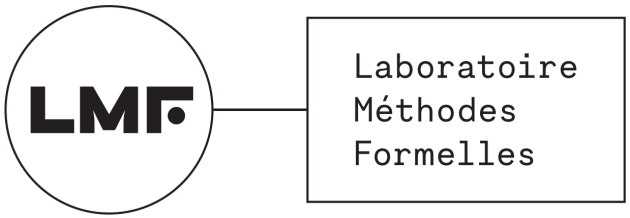
Natural language processing of subjectivity analysis in personal narratives

Gustave Cortal



Introduction

Research question: How to model subjective experience in narratives?

Steps:

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

Each step lead to a first-author article in an international conference

Definition of objectives and scope using cognitive science

Goal: Identify limitations and research directions

I review psychological theories of emotion and emotion annotation schemes in NLP

What are current **limitations**?

- Different emotion theories lead to divergences in how to annotate them in the text
- Some linguistic and cognitive science theories are not considered
- There is no benchmark that evaluates the richness of the emotional phenomenon

G. Cortal and C. Bonard. Improving Language Models for Emotion Analysis: Insights from Cognitive Science. *CMCL, ACL 2024*.

Definition of objectives and scope using cognitive science

How to integrate psychological theories of emotion?

I use the **integrated framework for emotion theories** (Scherer, 2022):

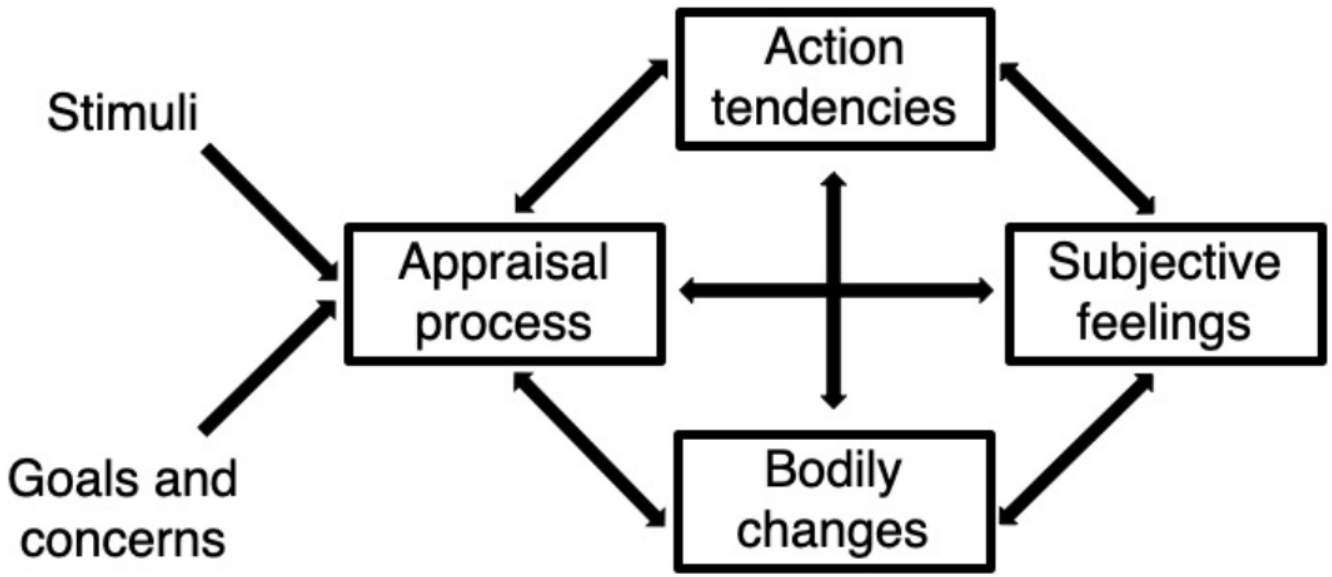


Figure: Emotional episodes are synchronized changes in four components.

Definition of objectives and scope using cognitive science

Which verbal signs are used to infer expressed emotions?

Raphaël Micheli categorizes a range of linguistic markers into three **emotion expression modes** (Micheli, 2014). 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.

Construction of an emotion dataset

Goal: 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.

More than 1,000 narratives were collected during emotion regulation sessions

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

Training language models for emotion analysis

Goal: 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

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

Training language models for emotion analysis

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

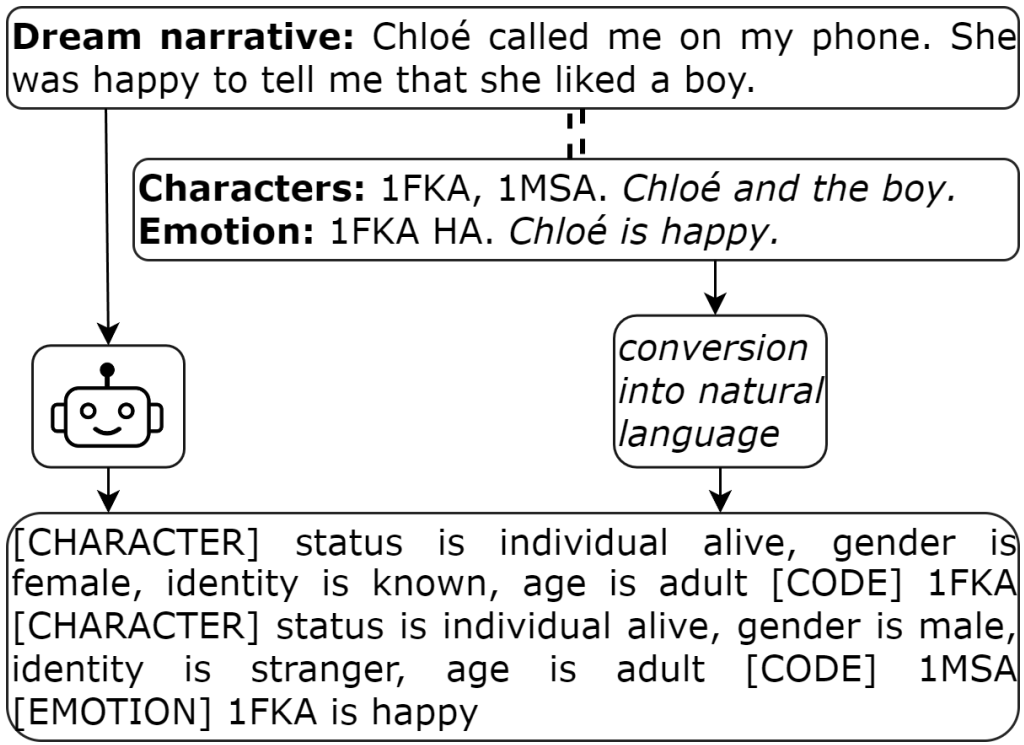
Quantitative dream analysis examines recurring patterns between narrative elements using a database of dream narratives and an annotation scheme (Domhoff, 2004)

The annotation process is complex and costly

How to automate the annotation process?

Training language models for emotion analysis

Goal: Character and emotion detection in dream narratives



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

Training language models for emotion analysis

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
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**	-
CROSS-VALIDATION	86.28	81.9	79.51	89.52	68.64	76.18

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

Formalization of style in narratives

How subjective experience is communicated?

Goal: Formalize style as patterns of linguistic choices that encode subjective experience

Clause	**Process (code)**	**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}

Table: Illustrative pipeline. Each narrative is mapped to a symbolic sequence using an alphabet based on extracted features

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

Introduction

Research question: How is subjective experience communicated in narratives?

We use narratives to express our representations of reality and make sense of the world

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

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

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

Contributions

Problem: Style is an intuitive notion; we need an operational definition

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

Research task: Formalize style as *patterns of linguistic choices that encode subjective experience*

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

Categorizing linguistic features

Our categorization is grounded in *systemic functional linguistics*: language represents experience through *processes, participants and circumstances*

Processes	**Examples**
Action: actions and events in the physical world.	[He]$_{\text{Actor}}$ [**takes**]$_{\text{Action}}$ [the valuable]$_{\text{Affected}}$ [Members of my cult]$_{\text{Actor}}$ [**have made**]$_{\text{Action}}$ [1500 euros]$_{\text{Result}}$ [I]$_{\text{Actor}}$ [**give**]$_{\text{Action}}$ [her]$_{\text{Recipient}}$ [a chance]$_{\text{Range}}$
Mental: internal experiences such as thoughts, perceptions, and feelings.	[We]$_{\text{Senser}}$ [**believe**]$_{\text{Mental}}$ [women are the leaders of change]$_{\text{Phenomenon}}$ [The moon]$_{\text{Senser}}$ [**sees**]$_{\text{Mental}}$ [the earth]$_{\text{Phenomenon}}$ [He]$_{\text{Senser}}$ [**disliked**]$_{\text{Mental}}$ [Gilbert's writing]$_{\text{Phenomenon}}$
Verbal: acts of communication.	[David]$_{\text{Sayer}}$ [**said**]$_{\text{Verbal}}$ [“the corrupt, criminals and money launderers”]$_{\text{Verbiage}}$
State: states of being, having, or existence.	There [**was**]$_{\text{Existential}}$ [a swimming pool]$_{\text{Existent}}$ [John]$_{\text{Carrier}}$ [**is**]$_{\text{State}}$ [an interesting teacher]$_{\text{Attribute}}$ [Hadrian's Wall]$_{\text{Possessor}}$ [**has**]$_{\text{State}}$ [something for everyone]$_{\text{Possessed}}$

Table: Processes with their participants.

Pipeline for our sequence-based framework

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}

1. We first segment *"I wake in a dark room. I feel a cold wind. I tell myself to move."* into clauses
2. Identify features (*e.g.*, processes and participants) for each clause
3. Each narrative is mapped to a symbolic sequence using an alphabet based on identified features

Case Study on Dream Narratives

We apply our framework to dream narratives as they possess a narrative structure and represent attempts to communicate subjective experience

We use DreamBank, a database of more than 27,000 narratives with 72 series of dreamers

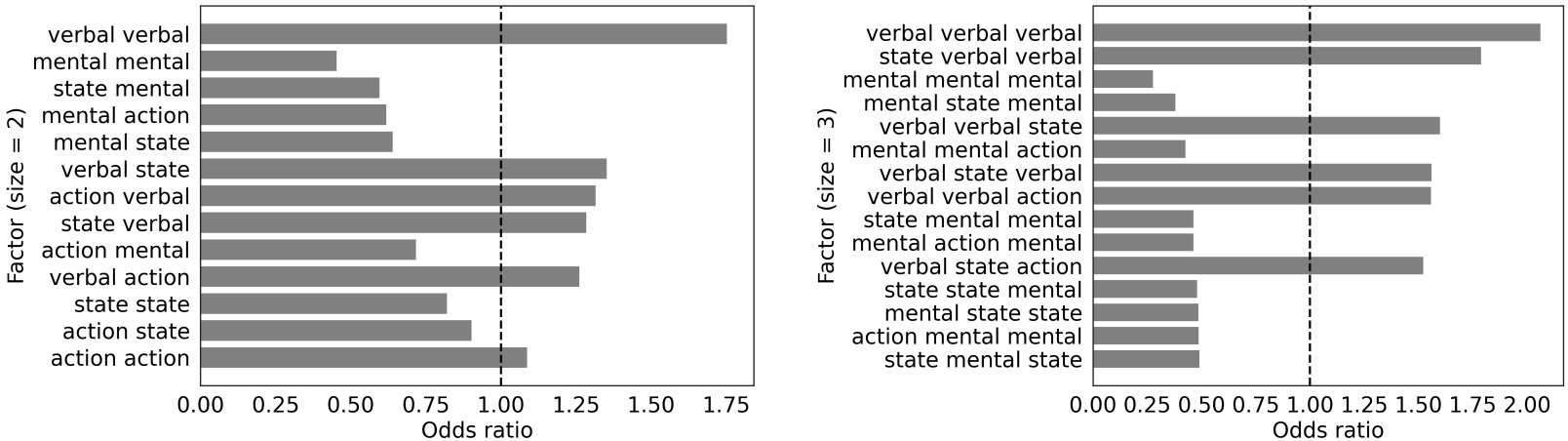
We analyze five series of dreamers: long-term blind dreamers (n=361), *ed* (a widower, n=139), *izzy* (a teenager, n=1091), *merri* (an artist, n=202), and *viet* (a Vietnam War veteran with PTSD, n=566)

We construct a *norm* (n=720) to compare how each series deviates from a hypothetical average dreamer

To identified features, we perform in-context learning with Llama 3 8B

Results on the Vietnam 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

We show a preference for *viet* to remain in a verbal process, as indicated by substrings such as *verbal.verbal* and *verbal.verbal.verbal* with high odds ratios (respectively 2.00 and 1.75)

Results on the Vietnam War veteran

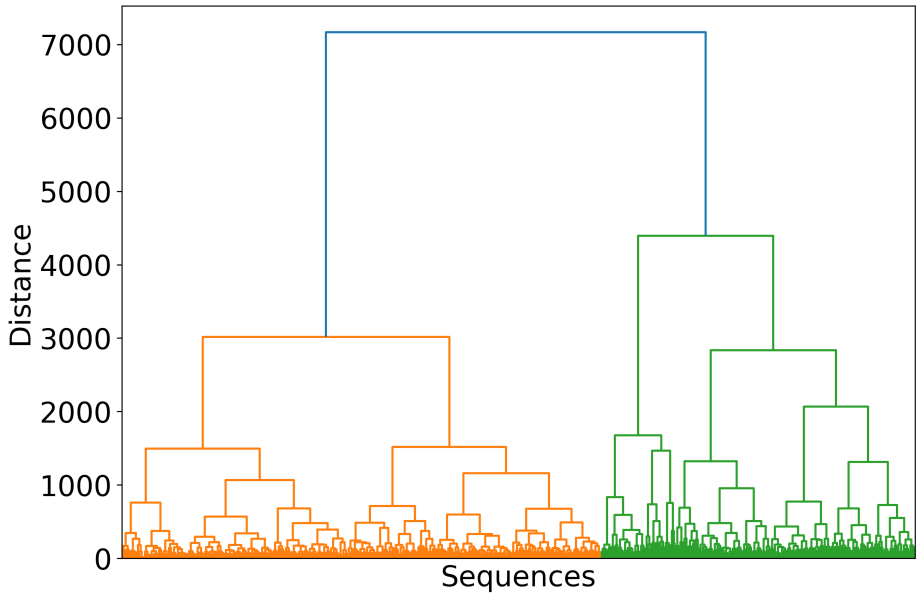


Figure: Dendrogram with Ward linkage and cosine similarity

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

Two templates: a highly action-oriented structure or a more varied structure alternating between state and action processes

Perspectives

- **Authorship profiling**: identifying signature patterns (*e.g.*, distinctive substrings) that characterize an author's unique way of constructing narratives
- **Style-conditioned narrative generation**: generating narratives from a sequence of linguistic features
- **Applying methods from complexity science and formal language theory**: analyzing subsequences, using complexity measures to quantify redundancies, etc.

Conclusion

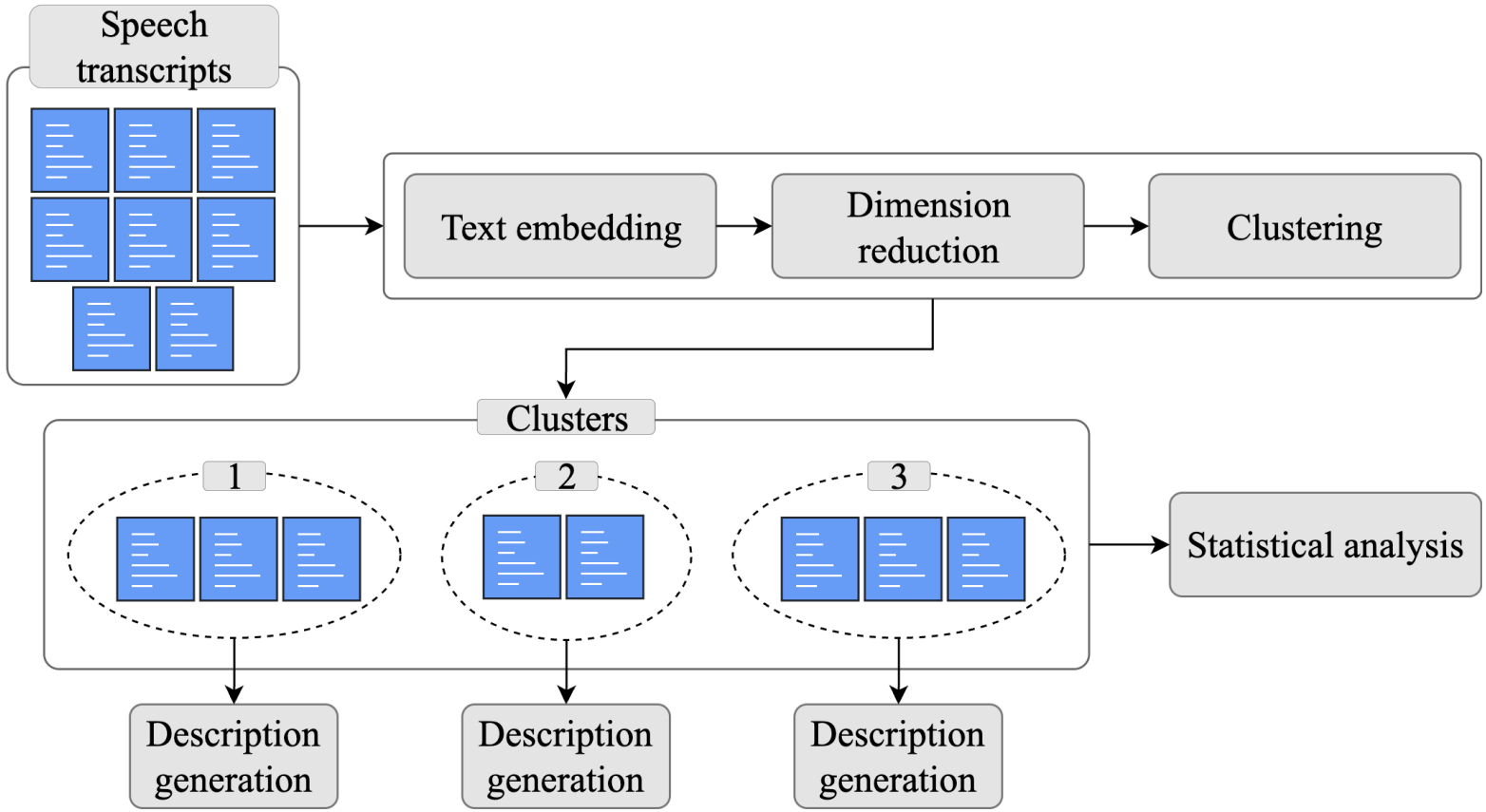
Research question: How to model subjective experience in narratives?

Steps:

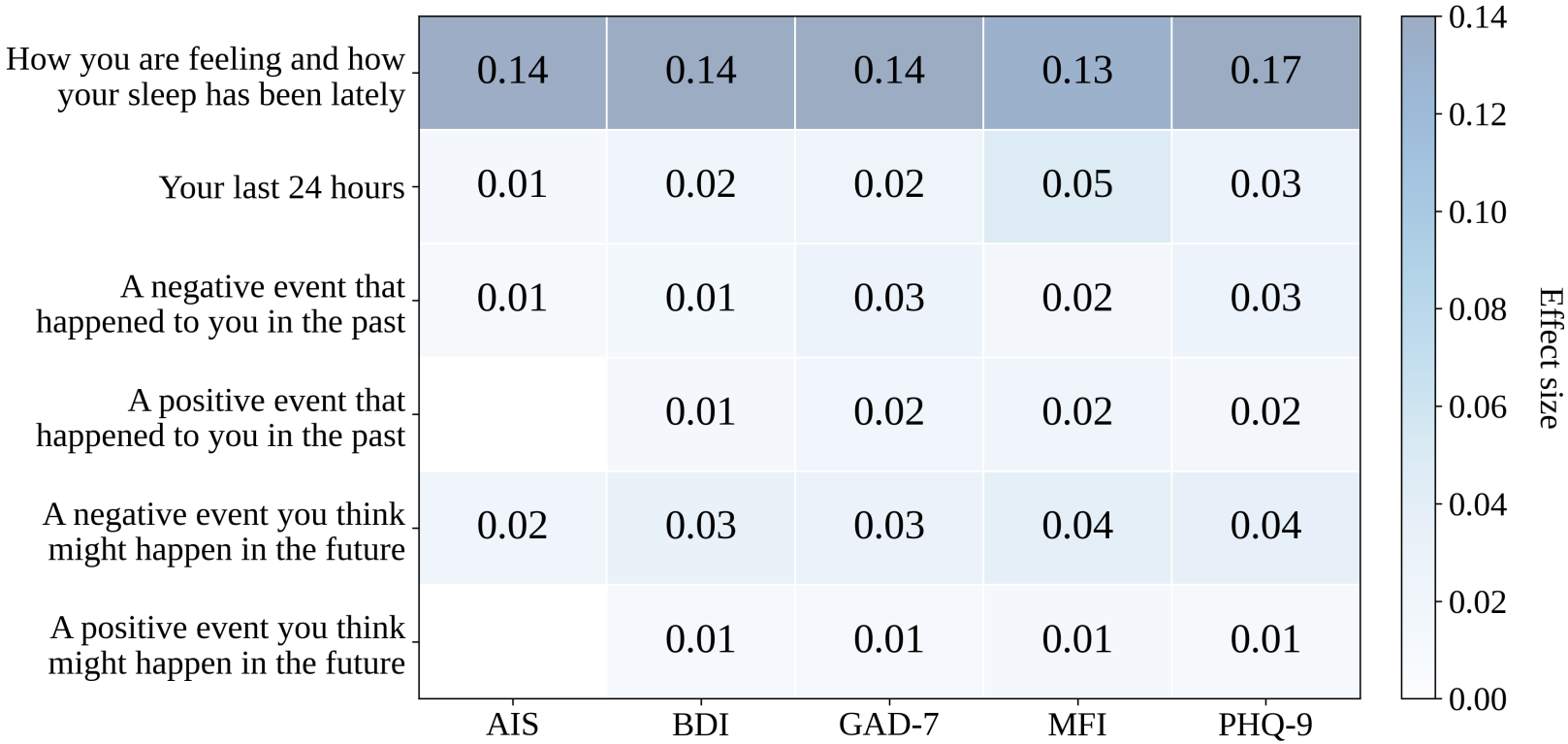
- Definition of objectives and scope using cognitive science
- Construction of an emotion dataset
- Training of language models for emotion analysis
- Formalization of style in narratives

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

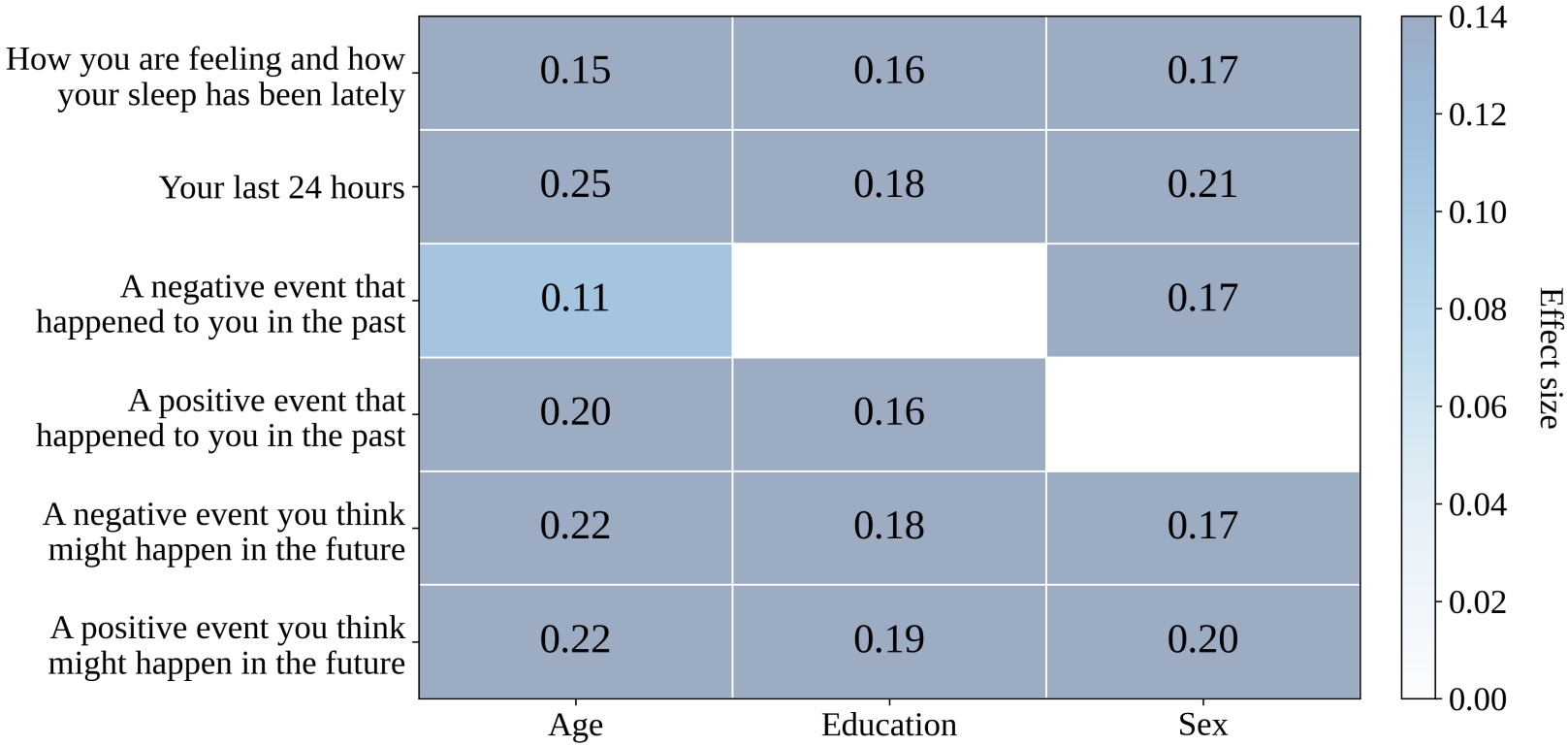
Pipeline for semantic clustering and description generation



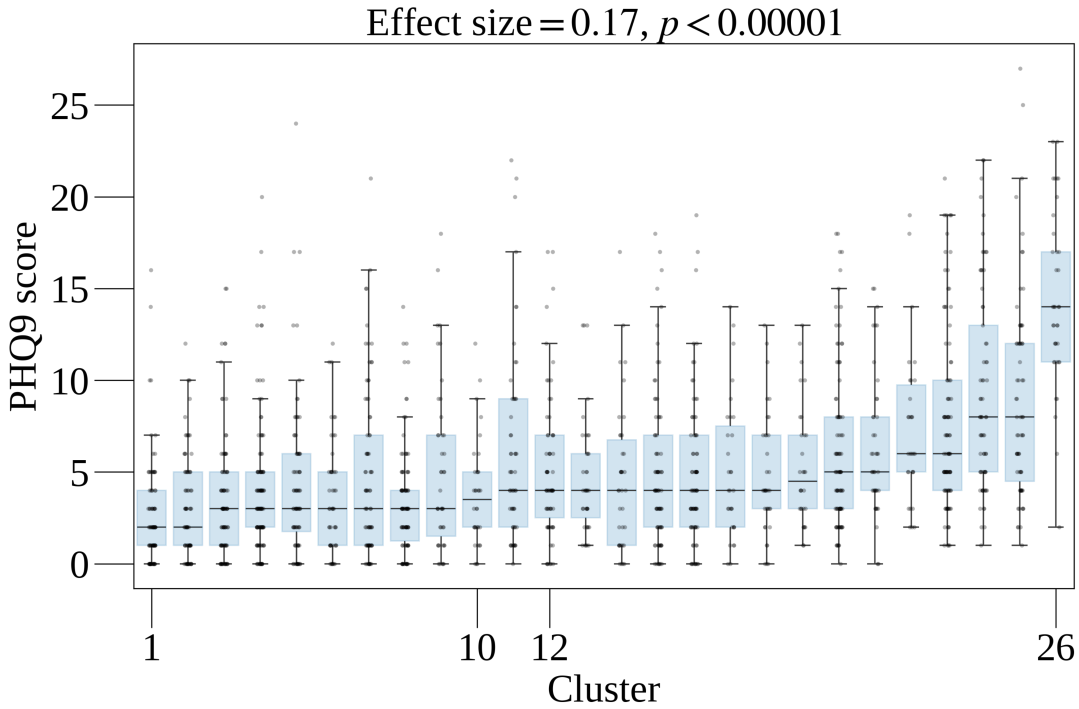
Effect size across tasks and clinical scores



Effect size across tasks and demographics scores



Fine-grained mental health topic modeling



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

Appendix

- *Fine-grained mental health topic modeling in different populations using large language models* (PhD internship, to be submitted for Nature Mental Health)
- Piaget, a language model for psychological and philosophical reasoning, 2025.
- N. Richet, S. Belharbi, H. Aslam, M. Schadt, M. González-González, **G. Cortal**, A. Koerich, M. Pedersoli, A. Finkel, S. Bacon, E. Granger. Textualized and Feature-based Models for Compound Multimodal Emotion Recognition in the Wild. *ABAW, ECCV 2024*.