UniCoRN: Unified Cognitive Signal ReconstructioN bridging cognitive signals and human language

Nuwa Xi, Sendong Zhao[*]**, Haochun Wang, Chi Liu, Bing Qin and Ting Liu**
Research Center for Social Computing and Information Retrieval,
Harbin Institute of Technology, China
{nwxi,sdzhao,hcwang,cliu,bqin,tliu}@ir.hit.edu.cn

Abstract

Decoding text stimuli from cognitive signals (e.g. fMRI) enhances our understanding of the human language system, paving the way for building versatile Brain-Computer Interface. However, existing studies largely focus on decoding individual word-level fMRI volumes from a restricted vocabulary, which is far too idealized for real-world application. In this paper, we propose fMRI2text, the first open-vocabulary task aiming to bridge fMRI time series and human language. Furthermore, to explore the potential of this new task, we present a baseline solution, UniCoRN: the **Uni**fied **Co**gnitive Signal **R**econstructio**N** for Brain Decoding. By reconstructing both individual time points and time series, UniCoRN establishes a robust encoder for cognitive signals (fMRI & EEG). Leveraging a pre-trained language model as decoder, UniCoRN proves its efficacy in decoding coherent text from fMRI series across various split settings. Our model achieves a 34.77% BLEU score on fMRI2text, and a 37.04% BLEU when generalized to EEG-to-text decoding, thereby surpassing the former baseline. Experimental results indicate the feasibility of decoding consecutive fMRI volumes, and the effectiveness of decoding different cognitive signals using a unified structure.

1 Introduction

Language serves as a window into the cognitive processes unfolding within our minds, communicating a vast amount of information through its syntax and semantics (Pagel, 2017). Advances in cognitive neuroscience have enabled us to directly observe the cognitive processes that underlie language use through the analysis of non-invasive cognitive signals, such as functional Magnetic Resonance Imaging (fMRI) and electroencephalogram (EEG). However, this also poses a challenge in understanding the relationship between these signals and the external stimuli that give rise to them within the mind. Deciphering cognitive signals into human language not only enhances our grasp of the linguistic system, but also facilitates the development of practical brain-computer interfaces (BCIs) by leveraging our comprehension of decoded signals (Wolpaw, 2007; Mudgal et al., 2020).

Although brain decoding has gained great success from word-level to sentence-level decoding on EEG (Panachakel and Ramakrishnan, 2021; Wang and Ji, 2022), relatively little research has been dedicated to directly generating text, particularly complete sentences, from fMRI volumes. This is largely attributed to the challenges posed by the relatively low temporal resolution of fMRI, which makes it challenging to acquire word-level fMRI frames within a sentence. In this study, we propose fMRI2text, the first open-vocabulary task that decodes fMRI time series into the corresponding texts under naturalistic settings.

Despite the early efforts in fMRI decoding (Mitchell et al., 2008; Palatucci et al., 2009; Wang et al., 2020; Zou et al., 2021), these methods are limited in the ways that they: (1) primarily rely on predefined regions of interest (ROIs) for feature extraction, underutilizing the rich spatial data inherent in full fMRI volumes. This may oversimplify the complex, distributed nature of cognitive processes (Ruiz et al., 2014). (2) do not effectively leverage the sequential information embedded in fMRI time series, missing valuable insights into the dynamics of cognitive processes (Du et al., 2022). (3) prioritize the role of the decoder while overlooking the importance of efficient encoding, particularly for high-dimensional signals like fMRI. These limitations extend beyond fMRI decoding and apply to other cognitive signal decoding methods as well. To address these issues and obviate the need for separate, complex pipelines to decode specific cognitive signals, we propose UniCoRN (**Uni**fied **Co**gnitive signal **R**econstructio**N** for brain decoding), a versatile brain decoding pipeline that

[*]Corresponding author

Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics
Volume 1: Long Papers, pages 13277–13291

can be applied to various types of cognitive signals.

As a standard encoder-decoder framework, UniCoRN leverages the robust decoding abilities of pre-trained language models. Crucially, it constructs an effective encoder through both snapshot and series reconstructions, harnessing the power of seq2seq models. This allows UniCoRN to analyze individual signal "snapshots" (such as a single fMRI volume or an EEG time point) and capture the "series" or temporal dependencies among these snapshots, thus maximizing the information extracted from the cognitive signals.

In summary, our contributions are as follows:

- We introduce a novel task, designated as fMRI2text, which is the first open-vocabulary task that decodes fMRI time series into human language in a naturalistic context.

- We present a baseline solution to further elucidate the potential of fMRI2text and demonstrate that our proposed method is effective across various split settings.

- We propose a unified framework UniCoRN (**Uni**fied **Co**gnitive signal **R**econstructio**N** for brain decoding) to translate cognitive signals into human language, and validate its effectiveness on both EEG and fMRI.

2 Related Work

Cognitive Signals Cognitive signals represent the dynamic neural activity associated with information processing and cognitive functions, and are crucial in building BCI systems (Mudgal et al., 2020). These signals are captured at individual time points or as part of a time series, with each data point providing a snapshot of brain activity at a specific point in time. While EcoG is often used in high-performance BCI systems (Akbari et al., 2019; Rapeaux and Constandinou, 2021; Metzger et al., 2022), its semi-invasive nature limits its potential for widespread application in healthy individuals. In non-invasive BCI systems, EEG is most commonly used due to its high temporal resolution and cost-effectiveness, while other techniques such as fMRI have also been employed in recent years (Saha et al., 2021; Martinek et al., 2021; Pitt and Dietz, 2022). In spite of its relatively lower temporal resolution, fMRI allows for the mapping of brain-wide responses to linguistic stimuli at a highly detailed spatial resolution of millimeters (Vouloumanos et al., 2001; Noppeney and Price, 2004; Binder et al., 2009). This makes fMRI particularly ideal for BCI systems that translate brain signals into text, a process that involves the participation of multiple brain regions (Ruiz et al., 2014).

Brain Decoding Recent research has been directed towards resolving the issue of decoding cognitive signals into human language through the introduction of new multi-modal tasks and models. Most recently-proposed tasks in this field focus on aligning cognitive signals with a limited vocabulary up to a thousand for word-level decoding (Bhattasali et al., 2019; Affolter et al., 2020; Défossez et al., 2022) or incorporating them into sentence embeddings for sentence-level decoding using pairwise classification (Pereira et al., 2018; Sun et al., 2019). Wang and Ji introduce a novel brain decoding task called EEG-To-Text decoding (EEG2text for short), which achieves sentence-level decoding by converting each word-level EEG signal into corresponding text stimuli using pre-trained language models, thereby extending the problem from a closed vocabulary to an open vocabulary.

3 Task Definition

As shown in Figure 1, the subject is instructed to read or listen to the text stimuli, while an fMRI volume is acquired every fixed repetition time (TR). Given an fMRI time series of length $\mathcal{T}$, $\mathcal{F} := \{f_1, f_2, ..., f_{\mathcal{T}}\}$, the task is to decode the corresponding text tokens $\mathcal{W} := \{w_1, w_2, ..., w_n\}$ of the stimuli used during the acquisition of the fMRI volumes from an open vocabulary $\mathcal{V}$.

As mentioned in Section 2, many studies have aimed to link cognitive signals with human language. We summarize three related tasks and compare them to fMRI2text in Table 1.

The three representative tasks share a common characteristic of relying on cognitive signals that operate at the word level. However, this approach may not be practical for real-world fMRI applications, as the poor temporal resolution of fMRI necessitates tightly controlled experimental manipulations under these settings (Nastase et al., 2020; Hamilton and Huth, 2020). In contrast, fMRI2text leverages cognitive signals from more naturalistic settings, using text and speech as stimuli in a manner closer to real-world language use (Huth et al., 2016). Here, each fMRI frame corresponds to a specific timeframe, and is aligned with an undetermined number of tokens rather than a fixed one,

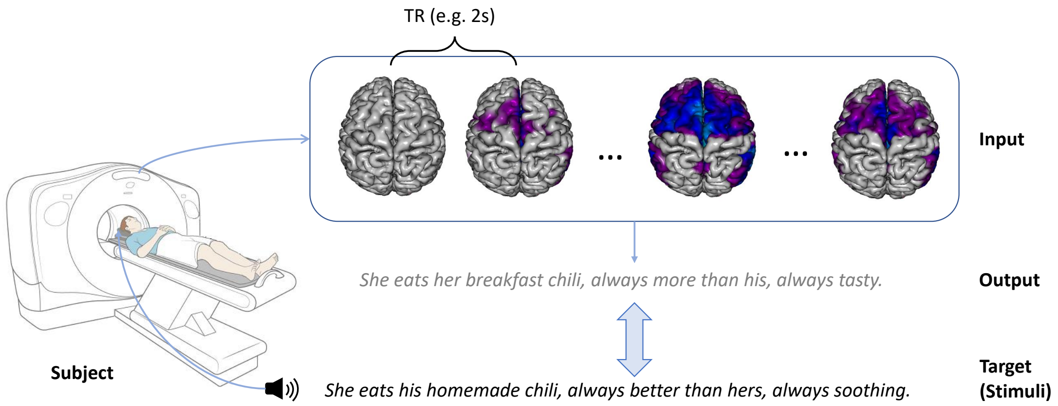

Figure 1: Task definition for fMRI2text

Task Name	Input	Target
Open Vocabulary EEG-To-Text Decoding	a sequence of word-level EEG features $\mathcal{E} := \{e_1, e_2, ..., e_n\}$	the corresponding text tokens $\mathcal{W} = \{w_1, w_2, ..., w_n\}$
fMRI-Conditioned Mask-Filling	an fMRI image $\mathcal{F}$ and a sentence $\mathcal{W} := \{w_1, w_2, ..., < mask >, ..., w_n\}$, where the corresponding word is masked	the word masked in sentence $\mathcal{W}$
fMRI-Conditioned Text Generation	an fMRI image $\mathcal{F}$ and a prefix $\mathcal{W}' := \{w_1, w_2, ..., w_k\}$	$\mathcal{W} := \{w_1, ..., w_k, ..., w_m\}$ where the corresponding word is contained
Open Vocabulary fMRI2text Decoding	a fixed-length sequence of $\mathcal{T}$ chronically consistent fMRI $\mathcal{F} := \{f_1, f_2, ..., f_{\mathcal{T}}\}$	the correspondent text tokens $\mathcal{W} := \{w_1, w_2, ...w_n\}$

Table 1: Input and target output for representative brain decoding tasks.

better reflecting the variable and dynamic nature of natural language processing.

Another distinct feature that differentiates fMRI2text from prior fMRI-related tasks is its incorporation of multiple sequential frames as input. The inherent low signal-to-noise ratio of fMRI has directed prior studies towards a focus on individual frames. However, this approach overlooks the valuable temporal information embedded within the interrelations of successive frames, which is particularly crucial when dealing with continuous data streams such as cognitive signals.

4 Method

In this section, we introduce the UniCoRN structure and use the fMRI2text task as an explicit demonstration. As shown in Figure 2, UniCoRN consists of two stages: (1) the cognitive signal reconstruction to train the encoder specifically for cognitive signals, and (2) the cog2text decoding to convert the embeddings of the cognitive signals from the first stage to human language.

4.1 Cognitive Signal Reconstruction

The cognitive signal reconstruction consists of two phases, snapshot reconstruction and series reconstruction, aiming to train the encoder of UniCoRN to integrate the individual characteristics of each fMRI volume (intra-volume information), as well as the temporal relationships among volumes in a time series (inter-volume information).

As shown in Figure 2, during the snapshot reconstruction, each fMRI frame is input into the Snapshot Encoder $\mathcal{E}_r$ ($\mathcal{E}_{reconstruction}$) respectively to obtain the snapshot embedding E^i, which will be used later for series reconstruction. In our case, we use a CNN-based model similar to Malkiel et al. (2021) as the Snapshot Encoder. During this phase, E^i is then fed to the Snapshot Decoder $\mathcal{D}_r$ ($\mathcal{D}_{reconstruction}$) to reconstruct the original fMRI frame f_k (The k-th frame in fMRI time series). Note that $\mathcal{D}_r$ is also CNN-based but simpler than

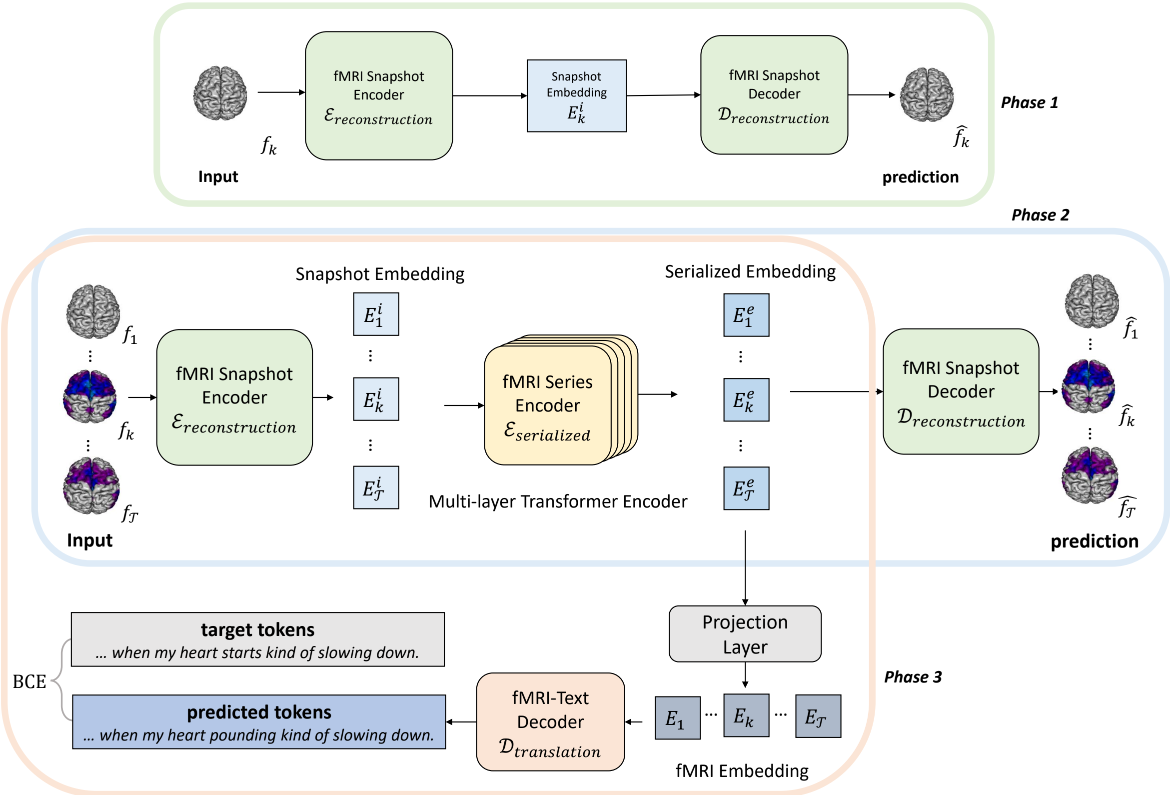

Figure 2: Illustration of UniCoRN structure. Here we refer to snapshot reconstruction, series reconstruction and cog2text decoding as *Phase 1*, *Phase 2*, *Phase 3* respectively. The snapshot encoder $\mathcal{E}_r$ from *Phase 1* (green box) and the series encoder $\mathcal{E}_s$ from *Phase 2* (blue box) together yield the fMRI embeddings for fMRI time series, which are then input to the decoder $\mathcal{D}_t$ to generate the corresponding text tokens in *Phase 3* (pink box).

$\mathcal{E}_r$ in structure, to ensure that the reconstruction of fMRI snapshots does not mostly rely on the decoding ability of $\mathcal{D}_r$. We use mean average error (MAE) as the loss function for both phases of cognitive signal reconstruction. *Phase 1* can be formulated as follows:

$$E_k^i = \mathcal{E}_r(f_k) \tag{1}$$

$$\mathcal{E}_r = \arg\min_{\mathcal{E}} \text{MAE}(\mathcal{D}_r(\mathcal{E}_r(f_k)), f_k) \tag{2}$$

During *Phase 2*, Series Encoder $\mathcal{E}_s$ ($\mathcal{E}_{serialized}$) takes the snapshot embedding E^i of $\mathcal{T}$ sequential fMRI frames to generate the corresponding serialized embedding E^e. We use multi-layer transformer encoder (Vaswani et al., 2017) as $\mathcal{E}_s$ to obtain information in time domain by applying self-attention to fMRI series. Serialized embedding E^e is then input into the same decoder as *Phase 1* for series reconstruction. We continue using $\mathcal{D}_r$ as the decoder to keep minimal effect of decoding process to signal reconstruction, as we will only be using $\mathcal{E}_r$ and $\mathcal{E}_s$ in the next stage. Denote $\{E_k^e, E_{k+1}^e, ..., E_{k+\mathcal{T}-1}^e\}$ as $E_{k\sim\mathcal{T}}^e$, $\{E_k^i, E_{k+1}^i, ..., E_{k+\mathcal{T}-1}^i\}$ as $E_{k\sim\mathcal{T}}^i$.

$$E_{k\sim\mathcal{T}}^e = \mathcal{E}_s(E_{k\sim\mathcal{T}}^i) \tag{3}$$

$$\mathcal{E}_s = \arg\min_{\mathcal{E}} \text{MAE}(\mathcal{D}_r(\mathcal{E}_s(E_{k\sim\mathcal{T}}^i)), E_{k\sim\mathcal{T}}^i) \tag{4}$$

4.2 Cog2text Decoding

The motivation of cognitive signal reconstruction is to get a decent representation of fMRI, which is quite so different from and more difficult than EEG since each fMRI frame has more spatial information as a 3D signal. Similar to Wang and Ji (2022), we use this representation as primary word embeddings for language models, except that these embeddings have been denoised and condensed through reconstruction. The high-level idea here is that we consider each original frame of fMRI as a word-level representation of “the foreign language spoken by the human brain”, and use the encoder constructed in Section 4.1 to obtain the embeddings of this “language”, which will be then decoded to real human language (English in our case) like traditional machine translation tasks.

Figure 2 gives a detailed demonstration of how fMRI embedding is acquired and how the two stages are concatenated together. After the two phases of cognitive signal reconstruction, the decoder $\mathcal{D}_r$ used in stage one is replaced with the fMRI-Text decoder $\mathcal{D}_t$ ($\mathcal{D}_{translation}$) for text generation. The serialized embeddings E^e are then projected into fMRI embedding E as the final representation of fMRI, which contains both intra-volume information and inter-volume information and will be used as the input for $\mathcal{D}_t$ to convert to texts. Here we use BART (Lewis et al., 2019) as the fMRI-Text decoder $\mathcal{D}_t$ and cross-entropy loss (CE) like most seq2seq tasks as the training target. Denote $\{E_k, E_{k+1}, ..., E_{k+\mathcal{T}-1}\}$ as $E_{k\sim\mathcal{T}}$, and the projection layer matrix as W^P.

$$E_{k\sim\mathcal{T}} = E^e_{k\sim\mathcal{T}} W^P \tag{5}$$

$$\mathcal{D}_t = \arg\min_{\mathcal{D}} \mathrm{CE}(\mathcal{D}(E_{k\sim\mathcal{T}}), \mathcal{W}) \tag{6}$$

4.3 UniCoRN Structure

Other than fMRI, UniCoRN is also capable of decoding other cognitive signals into human language. We generalize the same pipeline to EEG2text, without changing the overall structure but only moderately modifying the snapshot encoder $\mathcal{E}_r$ and snapshot decoder $\mathcal{D}_r$ due to the difference in spatial structure between EEG and fMRI. The detailed illustration is provided in Appendix D.

5 Experiments

5.1 Dataset

The "Narratives" dataset (Nastase et al., 2021) encompasses a range of fMRI data from individuals who were engaged in listening to spoken stories in the real-world setting. Given that various fMRI machines produce frames of different sizes, and considering the "Narratives" dataset comprises data from multiple machines, we focus solely on data with dimensions of $64 \times 64 \times 27$ voxels. The detailed information of the "Narratives" dataset we used in this paper is provided in Appendix C.

Most cognitive signals require pre-processing before putting into use. For fMRI, We follow the same pre-processing procedure as provided in Nastase et al. (2021). As for EEG, we use the same waves as in Wang and Ji (2022) for comparison.

Given that the "Narratives" dataset does not offer any pre-determined splits and the appropriate method for splitting fMRI data for this task is a matter of debate, we conduct experiments utilizing a variety of different split configurations.

Split Method	Test Set
random	$\{F^{ij}_{k\sim\mathcal{T}} \mid F^{ij}_{k\sim\mathcal{T}} \notin \mathbb{F}_{Tr}\}$
random time	$\{F^{ij}_{k\sim\mathcal{T}} \mid \forall j, k \notin T^j_{Tr}\}$
consecutive time	$\{F^{ij}_{k\sim\mathcal{T}} \mid \forall j, \forall t \in T^j_{Tr}, t < k\}$
by stimuli	$\{F^{ij}_{k\sim\mathcal{T}} \mid j \notin \mathcal{C}_{Tr}\}$
by subject	$\{F^{ij}_{k\sim\mathcal{T}} \mid i \notin \mathcal{S}_{Tr}\}$

Table 2: Splitting method for fMRI2text. Detailed notations are further explained in Appendix B.

Denote all subjects as $\mathcal{S} := \{S_1, S_2, ..., S_n\}$, all stimuli as $\mathcal{C} := \{C_1, C_2, ..., C_m\}$, where n and m stands for the total number of subjects and stimuli respectively. Note that the total number of stimuli given to individual subjects may vary. The fMRI series of subject S_i receiving stimuli C_j is represented as $\mathcal{F}^{ij} := \{f^{ij}_1, f^{ij}_2, ..., f^{ij}_{T_j}\}$. T_j here represents the total number of fMRI frames of stimuli j. For briefings, we use $F^{ij}_{k\sim\mathcal{T}}$ to represent the fMRI series of length $\mathcal{T}$ starting at the k th frame $\{f^{ij}_k, f^{ij}_{k+1}, ..., f^{ij}_{k+\mathcal{T}-1}\}$. Different split methods are formulated in detail in Table 2.

As for EEG2text, We use ZuCo1.0 datasets (Hollenstein et al., 2018), which comprises EEG recordings obtained from natural reading tasks, including both Normal Reading (NR) and Task-Specific Reading (TSR). The reading materials utilized for these tasks were sourced from movie reviews (Socher et al., 2013) and Wikipedia articles. The ZuCo1.0 dataset comprises a total of 1,107 unique sentences across 12 subjects, yielding a total of 10,258 samples. Given the limited number of training samples, we utilize a split method similar to the *random* method described above.

5.2 Implementation

Our model utilizes the Pytorch-based (Paszke et al., 2019) Huggingface Transformers (Wolf et al., 2020) packages and is designed to reconstruct sequences with a length of 5 for fMRI2text and 10 for EEG2text in *Phase 2*. Additional hyperparameters can be found in Appendix A. Both datasets are split into *train*, *validation*, *test* sets with a ratio of 70%, 15%, 15% respectively. We follow the same evaluation strategy as Wang and Ji (2022) to establish a fair comparison and gain insights into the optimal performance scenario of UniCoRN. The

Method	BLEU-N (%)				ROUGE-1 (%)		
	BLEU-1	BLEU-2	BLEU-3	BLEU-4	F	P	R
random	65.64	**52.51**	**44.96**	**39.74**	60.74	63.63	58.44
random time	62.90	49.00	40.59	34.77	59.52	62.65	56.91
consecutive time	28.21	9.23	4.27	1.83	21.88	25.84	19.12
by stimuli	26.29	6.66	2.26	0.53	23.72	30.74	19.40
by subject	**66.10**	52.32	43.78	37.78	**62.68**	**66.06**	**59.88**

Table 3: Results of UniCoRN for fMRI2text on different split settings.

Method	BLEU-N (%)				ROUGE-1 (%)		
	BLEU-1	BLEU-2	BLEU-3	BLEU-4	F	P	R
1	39.16	9.62	3.47	1.09	11.00	12.74	10.38
3	25.17	9.89	5.05	2.75	19.46	17.05	23.15
5	44.78	24.95	15.75	10.58	36.49	39.90	33.95
8	49.66	30.71	21.10	15.44	43.75	48.14	40.38
10	**62.90**	**49.00**	**40.59**	**34.77**	**59.52**	62.65	**56.91**
12	62.02	47.35	38.77	33.04	59.02	**63.09**	55.65
14	58.58	42.27	33.07	27.09	54.78	59.39	51.10
16	51.14	32.58	22.87	17.17	46.45	51.47	42.53

Table 4: Results of UniCoRN for fMRI2text on different series length $\mathcal{T}$.

results reported are the average of three separate runs. All experiments were conducted on NVIDIA A100-80GB-PCIe GPUs.

5.3 UniCoRN Structure for fMRI2text

fMRI2text across Different Splits We experiment with series length $\mathcal{T}$ of 10 and report the BLEU scores and ROUGE-1 scores for fMRI2text across different splits. As shown in Table 3, UniCoRN achieves fairly effective results across all splitting methods introduced in Section 5.1. Meanwhile, to have an intuitive grasp of the decoding quality, we present a few cases comparing the target tokens and the predicted tokens in Table 5.

The experiments conducted under the *random*, *random time*, and *by subject* settings resulted in BLEU-4 scores of 39.74%, 34.77%, and 37.78% respectively. These results shed light on the prospect of fMRI2text when viewing it as a translation-like task, particularly in comparison to state-of-the-art results in machine translation, such as 46.40% for English-French translation as reported by Liu et al. (2020) and 15.20% for English-Arabic translation as reported by Provilkov et al. (2019).

In contrast, the results obtained under the *by stimuli* and *consecutive time* settings are less so ideal. This may be attributed to the fact that the input fMRI frames do not correspond to a fixed and predetermined set of words. Consequently, the fMRI embeddings learned by the model may represent an imprecise combination of words rather than specific, individual words. Such variability might pose a challenge when the model encounters frames paired with unique word combinations unseen during training. Nonetheless, this does not preclude UniCoRN's ability to extract meaningful information under these conditions. As shown in Table 5, despite a decline in decoding quality under these two methods, UniCoRN is still successful in identifying key words within the text fragments, and maintains a semblance of polarity and structure that resonates with the target sentence.

One thing to notice is that the results under the *by subject* split setting do not show a significant deviation from those under the *random* and *random time* settings. This contrasts with previous studies that relied on individual fMRI frames for decoding, which suggests that UniCoRN's incorporation of inter-volume information can mitigate the effects of inter-subject variability on decoding performance.

Another interesting anomaly is that, despite that both the *random time* and *consecutive time* configurations have distinct text content across their *train*, *validation*, and cons*test* sets, the former setting performs significantly better than the latter. This discrepancy may be attributed to the robust

Split Method	$\mathcal{T}$	Results
consecutive time	10	T: the policeman, um, he **doesn't** even **say** anything to **Sherlock**... P: and first, the, she just **doesn't** *talk* though **Sherlock**...
by stimuli	10	T: I think it's **some sort of** mass hyp**nosis** or something... P: and you *a sort of* the You**nosis** session something...
random time	1	T: He woke up early the **next morning** P: I's up and **morning** *other day*
random time	3	T: she put her **arm** through mine and squeezed it a little bit. P: I says her *shoulder* through mine and I it a little bit
random time	5	T: Um, it was an **extremely Darwinian moment** for me, uh, **because**... P: I and, like *best* **Darwinian moment** for me, and, *for*...

Table 5: Case Analysis for fMRI2text. The target sentence is denoted as T, and the predicted sentence is represented by P. Text fragments in the target sentence to be compared are in **bold** font. Exact matches between the target and predicted sentences are indicated in **bold**, while semantic similarity is shown in *italic* font.

(1)	T: Stephen Rea, **Aidan Quinn**, and Alan Bates play Desmond's legal **eagles**... P: He Hara, **Aidan Quinn**, and Alan Bates play Desmond's legal **eagles**... B: He Baldwina, *Longan* shows, and Alan Lloyd play Hannibal's legal eternally...
(2)	T: the **sight** of this grandiloquent quartet **lolling** in pretty Irish settings is a **pleasant** enough **thing** P: the **sight** of this grandiloquent Shet **lolling** in pretty Irish American is a lot enough **thing** B: the real of this this asquent Shet filmolling's grand much American is a talented enough *film*

Table 6: Case Analysis for EEG2text. T, P, B denote the target sentence, UniCoRN predictions, and baseline predictions, respectively. Text fragments in **bold** represent the compared portions in the target sentence. **Bold** highlights or underlines indicate representative matches/mismatches, while *italicization* signifies semantic similarity.

decoding capabilities of BART, which effectively bridges the gap between frames that UniCoRN did not encounter during training.

The above results demonstrate an intrinsic characteristic when interpreting the fMRI2text task as a translation-like endeavor. The fMRI time series of different subjects can be likened to the unique accent or speaking style that each individual possesses. While variations among individuals exist, they usually do not present significant challenges in discerning the overall meaning, especially when contextual information is provided. This analogy extends to the case of *random time* and *consecutive time*: when a non-native speaker attempts to comprehend a foreign language, the chances of comprehending key information increase significantly when interpretation can be made from a broader context, as opposed to deciphering a sentence without any foresight of what follows.

Effect of Series Length $\mathcal{T}$ To further demonstrate the effectiveness of decoding fMRI by series, we conduct experiments on different series length $\mathcal{T}$ under *random time* split setting. As shown in Table 4, the length of fMRI series does have a major impact on decoding results when $\mathcal{T}$ is relatively small. However, this impact seems to reach a plateau and might even turn adverse as $\mathcal{T}$ increases. Such trend could be attributed to the inherent limitations of the transformer model in effectively learning long-term dependencies.

Meanwhile, although decoding results tend to be less optimal when $\mathcal{T}$ is small, experiments indicate that apart from frequently used phrases (such as catchphrases during pauses), UniCoRN can still decode semantically and syntactically similar tokens. This capability aligns with previous studies, affirming the feasibility of bridging fMRI and human language under naturalistic settings.

5.4 UniCoRN Structure for EEG2text

As shown in Table 7, the UniCoRN structure surpasses the former baseline on all metrics except when solely using snapshot reconstruction, which will be further discussed in Section 6.

Method	BLEU-N (%)				ROUGE-1 (%)		
	BLEU-1	BLEU-2	BLEU-3	BLEU-4	F	P	R
UniCoRN	57.68	47.93	41.73	**37.04**	**64.39**	60.37	**70.00**
w/o p1	**59.63**	**48.90**	**41.87**	36.51	62.40	59.92	66.25
w/o p2	48.51	37.15	30.25	25.28	52.49	47.48	60.94
w/o p1p2	57.78	46.40	39.10	33.69	62.42	**61.01**	64.44
baseline	54.02	44.93	39.09	34.65	58.78	52.75	67.87

Table 7: Results of EEG2text ablation study. *p1* and *p2* stands for *Phase 1* and *Phase 2* respectively.

Method	BLEU-N (%)				ROUGE-1 (%)		
	BLEU-1	BLEU-2	BLEU-3	BLEU-4	F	P	R
UniCoRN	**62.90**	**49.00**	**40.59**	**34.77**	**59.52**	**62.65**	**56.91**
w/o p1	60.74	46.02	37.27	31.25	57.41	60.69	54.69
w/o p2	61.91	47.36	38.66	32.66	58.33	61.36	55.78
w/o p1p2	53.58	35.53	25.78	19.68	48.75	53.39	45.08

Table 8: Results of fMRI2text ablation study. *p1* and *p2* stands for *Phase 1* and *Phase 2* respectively.

Here we take a closer look at the performance of UniCoRN on EEG2text in Table 6 and compare with the former baseline in Wang and Ji (2022). The results illustrate that UniCoRN outperforms the previous baseline in terms of capturing semantics and syntax in target tokens. Specifically, UniCoRN not only enhances the decoding accuracy of individual words but also maintains superior coherence in sentence structure, resulting in more fluent and comprehensible decoding outputs

6 Ablation Study

To further validate the effectiveness of UniCoRN, we conduct ablation studies on both fMRI2text and EEG2text, to assess how the two phases of signal reconstruction affect the model's performance.

As shown in Table 8, fMRI2text greatly benefits from both phases of fMRI reconstruction, resulting in an improvement of the BLEU score by approximately 20% when reconstruction is included. This indicates that for cognitive signals that are rich in spatial information like fMRI, it is important for the encoder to have a thorough understanding of these signals themselves, but not mainly rely on the ability of decoder. Comparatively, series reconstruction proves to be slightly more effective than snapshot reconstruction, which may be attributed to the nature of seq2seq tasks as the input of series reconstruction is more similar to that of cog2text decoding than snapshot reconstruction.

Conversely, Table 7 shows a decline in overall metrics when only *Phase 1* is used for EEG2text. This could be attributed to the noise introduced by the snapshot reconstruction, which might potentially compromise the ability of the model to process EEG sequences — a crucial aspect for cognitive signals with high temporal resolution like EEG. However, this doesn't undermine the importance of snapshot reconstruction for such signals. As evident in the results, combining snapshot and series reconstruction increases the BLEU-4 score from 36.51% to 37.04%, suggesting an enhancement in the model's performance for predicting longer n-grams. Thus, while the impact may vary depending on the spatial and temporal resolution of different cognitive signals, integrating both phases generally enhances the model's overall performance by developing a more sophisticated encoder.

7 Conclusion

In this paper, we introduce a novel open-vocabulary brain decoding task fMRI2text, aiming to decode linguistic stimuli from multiple fMRI frames collected under naturalistic conditions. Building upon this, we present UniCoRN, a two-stage framework that integrates both temporal and spatial aspects of cognitive signals through snapshot and series reconstruction. The efficacy of UniCoRN is validated under various split settings, illuminating the opportunities that this task provides. Furthermore, we adapt the framework to EEG2text, demonstrating its capacity to generate semantically and syntactically more accurate results, thereby introducing a fresh perspective to brain decoding tasks.

Limitation

The "Narratives" dataset provides a valuable fMRI resource, stimulated by language and obtained under naturalistic conditions. Further research opportunities can be pursued with the availability of more detailed datasets. For instance, comparative studies between instances of stuttering and non-stuttering in text stimuli can be conducted, as our experiments demonstrate that the model tends to retain frequently-used filler words (such as "um" and "like,") as a shortcut for higher accuracy. Meanwhile, the evaluation strategy applied for current research of open-vocabulary brain decoding presents an idealized condition and and serves as a starting point from which further exploration of how existing methods might perform under more real-world scenarios can commence. Although we use this setting for baseline comparison purposes and a testament to the feasibility of our fMRI2text task, additional tests under more practical conditions could be an essential step in future work, further elucidating the applicability and robustness of the methods. Furthermore, the structure of the snapshot encoder can be explored further, as exemplified by the use of transformer-based Vision Transformer (ViT) in Chen et al. (2022) for fMRI encoding.

Ethical Considerations

In this work, we introduce a new NLP task related to fMRI and a unified approach for decoding various types of cognitive signals into human language. We conduct our experiments on the public cognition datasets *Narratives* and *ZuCo1.0* with the authorization from the respective maintainers of the datasets. All experimental datasets involved have been de-identified by dataset providers and used for research only.

Acknowledgements

We express our sincere gratitude to the anonymous reviewers for their professional, insightful and constructive comments and gratefully acknowledge the support of the National Key R&D Program of China [2021ZD0113302]; and the National Natural Science Foundation of China [62206079]; and Heilongjiang Provincial Natural Science Foundation of China [YQ2022F006].

References

Nicolas Affolter, Beni Egressy, Damian Pascual, and Roger Wattenhofer. 2020. Brain2word: decoding brain activity for language generation. *arXiv preprint arXiv:2009.04765*.

Hassan Akbari, Bahar Khalighinejad, Jose L Herrero, Ashesh D Mehta, and Nima Mesgarani. 2019. Towards reconstructing intelligible speech from the human auditory cortex. *Scientific reports*, 9(1):874.

Shohini Bhattasali, Murielle Fabre, Wen-Ming Luh, Hazem Al Saied, Mathieu Constant, Christophe Pallier, Jonathan R Brennan, R Nathan Spreng, and John Hale. 2019. Localising memory retrieval and syntactic composition: an fmri study of naturalistic language comprehension. *Language, Cognition and Neuroscience*, 34(4):491–510.

Jeffrey R Binder, Rutvik H Desai, William W Graves, and Lisa L Conant. 2009. Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral cortex*, 19(12):2767–2796.

Zijiao Chen, Jiaxin Qing, Tiange Xiang, Wan Lin Yue, and Juan Helen Zhou. 2022. Seeing beyond the brain: Conditional diffusion model with sparse masked modeling for vision decoding. *arXiv preprint arXiv:2211.06956*.

Alexandre Défossez, Charlotte Caucheteux, Jérémy Rapin, Ori Kabeli, and Jean-Rémi King. 2022. Decoding speech from non-invasive brain recordings. *arXiv preprint arXiv:2208.12266*.

Bing Du, Xiaomu Cheng, Yiping Duan, and Huansheng Ning. 2022. fmri brain decoding and its applications in brain–computer interface: A survey. *Brain Sciences*, 12(2):228.

Liberty S Hamilton and Alexander G Huth. 2020. The revolution will not be controlled: natural stimuli in speech neuroscience. *Language, cognition and neuroscience*, 35(5):573–582.

Nora Hollenstein, Jonathan Rotsztejn, Marius Troendle, Andreas Pedroni, Ce Zhang, and Nicolas Langer. 2018. Zuco, a simultaneous eeg and eye-tracking resource for natural sentence reading. *Scientific data*, 5(1):1–13.

Alexander G Huth, Wendy A De Heer, Thomas L Griffiths, Frédéric E Theunissen, and Jack L Gallant. 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600):453–458.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

Xiaodong Liu, Kevin Duh, Liyuan Liu, and Jianfeng Gao. 2020. Very deep transformers for neural machine translation. *arXiv preprint arXiv:2008.07772*.

Itzik Malkiel, Gony Rosenman, Lior Wolf, and Talma Hendler. 2021. Pre-training and fine-tuning transformers for fmri prediction tasks. *arXiv preprint arXiv:2112.05761*.

Radek Martinek, Martina Ladrova, Michaela Sidikova, Rene Jaros, Khosrow Behbehani, Radana Kahankova, and Aleksandra Kawala-Sterniuk. 2021. Advanced bioelectrical signal processing methods: Past, present and future approach—part ii: Brain signals. *Sensors*, 21(19):6343.

Sean L Metzger, Jessie R Liu, David A Moses, Maximilian E Dougherty, Margaret P Seaton, Kaylo T Littlejohn, Josh Chartier, Gopala K Anumanchipalli, Adelyn Tu-Chan, Karunesh Ganguly, et al. 2022. Generalizable spelling using a speech neuroprosthesis in an individual with severe limb and vocal paralysis. *Nature Communications*, 13(1):1–15.

Tom M Mitchell, Svetlana V Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L Malave, Robert A Mason, and Marcel Adam Just. 2008. Predicting human brain activity associated with the meanings of nouns. *science*, 320(5880):1191–1195.

Shiv Kumar Mudgal, Suresh K Sharma, Jitender Chaturvedi, and Anil Sharma. 2020. Brain computer interface advancement in neurosciences: Applications and issues. *Interdisciplinary Neurosurgery*, 20:100694.

Samuel A Nastase, Ariel Goldstein, and Uri Hasson. 2020. Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, 222:117254.

Samuel A Nastase, Yun-Fei Liu, Hanna Hillman, Asieh Zadbood, Liat Hasenfratz, Neggin Keshavarzian, Janice Chen, Christopher J Honey, Yaara Yeshurun, Mor Regev, et al. 2021. The "narratives" fmri dataset for evaluating models of naturalistic language comprehension. *Scientific data*, 8(1):1–22.

Uta Noppeney and Catherine J Price. 2004. An fmri study of syntactic adaptation. *Journal of Cognitive Neuroscience*, 16(4):702–713.

Mark Pagel. 2017. Q&a: What is human language, when did it evolve and why should we care? *BMC biology*, 15(1):1–6.

Mark Palatucci, Dean Pomerleau, Geoffrey E Hinton, and Tom M Mitchell. 2009. Zero-shot learning with semantic output codes. *Advances in neural information processing systems*, 22.

Jerrin Thomas Panachakel and Angarai Ganesan Ramakrishnan. 2021. Decoding covert speech from eeg-a comprehensive review. *Frontiers in Neuroscience*, page 392.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Francisco Pereira, Bin Lou, Brianna Pritchett, Samuel Ritter, Samuel J Gershman, Nancy Kanwisher, Matthew Botvinick, and Evelina Fedorenko. 2018. Toward a universal decoder of linguistic meaning from brain activation. *Nature communications*, 9(1):1–13.

Kevin M Pitt and Aimee Dietz. 2022. Applying implementation science to support active collaboration in noninvasive brain–computer interface development and translation for augmentative and alternative communication. *American Journal of Speech-Language Pathology*, 31(1):515–526.

Ivan Provilkov, Dmitrii Emelianenko, and Elena Voita. 2019. Bpe-dropout: Simple and effective subword regularization. *arXiv preprint arXiv:1910.13267*.

Adrien B Rapeaux and Timothy G Constandinou. 2021. Implantable brain machine interfaces: first-in-human studies, technology challenges and trends. *Current opinion in biotechnology*, 72:102–111.

Sergio Ruiz, Korhan Buyukturkoglu, Mohit Rana, Niels Birbaumer, and Ranganatha Sitaram. 2014. Real-time fmri brain computer interfaces: self-regulation of single brain regions to networks. *Biological psychology*, 95:4–20.

Simanto Saha, Khondaker A Mamun, Khawza Ahmed, Raqibul Mostafa, Ganesh R Naik, Sam Darvishi, Ahsan H Khandoker, and Mathias Baumert. 2021. Progress in brain computer interface: Challenges and opportunities. *Frontiers in Systems Neuroscience*, 15:578875.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642.

Jingyuan Sun, Shaonan Wang, Jiajun Zhang, and Chengqing Zong. 2019. Towards sentence-level brain decoding with distributed representations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7047–7054.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Athena Vouloumanos, Kent A Kiehl, Janet F Werker, and Peter F Liddle. 2001. Detection of sounds in the auditory stream: event-related fmri evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience*, 13(7):994–1005.

Shaonan Wang, Jiajun Zhang, Haiyan Wang, Nan Lin, and Chengqing Zong. 2020. Fine-grained neural decoding with distributed word representations. *Information Sciences*, 507:256–272.

Zhenhailong Wang and Heng Ji. 2022. Open vocabulary electroencephalography-to-text decoding and zero-shot sentiment classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 5350–5358.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pages 38–45.

Jonathan R Wolpaw. 2007. Brain-computer interfaces (bcis) for communication and control. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, pages 1–2.

Shuxian Zou, Shaonan Wang, Jiajun Zhang, and Chengqing Zong. 2021. Towards brain-to-text generation: Neural decoding with pre-trained encoder-decoder models. In *NeurIPS 2021 AI for Science Workshop*.

A Implementation Details

The hyperparameters for the experiments in this paper are shown in Table 9.

Task		**Initial LR**	**Batch Size**	**Epoch**
fMRI2text	p1	1e-3	512	10
	p2	1e-3	256	5
	p3	1e-3	224	10
EEG2text	p1	1e-4	768	30
	p2	5e-4	292	30
	p3	1e-4	16	50

Table 9: Hyperparameters used in this paper.

B Notation Table

The notation for the variables mentioned in this paper is presented in Table 10.

C Details of Dataset

The detailed information of the "Narratives" datasets that are used for fMRI2text experiments in this paper is shown in Table 11.

D Details of UniCoRN for EEG2text

As depicted in Figure 3, the snapshot encoder $\mathcal{E}_r$ begins by partitioning the original EEG signal into smaller patches. Subsequently, a multi-layer transformer encoder is utilized to analyze the connections between these patches. The resulting output of $\mathcal{E}_r$ is then concatenated and transformed into a vector with a dimensionality of 1024, serving as the snapshot embedding. The subsequent steps in the process are analogous to those used in the fMRI2text scenario.

E Case Analysis

In this section, we present several cases from our ablation study in Table 12 for fMRI2text and Table 13 for EEG2text to provide a more comprehensive understanding of the variations in decoding quality and the impact of different phases.

As demonstrated in Table 12, UniCoRN effectively decodes "key information" ranging from verbs (such as "swallowing" and "smiled") to nouns ("chocolate" in this example). Without the series reconstruction in *Phase 2*, the model still demonstrates the ability to decode some nouns, but its performance in predicting verbs is significantly impaired. The performance further deteriorates when the snapshot reconstruction in *Phase 1* is removed, although the model still retains sentence structure that is more similar to the target sentence than the model without *Phase 1* and *Phase 2*.

In contrast, the differences in decoding quality are less pronounced in the case of EEG2text. Although UniCoRN is still able to decode some accurate information such as "Einstein" and "Soviet", it fails to correctly decode "physicist" like other methods, and instead generates "government". This discrepancy could be attributed to the fact that EEG signals are aligned at the word level, making the task of decoding EEG less challenging than fMRI2text and thus not showcasing the superiority of UniCoRN as much. Additionally, it could be attributed to UniCoRN's efficient encoder which allows for better utilization of pre-trained language models, since "government" might be mentioned more frequently in the context of "Soviet" than "physicist".

f_k^{ij}	the k th fMRI frame taken when subject i receives stimuli j
S_k	the subject indexed with k
C_k	the stimuli indexed with k
$\mathcal{F}^{ij}$	the collection of all the fMRI frames acquired when subject i receives stimuli j
$\mathcal{F}_{k\sim\mathcal{T}}^{ij}$	the fMRI time series of length $\mathcal{T}$ starting at the k th frame
$\mathbb{F}_{Tr}$	the collection of the fMRI time series contained in the training set
T_{Tr}^{j}	the collection of the index of the starting frames of the input fMRI time series from stimuli j in the training set
$\mathcal{C}_{Tr}$	the collection of the index of the stimuli in the training set
$\mathcal{S}_{Tr}$	the collection of the index of the subjects in the training set
E_k^i	the snapshot embedding for the k th fMRI frame
E_k^e	the serialized embedding for the k th fMRI frame
$E_{k\sim\mathcal{T}}^i$	the snapshot embeddings for fMRI time series of length $\mathcal{T}$ starting at the k th fMRI frame
$E_{k\sim\mathcal{T}}^e$	the serialized embeddings for fMRI time series of length $\mathcal{T}$ starting at the k th fMRI frame
$E_{k\sim\mathcal{T}}$	the fMRI embeddings for fMRI time series of length $\mathcal{T}$ starting at the k th fMRI frame

Table 10: Notations for the main variables used in this paper.

Stimuli	**Duration**	**TRs**	**Words**	**Subjects**
"Pie Man"	07:02	282	957	82
"Tunnel Under the World"	25:34	1,023	3,435	23
"Lucy"	09:02	362	1,607	16
"Pretty Mouth and Green My Eyes"	11:16	451	1,970	40
"Milky Way"	06:44	270	1,058	53
"Slumlord"	15:03	602	2,715	18
"Reach for the Stars One Smal Step at a Time"	13:45	550	2,629	18
"It's Not the Fall That Gets You"	09:07	365	1,601	56
"Merlin"	14:46	591	2,245	36
"Sherlock"	17:32	702	2,681	36
"The 21st Year"	55:38	2,226	8,267	25
Total	3.1 hours	7,424	29,174	
Total across subjects	5.0 days	228,169	887,924	

Table 11: Details of the "Narratives" dataset used in this paper.

Target Sentence	On his way to seat, while **swallowing** what was left of his **chocolate**, he **smiled** to himself.
UniCoRN	On the way to seat, while **swallowing** what was left hand his **chocolate**, Mr **smiled** to himself.
w/o p1	On his way he get, while <u>the</u> what was left with the <u>lesson</u>, Mr <u>was</u> to be.
w/o p2	What the way to the, while <u>they</u> what was left hand his **chocolate**, he<u>'d</u> to himself.
w/o p1p2	and the heart to the, while <u>she</u> what was saying hand the mother, Mr <u>was</u> to.

Table 12: Case Analysis for fMRI2text ablation study.

Target Sentence	**Abram Joffe**, a **Soviet physicist** who knew Einstein, in an obituary of **Einstein**, wrote...
UniCoRN	*Heram Joff*, a **Soviet** government who wrote Einstein, in an Americanitken of **Einstein**, and, wrote...
w/o p1	*Heram J.* -, a Bachelor **physicist** who has Einstein, in a Academyitken of Win, was, wasH most...
w/o p2	*Heram Jia (*, a grades **physicist** of is Einstein, and an Americanitken of his, and, andB film...
w/o p1p2	*Heram Joff about*, a family **physicist** who was Einstein, in an Americanitken of an, in, NewC...

Table 13: Case Analysis for EEG2text ablation study.

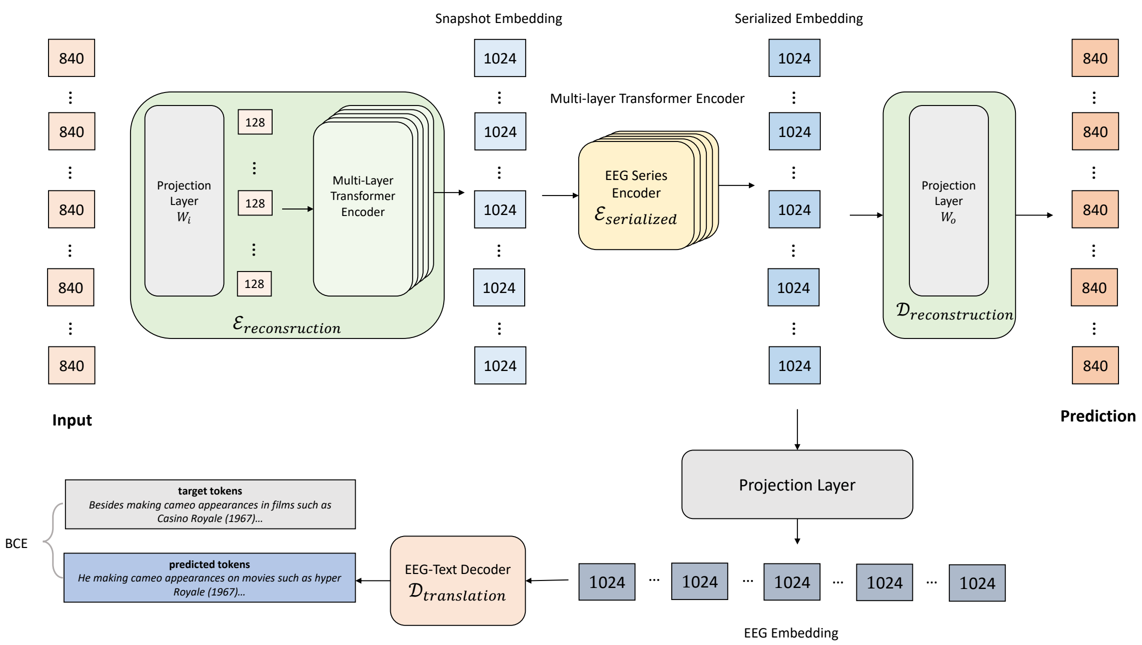

Figure 3: UniCoRN for EEG-To-Text Decoding

ACL 2023 Responsible NLP Checklist

A For every submission:

☑ A1. Did you describe the limitations of your work?
the limitation section

☐ A2. Did you discuss any potential risks of your work?
Not applicable. Left blank.

☑ A3. Do the abstract and introduction summarize the paper's main claims?
the abstrct and section 1

☒ A4. Have you used AI writing assistants when working on this paper?
Left blank.

B ☑ Did you use or create scientific artifacts?

5

☑ B1. Did you cite the creators of artifacts you used?
5

☐ B2. Did you discuss the license or terms for use and / or distribution of any artifacts?
Not applicable. The detailed of the datasets used for the paper is explicitly explained in the original dataset paper, which are cited in section 5

☐ B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
Not applicable. Left blank.

☐ B4. Did you discuss the steps taken to check whether the data that was collected / used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect / anonymize it?
Not applicable. Left blank.

☐ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
Not applicable. Left blank.

☑ B6. Did you report relevant statistics like the number of examples, details of train / test / dev splits, etc. for the data that you used / created? Even for commonly-used benchmark datasets, include the number of examples in train / validation / test splits, as these provide necessary context for a reader to understand experimental results. For example, small differences in accuracy on large test sets may be significant, while on small test sets they may not be.
5

C ☑ Did you run computational experiments?

5

☑ C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
5.2

The Responsible NLP Checklist used at ACL 2023 is adopted from NAACL 2022, with the addition of a question on AI writing assistance.

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Appendix A

☑ C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?
5.2

☑ C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation), did you report the implementation, model, and parameter settings used (e.g., NLTK, Spacy, ROUGE, etc.)?
5.2

D ☒ Did you use human annotators (e.g., crowdworkers) or research with human participants?

Left blank.

☐ D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?
No response.

☐ D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?
No response.

☐ D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating? For example, if you collected data via crowdsourcing, did your instructions to crowdworkers explain how the data would be used?
No response.

☐ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
No response.

☐ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
No response.