	**84.7**
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

	Logistic Regression			CamemBERT		
Component	Precision	Recall	F_1	Precision	Recall	F_1
All	71.2 (2.6)	69.1 (2.2)	67.8 (2.3)	**85.1**	**84.8**	**84.7**
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

	Logistic Regression			CamemBERT		
Component	Precision	Recall	F_1	Precision	Recall	F_1
All	71.2 (2.6)	69.1 (2.2)	67.8 (2.3)	**85.1**	**84.8**	**84.7**
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 Precision	Logistic Regression Recall	Logistic Regression F_1	CamemBERT Precision	CamemBERT Recall	CamemBERT F_1
All	71.2 (2.6)	69.1 (2.2)	67.8 (2.3)	**85.1**	**84.8**	**84.7**
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

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

Quantitative analysis of dream narratives

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

Quantitative dream analysis relies on dream databases and annotation schemes (Domhoff and Schneider, 2008; Winget and Kramer, 1979)

Quantitative analysis of dream narratives

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

Quantitative dream analysis relies on dream databases and annotation schemes (Domhoff and Schneider, 2008; Winget and Kramer, 1979)

The annotation process is complex and costly

Quantitative analysis of dream narratives

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

Quantitative dream analysis relies on dream databases and annotation schemes (Domhoff and Schneider, 2008; Winget and Kramer, 1979)

The annotation process is complex and costly

How to automate the annotation process using language models?

Hall and Van de Castle annotation scheme

DreamBank contains 27,000 narratives, only 1823 annotated with the Hall and Van de Castle (HVdC) scheme

(Domhoff and Schneider, 2008; Flanagan, 1966)

Series: **Girls (tutorial)** Number: **0039**

<table>
<tr><th>CHAR.</th><th colspan="2">AGGRESSION</th><th colspan="2">FRIENDLINESS</th><th>SEXUALITY</th><th>SET.</th><th>OBJ.</th></tr>
<tr><td rowspan="5">2MUT
1MUT
1FKT</td><td colspan="2">1MUT 3> 1FKT
D 2= 1MUT</td><td colspan="2">D 5= 1MUT</td><td></td><td>OU</td><td rowspan="5">[not coded]</td></tr>
<tr><td colspan="5">ACTIVITIES</td><td>MOD.</td></tr>
<tr><td colspan="5">[not coded]</td><td rowspan="3">[not coded]</td></tr>
<tr><td>FAILURE</td><td>SUCCESS</td><td>MISFORTUNE</td><td>GOOD FORT.</td><td>EMOTIONS</td></tr>
<tr><td></td><td></td><td></td><td></td><td>AP, D</td></tr>
</table>

#0039
It was my birthday and I was having a party but in a place I've never been before. It was in a forest type area. All I remember is that at the same time I had two boyfriends. Only one was at my party, though he had just broken up with my best friend so I kinda felt uncomfortable being with him. We had got in an argument so he left. I don't quite remember how but we did make up but I don't remember when or why even got in the argument. I woke up when I heard the telephone ringing. (103 words)

Figure: **Example of an annotated dream narrative with HVdC.**

Existing research on computational dream analysis

Lexical-based approaches associate text spans with specific categories (*e.g.*, type of interactions) (Miller, 1994; Fogli, Aiello, and Quercia, 2020)

Existing research on computational dream analysis

Lexical-based approaches associate text spans with specific categories (*e.g.*, type of interactions) (Miller, 1994; Fogli, Aiello, and Quercia, 2020)

Distributional semantic-based approaches represent text spans in a vector space to identify prototypical situations
(Gutman Music, Holur, and Bulkeley, 2022)

Existing research on computational dream analysis

Lexical-based approaches associate text spans with specific categories (*e.g.*, type of interactions) (Miller, 1994; Fogli, Aiello, and Quercia, 2020)

Distributional semantic-based approaches represent text spans in a vector space to identify prototypical situations
(Gutman Music, Holur, and Bulkeley, 2022)

McNamara et al. (2019) and Yu (2022) combine the lexical-based and distributional semantic-based approaches with machine learning

Existing research on computational dream analysis

Bertolini et al. (2023) are the first to fine-tune a pre-trained language model for detecting the presence of emotions

Existing research on computational dream analysis

Bertolini et al. (2023) are the first to fine-tune a pre-trained language model for detecting the presence of emotions

This approach uses full narrative context and compares predictions with HVdC gold annotations

Existing research on computational dream analysis

Bertolini et al. (2023) are the first to fine-tune a pre-trained language model for detecting the presence of emotions

This approach uses full narrative context and compares predictions with HVdC gold annotations

Limitations: emotions are not linked to characters; emotion frequency is not modeled

Existing research on computational dream analysis

Bertolini et al. (2023) are the first to fine-tune a pre-trained language model for detecting the presence of emotions

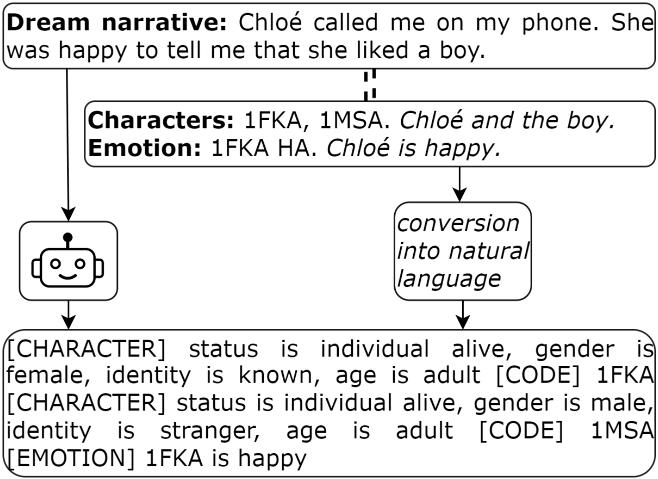
This approach uses full narrative context and compares predictions with HVdC gold annotations

Limitations: emotions are not linked to characters; emotion frequency is not modeled

→ We address this by identifying characters and their emotions with transformer-based sequence-to-sequence models

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
$\texttt{No}_{\texttt{semantics}}$	71.37	56.54*	61.0	90.51	41.79*	75.79
$\texttt{No}_{\texttt{names}}$	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
$\texttt{No}_{\texttt{semantics}}$	71.37	56.54*	61.0	90.51	41.79*	75.79
$\texttt{No}_{\texttt{names}}$	80.66*	74.32**	74.2	83.95*	60.93**	73.04*
$\texttt{Size}_{\texttt{small}}$	78.35**	72.13**	70.25**	81.66**	56.79**	70.15**
$\texttt{Size}_{\texttt{large}}$	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
$\texttt{No}_{\texttt{semantics}}$	71.37	56.54*	61.0	90.51	41.79*	75.79
$\texttt{No}_{\texttt{names}}$	80.66*	74.32**	74.2	83.95*	60.93**	73.04*
$\texttt{Size}_{\texttt{small}}$	78.35**	72.13**	70.25**	81.66**	56.79**	70.15**
$\texttt{Size}_{\texttt{large}}$	84.51*	80.3**	78.63**	87.29	67.63**	74.71
$\texttt{First}_{\texttt{group}}$	82.33	77.71	74.86	85.61	63.71	71.94
$\texttt{First}_{\texttt{individual}}$	80.59**	76.14	74.22*	83.87**	62.67	67.32
$\texttt{First}_{\texttt{emotion}}$	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
$\text{No}_{\text{semantics}}$	71.37	56.54*	61.0	90.51	41.79*	75.79
No_{names}	80.66*	74.32**	74.2	83.95*	60.93**	73.04*
$\text{Size}_{\text{small}}$	78.35**	72.13**	70.25**	81.66**	56.79**	70.15**
$\text{Size}_{\text{large}}$	84.51*	80.3**	78.63**	87.29	67.63**	74.71
$\text{First}_{\text{group}}$	82.33	77.71	74.86	85.61	63.71	71.94
$\text{First}_{\text{individual}}$	80.59**	76.14	74.22*	83.87**	62.67	67.32
$\text{First}_{\text{emotion}}$	83.92	78.74	77.06	87.63	64.97	72.03
$\text{Conversion}_{\text{comma}}$	84.02**	79.84**	77.67**	87.08*	66.69**	73.68
$\text{Conversion}_{\text{marker}}$	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
$\texttt{No}_{\texttt{semantics}}$	71.37	56.54*	61.0	90.51	41.79*	75.79
$\texttt{No}_{\texttt{names}}$	80.66*	74.32**	74.2	83.95*	60.93**	73.04*
$\texttt{Size}_{\texttt{small}}$	78.35**	72.13**	70.25**	81.66**	56.79**	70.15**
$\texttt{Size}_{\texttt{large}}$	84.51*	80.3**	78.63**	87.29	67.63**	74.71
$\texttt{First}_{\texttt{group}}$	82.33	77.71	74.86	85.61	63.71	71.94
$\texttt{First}_{\texttt{individual}}$	80.59**	76.14	74.22*	83.87**	62.67	67.32
$\texttt{First}_{\texttt{emotion}}$	83.92	78.74	77.06	87.63	64.97	72.03
$\texttt{Conversion}_{\texttt{comma}}$	84.02**	79.84**	77.67**	87.08*	66.69**	73.68
$\texttt{Conversion}_{\texttt{marker}}$	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; there is room for improvement

Case study on the war veteran

Group	Category	% Vet	% Total	Δ
	known*	24.9	51.6	-26.7
	prominent	1.9	2.5	-0.6
Identity	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*.

Introduction

Scholarly work has examined personal modes of reasoning and expression

(Husserl, 2012; Hadamard, 1945; Dilts, 1994; Granger, 1968)

Introduction

Scholarly work has examined personal modes of reasoning and expression
(Husserl, 2012; Hadamard, 1945; Dilts, 1994; Granger, 1968)

$\rightarrow$ They describe "styles of thought" but lack operational tools

Introduction

Scholarly work has examined personal modes of reasoning and expression (Husserl, 2012; Hadamard, 1945; Dilts, 1994; Granger, 1968)

→ They describe "styles of thought" but lack operational tools

Style is central to how authors express themselves: stylistics (Wales, 2014), stylometry (Neal et al., 2017)

Introduction

Scholarly work has examined personal modes of reasoning and expression (Husserl, 2012; Hadamard, 1945; Dilts, 1994; Granger, 1968)

→ They describe "styles of thought" but lack operational tools

Style is central to how authors express themselves: stylistics (Wales, 2014), stylometry (Neal et al., 2017)

→ They provide operational tools to capture linguistic form, but do not focus on how such forms encode subjective experience

Introduction

Limitation: A formalization of style that captures how subjective experience is linguistically communicated is lacking

Introduction

Limitation: A formalization of style that captures how subjective experience is linguistically communicated is lacking

Formalization could enable more precise identification of linguistic patterns associated with psychological states and may support interventions (White and Epston, 1990)

Introduction

Limitation: A formalization of style that captures how subjective experience is linguistically communicated is lacking

Formalization could enable more precise identification of linguistic patterns associated with psychological states and may support interventions (White and Epston, 1990)

$\rightarrow$ Our approach complements existing works by providing a formal framework grounded in systemic functional linguistics

Introduction

Limitation: A formalization of style that captures how subjective experience is linguistically communicated is lacking

Formalization could enable more precise identification of linguistic patterns associated with psychological states and may support interventions (White and Epston, 1990)

→ Our approach complements existing works by providing a formal framework grounded in systemic functional linguistics

→ We aim to create a accessible framework that researchers can build upon in future studies

How to give an operational definition of style?

Intuitive definition: a distinctive manner of communicating subjective experience in narratives

How to give an operational definition of style?

Intuitive definition: a distinctive manner of communicating subjective experience in narratives

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

How to give an operational definition of style?

Intuitive definition: a distinctive manner of communicating subjective experience in narratives

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 that encode subjective experience*

How to give an operational definition of style?

Intuitive definition: a distinctive manner of communicating subjective experience in narratives

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 that encode subjective experience*
2. A methodology for identifying patterns using sequence analysis

How to give an operational definition of style?

Intuitive definition: a distinctive manner of communicating subjective experience in narratives

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 that encode subjective experience*
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)

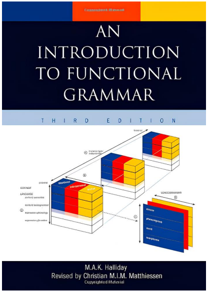


Figure: Halliday et al. (2014). $+57,000$ citations.

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

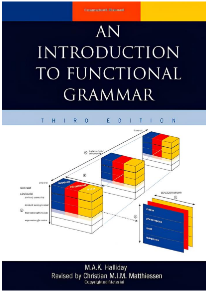


Figure: Halliday et al. (2014). $+57,000$ citations.

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)

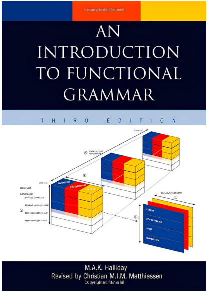


Figure: Halliday et al. (2014). $+57,000$ citations.

What linguistic features encode subjective experience?

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

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

Methodology for our sequence-based framework

Goal: Map narratives to sequences based on extracted linguistic features:

Methodology for our sequence-based framework

Goal: Map narratives to sequences based on extracted linguistic features:

"I wake in a dark room. I feel a cold wind. I tell myself to move." $\rightarrow$ *amv*

Methodology for our sequence-based framework

Goal: Map narratives to sequences based on extracted linguistic features:

"I wake in a dark room. I feel a cold wind. I tell myself to move." $\rightarrow$ *amv*

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}

Substrings are contiguous sequences of symbols within a sequence

Results on the war veteran

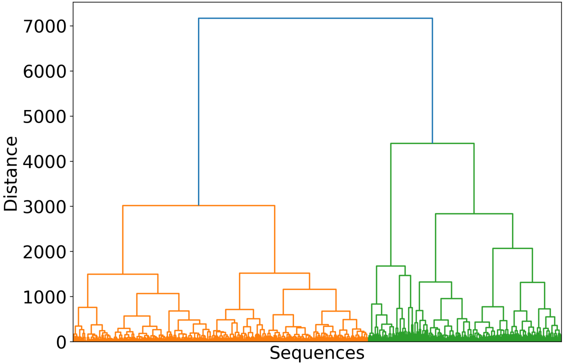


Figure: Dendrogram with Ward linkage and cosine similarity

Results on the war veteran

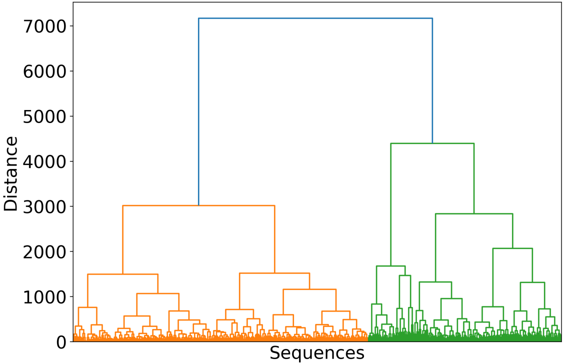


Figure: Dendrogram with Ward linkage and cosine similarity

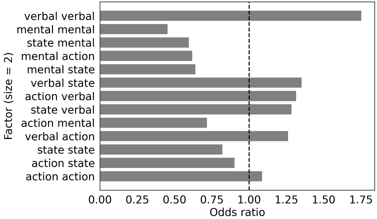
Representative sequences: *savamasasaaamaaasavvvaaaaaaavssaaaaa* and *sssssavaavssvsavvvvsmasasaasasaamaamvmsss* with
a = action, m = mental, s = state, v = verbal

Results on the war veteran

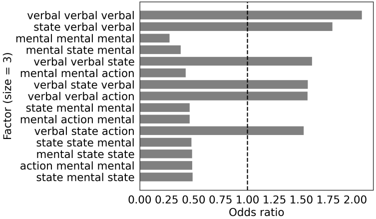
We compare the proportion of sequences containing a given substring

Results on the war veteran

We compare the proportion of sequences containing a given substring



(a) Size 2.

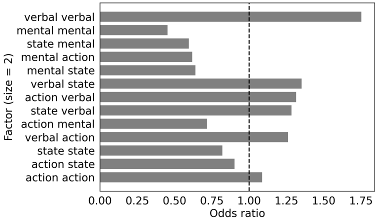


(b) Size 3.

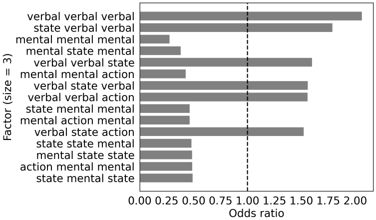
Figure: Top substring odds ratio between the veteran and the norm

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

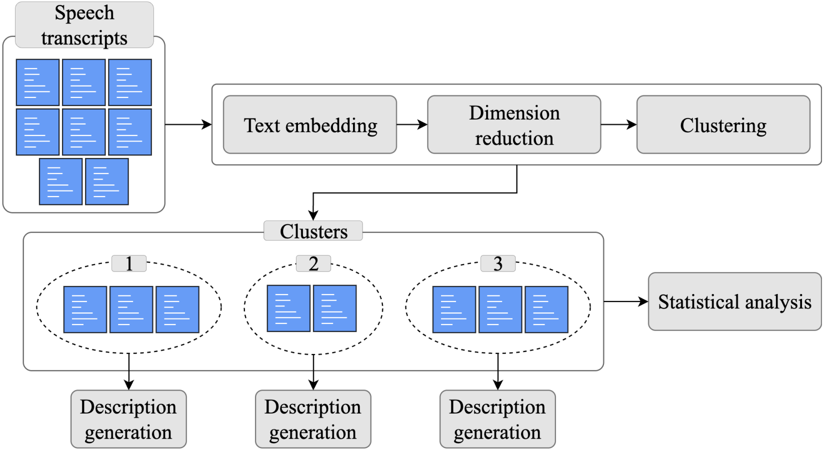
Introduction

- 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

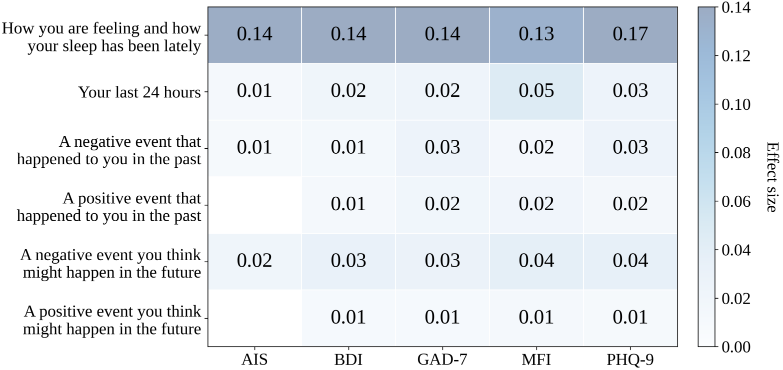
Introduction

- 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

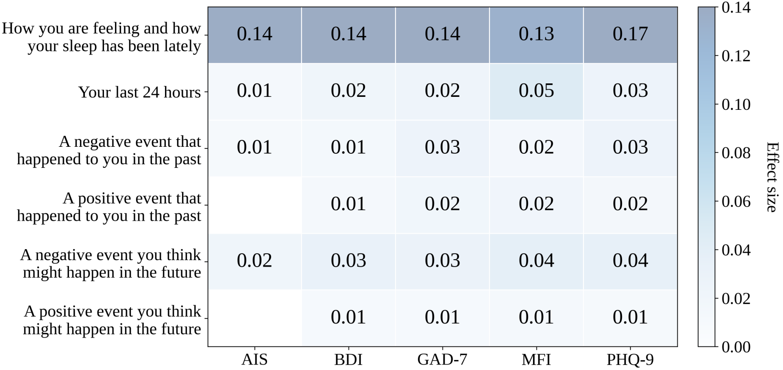
Semantic clustering and description generation



Effect size across questions and clinical scores

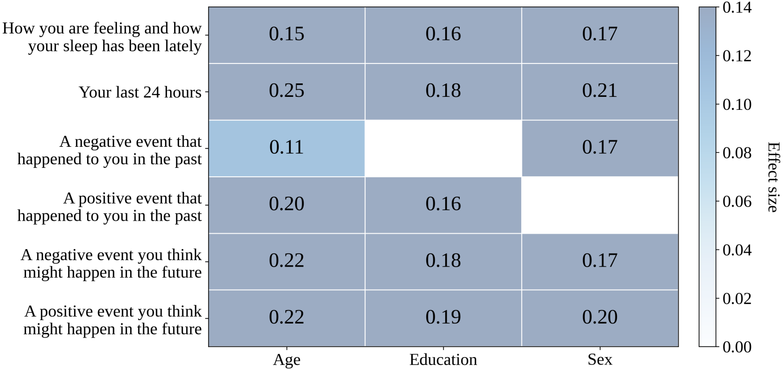


Effect size across questions and clinical scores

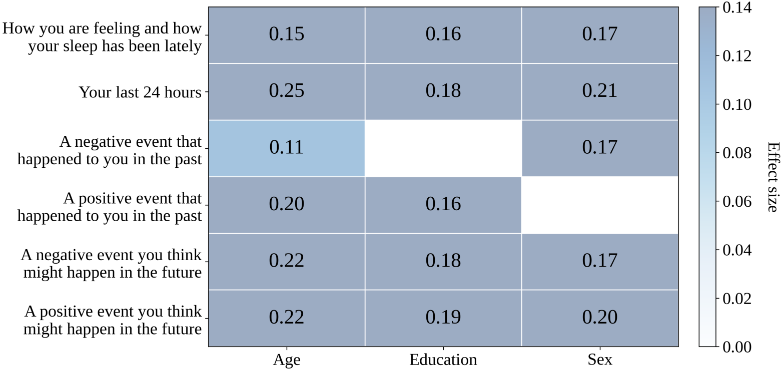


→ Certain questions better discriminate clinical scores

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 (Sharma et al., 2023), cognitive distortions, theory of mind (Zhou et al., 2023; Ma et al., 2023; H. Kim et al., 2023; Gandhi et al., 2023)

Perspectives

- **Emotion analysis for mental health**: empathic support (Sharma et al., 2023), cognitive distortions, theory of mind (Zhou et al., 2023; Ma et al., 2023; H. Kim et al., 2023; Gandhi et al., 2023)
- **Psychology of language models**: sycophancy, thought operations (Didolkar et al., 2025) [add refs]

Perspectives

- **Emotion analysis for mental health**: empathic support (Sharma et al., 2023), cognitive distortions, theory of mind (Zhou et al., 2023; Ma et al., 2023; H. Kim et al., 2023; Gandhi et al., 2023)
- **Psychology of language models**: sycophancy, thought operations (Didolkar et al., 2025) [add refs]
- **Post-training for psychology** (preferences and reasoning data) [add refs]

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 Kuribayashi et al. Bangkok, Thailand: Association for Computational Linguistics, pp. 264–277. DOI: 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.latechclfl-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

What are the psychological mechanisms used to infer what is communicated?

A *code* is a pre-established pairing between stimuli and sets of information

What are the psychological mechanisms used to infer what is communicated?

A *code* is a pre-established pairing between stimuli and sets of information

The Morse code is a pairing between <combination of short and long signals> and [letters]

What are the psychological mechanisms used to infer what is communicated?

A *code* is a pre-established pairing between stimuli and sets of information

The Morse code is a pairing between <combination of short and long signals> and [letters]

The formal semantics of a language is made of syntactical and lexical rules that pairs <strings of words> with [sentential meanings]

What are the psychological mechanisms used to infer what is communicated?

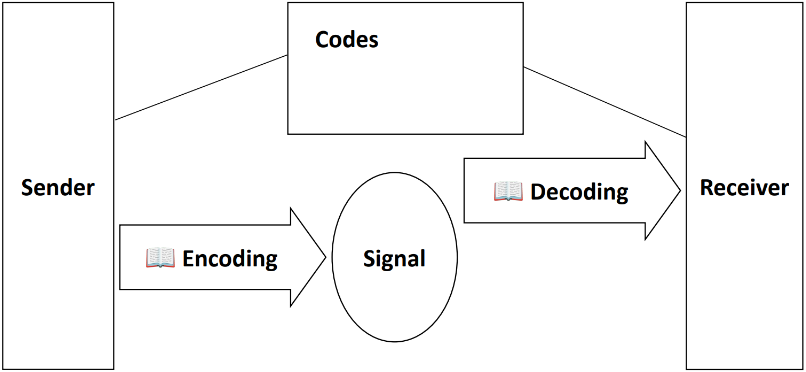


Figure: Dictionary analysis in cognitive pragmatics.

Codes underdetermine emotion meaning

Codes underdetermine emotion meaning

Let's take emotion expression modes as an example:

Codes underdetermine emotion meaning

Let's take emotion expression modes as an example:

- *Labeled*: "I am happy now" is explicit about the feeling but does not encode what the emotion is about

Codes underdetermine emotion meaning

Let's take emotion expression modes as an example:

- *Labeled*: "I am happy now" is explicit about the feeling but does not encode what the emotion is about
- *Displayed*: interjections ("Wow!", "Ah!", "Damn!") show affect yet leave valence and focus unclear

Codes underdetermine emotion meaning

Let's take emotion expression modes as an example:

- *Labeled*: "I am happy now" is explicit about the feeling but does not encode what the emotion is about
- *Displayed*: interjections ("Wow!", "Ah!", "Damn!") show affect yet leave valence and focus unclear
- *Suggested*: "The ship has black sails." can communicate any kind of emotion

Codes underdetermine emotion meaning

Let's take emotion expression modes as an example:

- *Labeled*: "I am happy now" is explicit about the feeling but does not encode what the emotion is about
- *Displayed*: interjections ("Wow!", "Ah!", "Damn!") show affect yet leave valence and focus unclear
- *Suggested*: "The ship has black sails." can communicate any kind of emotion

Codes underdetermine emotion meaning

Let's take emotion expression modes as an example:

- *Labeled*: "I am happy now" is explicit about the feeling but does not encode what the emotion is about
- *Displayed*: interjections ("Wow!", "Ah!", "Damn!") show affect yet leave valence and focus unclear
- *Suggested*: "The ship has black sails." can communicate any kind of emotion

→ We rely on other sources of evidence to infer what is communicated

What are the psychological mechanisms used to infer what is communicated?

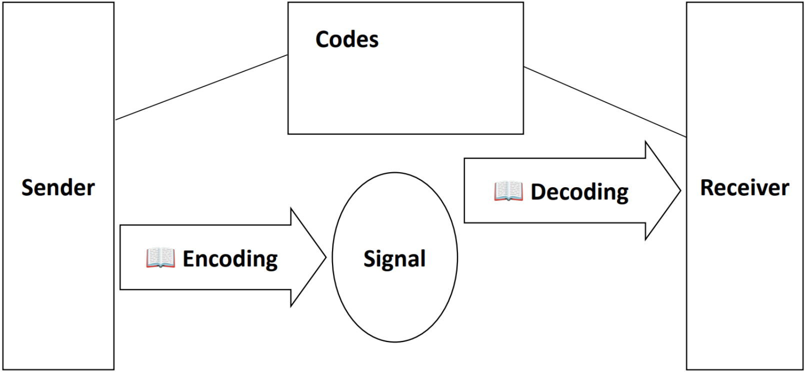


Figure: Dictionary analysis in cognitive pragmatics.

What are the psychological mechanisms used to infer what is communicated?

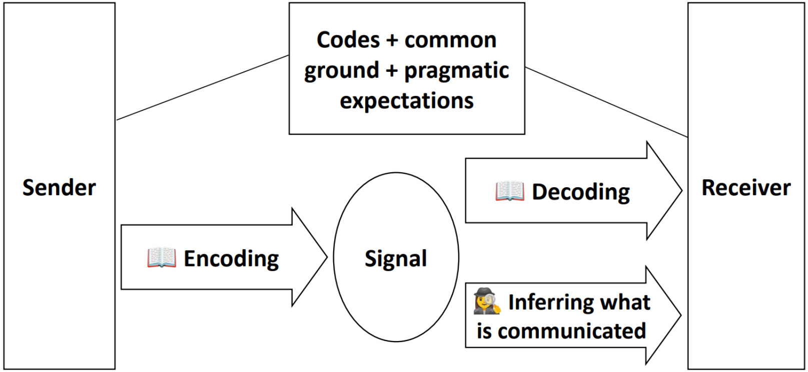


Figure: Detective analysis in cognitive pragmatics.

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	**93.2**	**93.0**	**93.1**

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

Component classification in emotional narratives

Model	**Precision**	**Recall**	$\boldsymbol{F_1}$
Logistic Regression	84.9 (0.3)	84.3 (0.3)	84.4 (0.3)
CamemBERT	**93.2**	**93.0**	**93.1**

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

$\rightarrow$ Models can be used to automatically classify unstructured narratives

Results

$\texttt{StableBeluga}_\texttt{i}$ is a 7B model with in-context learning using *i* examples

Model	Status	Gender	Identity	Age	Character	Emotion
Baseline	82.87	78.02	76.17	86.21	64.74	75.13
$\texttt{StableBeluga}_1$	43.95**	39.76**	31.25**	56.16**	15.65**	-
$\texttt{StableBeluga}_3$	52.44**	46.49**	38.46**	63.88**	21.06**	-
$\texttt{StableBeluga}_5$	55.89**	46.29**	42.61**	63.73**	24.86**	-

Table: F_1-scores for character and emotion detection. Significant differences from baseline: ** ($p < 0.01$), * ($p < 0.05$).

Results

`StableBeluga`$_i$ is a 7B model with in-context learning using *i* examples

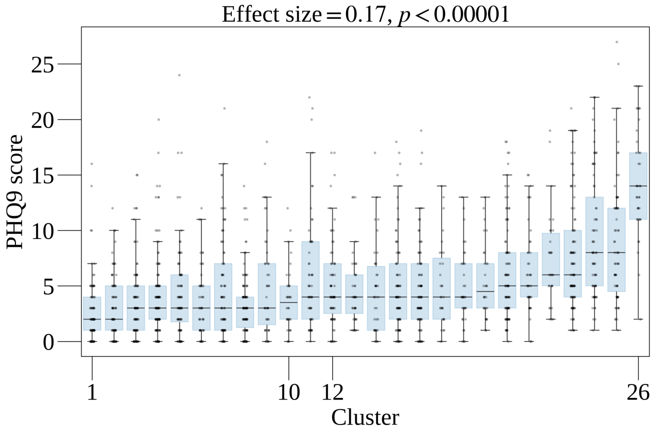
Model	**Status**	**Gender**	**Identity**	**Age**	**Character**	**Emotion**
`Baseline`	82.87	78.02	76.17	86.21	64.74	75.13
`StableBeluga`$_1$	43.95**	39.76**	31.25**	56.16**	15.65**	-
`StableBeluga`$_3$	52.44**	46.49**	38.46**	63.88**	21.06**	-
`StableBeluga`$_5$	55.89**	46.29**	42.61**	63.73**	24.86**	-

Table: F_1-scores for character and emotion detection. Significant differences from `baseline`: ** ($p < 0.01$), * ($p < 0.05$).

→ Compared to `StableBeluga`, our supervised models perform better while having 28 times fewer parameters (248M *vs.* 7B)

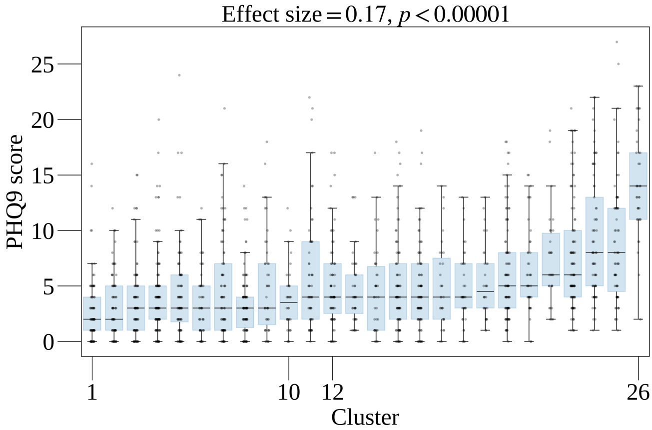
Distribution of depression scores across clusters

How you are feeling and how your sleep has been lately



Distribution of depression scores across clusters

How you are feeling and how your sleep has been lately



→ Depression scores vary significantly: cluster 26 highest (13.4±5.4), cluster 1 lowest (2.6±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
Demographics				
Language	French	Italian	Chinese	Spanish
Age	***	*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
Sex, n (%)	*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)
Education, n (%)	*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
C-SSRS	*n.s.*	*n.s.*	*n.s.*	*n.s.*
Suicidal risk, n (%)	-	-	-	60 (66.7)
No suicidal risk, n (%)	-	-	-	30 (33.3)
MADRS / MDD	*n.s.*	*n.s.*	*n.s.*	*n.s.*
Depression, n (%)	-	64 (55.2)	23 (44.2)	-
No depression, n (%)	-	52 (44.8)	29 (55.8)	-
PHQ-9	*n.s.*	*n.s.*	***	***
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 I

Arnold, Magda B. (1960). *Emotion and Personality*. New York: Columbia University Press.

Bertolini, Lorenzo et al. (2023). *Automatic scoring of dream reports' emotional content with large language models*. arXiv: 2302.14828[cs.CL].

Blanc, Nathalie (2010). "La Compréhension Des Contes Entre 5 et 7 Ans: Quelle Représentation Des Informations Émotionnelles? [The Comprehension of the Tales between 5 and 7 Year-Olds: Which Representation of Emotional Information?]" In: *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale* 64.4, pp. 256–265. ISSN: 1878-7290. DOI: 10.1037/a0021283.

References II

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 Kuribayashi et al. Bangkok, Thailand: Association for Computational Linguistics, pp. 264–277. DOI: 10.18653/v1/2024.cmcl-1.23.

Bostan, Laura-Ana-Maria and Roman Klinger (2018). "An Analysis of Annotated Corpora for Emotion Classification in Text". In: *Proceedings of the 27th International Conference on Computational Linguistics*. Santa Fe, New Mexico, USA: Association for Computational Linguistics, pp. 2104–2119. URL: https://aclanthology.org/C18-1179.

References III

Buechel, Sven and Udo Hahn (2017). "EmoBank: Studying the Impact of Annotation Perspective and Representation Format on Dimensional Emotion Analysis". In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. Valencia, Spain: Association for Computational Linguistics, pp. 578–585. URL: https://aclanthology.org/E17-2092.

Campagnano, Cesare, Simone Conia, and Roberto Navigli (2022). "SRL4E – Semantic Role Labeling for Emotions: A Unified Evaluation Framework". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Dublin, Ireland: Association for Computational Linguistics, pp. 4586–4601. DOI: 10.18653/v1/2022.acl-long.314. URL: https://aclanthology.org/2022.acl-long.314.

References IV

Casel, Felix, Amelie Heindl, and Roman Klinger (2021). "Emotion Recognition under Consideration of the Emotion Component Process Model". In: *Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021)*. Düsseldorf, Germany: KONVENS 2021 Organizers, pp. 49–61. URL: https://aclanthology.org/2021.konvens-1.5.

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.

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.

References V

- 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.latechclfl-1.8`.
- Creissen, S. and N. Blanc (2017). "Quelle Représentation Des Différentes Facettes de La Dimension Émotionnelle d'une Histoire Entre l'âge de 6 et 10 Ans ? Apports d'une Étude Multimédia". In: *Psychologie Française*. Cognition et Multimédia : Les Atouts Du Numérique En Situation d'apprentissage 62.3, pp. 263–277. ISSN: 0033-2984. DOI: `10.1016/j.psfr.2015.07.006`.
- Darwin, Charles (1872). *The expression of the emotions in man and animals*. London: John Murray.

References VI

Davidson, Denise, Zupei Luo, and Matthew J. Burden (2001). "Children's Recall of Emotional Behaviours, Emotional Labels, and Nonemotional Behaviours: Does Emotion Enhance Memory?" In: *Cognition and Emotion* 15.1, pp. 1–26. ISSN: 1464-0600. DOI: 10.1080/02699930042000105.

Demszky, Dorottya et al. (2020). "GoEmotions: a dataset of fine-grained emotions". In: *Proceedings of the 58th annual meeting of the association for computational linguistics*. Online: Association for Computational Linguistics, pp. 4040–4054. DOI: 10.18653/v1/2020.acl-main.372. URL: https://aclanthology.org/2020.acl-main.372.

Didolkar, Aniket et al. (2025). *Metacognitive Reuse: Turning Recurring LLM Reasoning Into Concise Behaviors*. arXiv: 2509.13237 [cs.LG]. URL: https://arxiv.org/abs/2509.13237.

References VII

Dilts, Robert (1994). *Strategies of Genius, Volume 1: Aristotle, Sherlock Holmes, Walt Disney, Wolfgang Amadeus Mozart*. Capitola, CA: Meta Publications. ISBN: 0-916990-32-X. URL: https://books.google.com/books/about/Strategies_of_Genius_Aristotle_Sherlock.html?id=jAQRAQAAIAAJ.

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. Publisher: Elsevier BV, pp. 1238–1247. DOI: 10.1016/j.concog.2008.06.010. URL: https://doi.org/10.1016/j.concog.2008.06.010.

Dragos, Valentina et al. (2022). "Angry or sad ? Emotion annotation for extremist content characterisation". In: *Proceedings of the thirteenth language resources and evaluation conference*. Marseille, France: European Language Resources Association, pp. 193–201. URL: https://aclanthology.org/2022.lrec-1.21.

References VIII

Ekman, Paul (1999). "Basic emotions". In: *Handbook of cognition and emotion*. Ed. by Tim Dalgleish and Mike J. Power. Chichester: John Wiley & Sons Ltd, pp. 45–60. (Visited on Feb. 12, 2024).

Etienne, Aline, Delphine Battistelli, and Gwénolé Lecorvé (2022). "A (psycho-)linguistically motivated scheme for annotating and exploring emotions in a genre-diverse corpus". In: *Proceedings of the thirteenth language resources and evaluation conference*. Marseille, France: European Language Resources Association, pp. 603–612. URL: https://aclanthology.org/2022.lrec-1.64.

Finkel, Alain (2022). *Manuel d'analyse cognitive des émotions: Théorie et applications*. Paris: Dunod.

Flanagan, H. M. (1966). "The content analysis of dreams. By calvin S. Hall and robert L. Van de castle new york: The century psychology series. 1966. Pp. 320. Price not given.". In: *The British Journal of Psychiatry* 112.490. Publisher: Cambridge University Press, pp. 963–964. DOI: 10.1192/bjp.112.490.963.

References IX

Fogli, Alessandro, Luca Maria Aiello, and Daniele Quercia (2020). "Our dreams, our selves: automatic analysis of dream reports". In: *Royal Society Open Science* 7.8. Publisher: The Royal Society, p. 192080. DOI: 10.1098/rsos.192080. URL: https://doi.org/10.1098/rsos.192080.

Foppolo, Francesca and Greta Mazzaggio (2024). "Conversational Implicature and Communication Disorders". en. In: *The Handbook of Clinical Linguistics, Second Edition*. Ed. by Martin J. Ball, Nicole Müller, and Elizabeth Spencer. 1st ed. Wiley, pp. 15–27. ISBN: 978-1-119-87590-1 978-1-119-87594-9. DOI: 10.1002/9781119875949.ch2. URL: https://onlinelibrary.wiley.com/doi/10.1002/9781119875949.ch2 (visited on Feb. 13, 2024).

Gandhi, Kanishk et al. (2023). *Understanding Social Reasoning in Language Models with Language Models*. arXiv. DOI: 10.48550/arXiv.2306.15448. arXiv: 2306.15448 [cs].

References X

Granger, Gilles-Gaston (1968). *Essai d'une Philosophie Du Style*. Paris: A. Colin.

Gutman Music, Maja, Pavan Holur, and Kelly Bulkeley (2022). "Mapping dreams in a computational space: A phrase-level model for analyzing Fight/Flight and other typical situations in dream reports". In: *Consciousness and Cognition* 106, p. 103428. ISSN: 1053-8100. DOI: https://doi.org/10.1016/j.concog.2022.103428. URL: https://www.sciencedirect.com/science/article/pii/S105381002200160X.

Hadamard, Jacques (1945). *An Essay on the Psychology of Invention in the Mathematical Field*. Princeton, N.J.: Princeton University Press. URL: https://books.google.com/books?id=W2U4AAAAMAAJ.

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.

References XI

Hofmann, Jan et al. (2020). "Appraisal theories for emotion classification in text". In: *Proceedings of the 28th international conference on computational linguistics*. Barcelona, Spain (Online): International Committee on Computational Linguistics, pp. 125–138. DOI: 10.18653/v1/2020.coling-main.11. URL: https://aclanthology.org/2020.coling-main.11.

Husserl, Edmund (2012). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: First Book: General Introduction to a Pure Phenomenology*. illustrated. Dordrecht: Springer Science & Business Media. ISBN: 978-94-009-7445-6. URL: https://books.google.com/books/about/Ideas_Pertaining_to_a_Pure_Phenomenology.html?id=geEsBAAAQBAJ.

References XII

Kim, Evgeny and Roman Klinger (2019). "An Analysis of Emotion Communication Channels in Fan-Fiction: Towards Emotional Storytelling". In: *Proceedings of the Second Workshop on Storytelling*. Florence, Italy: Association for Computational Linguistics, pp. 56–64. DOI: 10.18653/v1/W19-3406. URL: https://aclanthology.org/W19-3406.

Kim, Hyunwoo et al. (2023). *FANToM: A Benchmark for Stress-testing Machine Theory of Mind in Interactions*. arXiv. DOI: 10.48550/arXiv.2310.15421. arXiv: 2310.15421 [cs].

Lazarus, Richard S. (1991). "Progress on a cognitive-motivational-relational theory of emotion.". In: *American psychologist* 46.8, p. 819.

Ma, Ziqiao et al. (2023). *Towards A Holistic Landscape of Situated Theory of Mind in Large Language Models*. arXiv. DOI: 10.48550/arXiv.2310.19619. arXiv: 2310.19619 [cs].

References XIII

McNamara, Patrick et al. (2019). "Dream content analysis using Artificial Intelligence". In: *International Journal of Dream Research* 12.1, pp. 42–52. DOI: 10.11588/ijodr.2019.1.48744. URL: https://journals.ub.uni-heidelberg.de/index.php/IJoDR/article/view/48744.

Micheli, Raphaël (2013a). "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.

— (2013b). "Esquisse dune typologie des différents modes de sémiotisation verbale de lémotion". In: *Semen* 35. DOI: 10.4000/semen.9795. URL: https://doi.org/10.4000/semen.9795.

References XIV

Miller, George A. (1994). "WordNet: A Lexical Database for English". In: *Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey, March 8-11, 1994*. URL: https://aclanthology.org/H94-1111.

Neal, Tempestt et al. (2017). "Surveying Stylometry Techniques and Applications". In: *Acm Computing Surveys* 50.6, 86:1–86:36. ISSN: 0360-0300.

Paech, Samuel J. (2024). *EQ-Bench: An Emotional Intelligence Benchmark for Large Language Models*. arXiv. DOI: 10.48550/arXiv.2312.06281. arXiv: 2312.06281 [cs].

Plutchik, Robert (2001). "The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice". In: *American Scientist* 89.4. Publisher: Sigma Xi, The Scientific Research Society, pp. 344–350. ISSN: 00030996. URL: http://www.jstor.org/stable/27857503 (visited on June 13, 2022).

References XV

Russell, James A. and Lisa Barrett (1999). "Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant". In: *Journal of personality and social psychology* 76, pp. 805–19. DOI: 10.1037//0022-3514.76.5.805.

Schachter, Stanley and Jerome Singer (1962). "Cognitive, social, and physiological determinants of emotional state.". In: *Psychological review* 69.5. Publisher: American Psychological Association, p. 379.

Scherer, Klaus R. (2022a). "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. PMID: 35188091.

— (2022b). "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.

References XVI

Sharma, Ashish et al. (2023). "Cognitive Reframing of Negative Thoughts through Human-Language Model Interaction". In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki. Toronto, Canada: Association for Computational Linguistics, pp. 9977–10000.

Tomkins, Silvan (1962). *Affect imagery consciousness*. Vol. Volume I: The positive affects. New York: Springer. (Visited on Feb. 12, 2024).

Troiano, Enrica, Laura Oberländer, and Roman Klinger (2022). "Dimensional Modeling of Emotions in Text with Appraisal Theories: Corpus Creation, Annotation Reliability, and Prediction". In: *Computational Linguistics*, pp. 1–71. ISSN: 0891-2017, 1530-9312. DOI: `10.1162/coli_a_00461`. arXiv: `2206.05238 [cs]`.

Wales, Katie (2014). *A Dictionary of Stylistics*. Routledge. ISBN: 978-1-315-83350-7.

References XVII

White, Michael and David Epston (1990). *Narrative Means to Therapeutic Ends*. WW Norton. ISBN: 978-0-393-70098-5.

Winget, Carolyn and Milton Kramer (1979). *Dimensions of dreams*. Gainesville: University of Florida Press.

Yu, Calvin Kai-Ching (2022). "Automated analysis of dream sentiment—The royal road to dream dynamics?" In: *Dreaming : journal of the Association for the Study of Dreams* 32.1. Publisher: American Psychological Association (APA), pp. 33–51. DOI: 10.1037/drm0000189. URL: https://doi.org/10.1037/drm0000189.

Zhan, Hongli, Desmond Ong, and Junyi Jessy Li (2023). "Evaluating Subjective Cognitive Appraisals of Emotions from Large Language Models". In: *Findings of the Association for Computational Linguistics: EMNLP 2023*. Ed. by Houda Bouamor, Juan Pino, and Kalika Bali. Singapore: Association for Computational Linguistics, pp. 14418–14446. DOI: 10.18653/v1/2023.findings-emnlp.962.

References XVIII

- Zhang, Wenxuan, Yue Deng, et al. (2023). *Sentiment Analysis in the Era of Large Language Models: A Reality Check*. arXiv. DOI: 10.48550/arXiv.2305.15005. arXiv: 2305.15005 [cs].
- Zhang, Wenxuan, Xin Li, et al. (2022). "A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges". arXiv: 2203.01054 [cs].
- Zhou, Pei et al. (2023). *How FaR Are Large Language Models From Agents with Theory-of-Mind?* arXiv. DOI: 10.48550/arXiv.2310.03051. arXiv: 2310.03051 [cs].