Formalizing Style in Personal Narratives

Gustave Cortal[1,2], Alain Finkel[1]

[1]Université Paris-Saclay, CNRS, ENS Paris-Saclay, LMF, 91190, Gif-sur-Yvette, France

[2]Université Paris-Saclay, CNRS, LISN, 91400, Orsay, France

Introduction

Research question: How is subjective experience communicated in narratives?

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
Task: Formalize style as *patterns of linguistic choices encoding subjective experience*

Our contributions are:

- A sequence-based framework defining style as patterns of linguistic choices;
- A methodology for automatically identifying patterns using sequence analysis;
- A case study on dream narratives.

Categorizing linguistic features

According to *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]$_{\text{Affected}}$ [I]$_{\text{Actor}}$ [**give**]$_{\text{Action}}$ [her]$_{\text{Recipient}}$ [a chance]$_{\text{Range}}$
Mental: internal experiences such as thoughts, perceptions, and feelings.	[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.	[Gustave]$_{\text{Sayer}}$ [**said**]$_{\text{Verbal}}$ ["everything will be okay"]$_{\text{Verbiage}}$
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]$_{\text{Attribute}}$

To identify features, we perform in-context learning with a language model (Llama 3 8B)

Sequence-based framework

We first *segment a narrative* into clauses, then *identify features* such as processes and participants for each clause. Each narrative is mapped to a *symbolic sequence* based on identified features

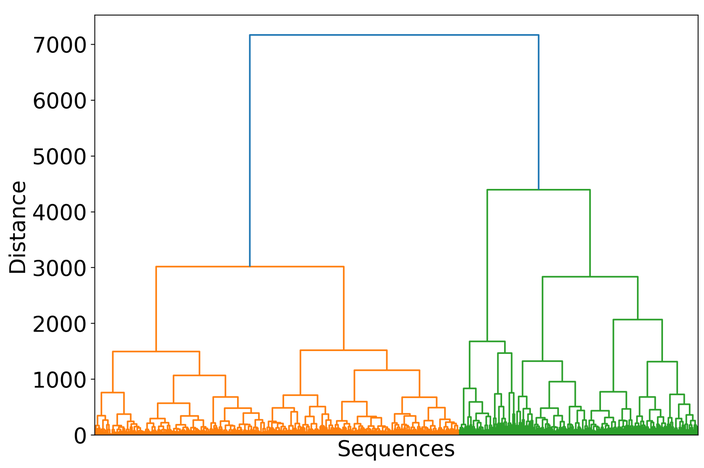
Clause	**Process (symbol)**	**Participants**
I wake in a dark room	Action (**a**)	Actor
I feel a cold wind	Mental (**m**)	Senser, Phenomenon
I tell myself to move	Verbal (**v**)	Sayer, Recipient

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

Representative sequences

Hierarchical clustering with cosine similarity between substrings of sequences

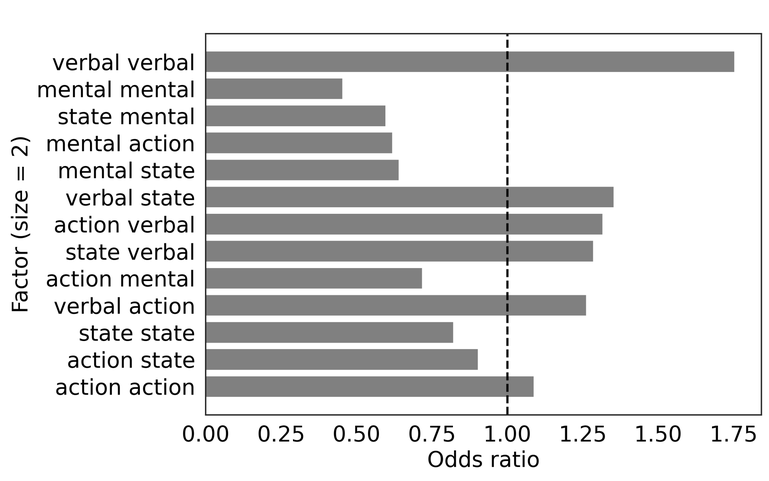
Representative sequences for the two clusters: savamasasaaamaaasavvvaaaaaaavssaaaaa and sssssavaavssvsavvvvsmasasaasasaamaamvmsss with $a = action, m = mental, s = state, v = verbal$



Proportions of substrings deviating from the average

We compare the proportions of sequences containing a given substring between the War veteran and the average dreamer

The *odds ratio* measures how much more (or less) likely it is to observe a given substring in one dreamer compared to another



Perspectives

- *Authorship profiling*: identifying signature patterns 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.

contact: gustave.cortal@ens-paris-saclay.fr