Mapping Communication Disruption in Traumatic Brain Injury with Transformer Embeddings

Rick Dale, Zachary P. Rosen
Department of Communication, University of California, Los Angeles
{rdale, z.p.rosen}@ucla.edu

Abstract

Recent advances in computational modeling have expanded our capacity to analyze language and communication, particularly through transformer models. The present work investigates how such computational frameworks can be leveraged to address clinical domains in communication disorders. We used semantic embeddings from BERT's layers to analyze language-related adjustments used by participants with traumatic brain injury (TBI) in conversational transcripts. By examining semantic convergence patterns across different layers of the BERT model, we found that TBI participants demonstrated more pronounced "self" convergence —- they tended to stay closer to their own semantic contributions in the conversation —- compared to controls. This effect was particularly noticeable at earlier layers of the BERT model, suggesting that surface-level semantics play a significant role. The findings highlight the potential for language models to enhance our understanding of social interaction dynamics. We further discuss how bridging computational linguistics with clinical domains can address analytic challenges in the study of natural cognition and communication.

Keywords: natural language processing; deep neural network; semantic analysis; traumatic brain injury; discourse; conversation

Introduction

In this paper, we use variation across layers of a deep neural network (DNN) to investigate how language and communication may be disrupted. Specifically, we analyze transcripts of natural interaction among participants who have experienced traumatic brain injury (TBI). Below we provide some background on the clinical context of TBI to justify our focus: There is considerable debate regarding how language and communication are disrupted in TBI, and the layers of a DNN provide a feature space in which to 'map out' how someone with TBI uses language in a way that may diverge from controls. Our goal is a proof-of-concept study showing that DNNs are a means to map out the disruption to communication in clinical contexts and so can reveal the potential locus of cognitive or linguistic organization that is impacted.

Background

The ability to converse with others is critical to social life: among friends and family, in education, in the workplace and much more (Turkle, 2015). If the ability to converse is disrupted, it can have deleterious effects on quality of life. Adults with social communication disorders can be profoundly impacted (ASHA, 2023). An important goal of the cognitive science of language and communication is to understand what underlies our ability to converse and how this ability can be disrupted. Studying the factors that are affected by injury or some other disruption has wide implications, both in understanding the cognitive basis of language and communication, but also in relevant clinical domains: It could contribute ideas for the development of novel assessments and treatment (Steel & Togher, 2019).

The complexity of language and communication itself can make it difficult to ascertain the underlying impacts of any disruption (Perkins, 2005; Steel & Togher, 2019; Turkstra, Brehm, & Montgomery, 2006; Turkstra & Politis, 2017). In particular, conversation typically involves many behaviors simultaneously. These can include eye movements and body movements, gesture and facial expressions and more (Alviar, Kello, & Dale, 2023; Cassell et al., 1994; Stivers, 2021; Streeck & Jordan, 2009; Hidaka & Yu, 2010; Fusaroli, Bjørndahl, Roepstorff, & Tylén, 2016; Mondada, 2016, 2019). Even in the verbal domain there is considerable need for finely-tuned coordination. Verbal behavior in conversation involves the selection of words, formulating coherent descriptions that are topically relevant, all while tracking the flow of the interaction (Riordan, Kreuz, & Olney, 2014). A disruption to underlying cognitive processes could impact this coordination significantly.

This paper focuses on one domain of such disruption: traumatic brain injury (TBI). TBI is typically defined as an acquired condition elicited by head injury (CDC, 2023; Sarno, Buonaguro, & Levita, 1986; Turkstra & Politis, 2017). It can present with varying degrees of severity (Sarno et al., 1986). Closed-head injury, such as from a fall or sports injury, can cause diffuse brain damage that disrupts circuitry important for complex conversational coordination and language processing. Its impacts have traditionally been studied using standardized language tests (Steel & Togher, 2019). TBI has also been found to impact social communication (Turkstra & Politis, 2017). The subtlety and complexity of TBI and associated language impacts is the subject of continuing investigation. Among the many findings in this area, some have found that TBI may cause disruption to turn-taking, to quantity of information sharing, and to informational entrainment with a conversation partner and more (Coelho, Youse, & Le, 2002; Gordon, Rigon, & Duff, 2015; Turkstra et al., 2006).

The goal of the present paper is to use a well-known cor-

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

```
*INV: so ‡ you_know I read an article the other day in [/] &-uh in the
paper that at time where &-uh hospitals are closing services Milford
keeps adding stuff and expanding a_lot_of xxx . 258594_276781
*PAR: that’s correct . 276781_277397
```

Figure 1: Transcript sample of conversation partners, investigator and participant, from Coelho et al. (1991, 2002) in TBIBank.

pus of TBI to examine how this conversational coordination is impacted across multiple aspects of language. As noted above, this requires an examination of the many aspects of behavior that underlie language and communication. We look to a statistical NLP pipeline intrinsic to language DNNs. These models may be best known for their generative capacities, as they have captivated the public's attention with tools such as GPT and others (Dale, 2021; Rogers, Kovaleva, & Rumshisky, 2020). However any such model has underneath it an array of numeric data representing different aspects of language. In this paper, we utilize BERT, among the first and best-known DNNs for language. BERT is composed of a dozen layers of artificial neural ensembles, known as transformers (Vaswani et al., 2017). These layers of transformers act like an NLP pipeline, and each layer may "pay attention" to different aspects of language, such as specific words, parts of speech, syntax and so on (de Vries, van Cranenburgh, & Nissim, 2020; Rogers et al., 2020).

DNNs like BERT could be considered integrated measurement devices for examining conversational coordination of many aspects of the verbal level. Our aim here is to examine how someone with TBI uses language in a natural communication context. A DNN like BERT could assay how language use with TBI differs from comparison control participants. This approach may link in intriguing ways to psycholinguistic theory too, fostering greater rapprochement between the study of language processing and its disruption. For example, we may find that executive control disruption in TBI aligns with the capacity for other-oriented perspective-taking related to work on language use and comprehension (e.g., Brown-Schmidt, 2009).

In the next section, we summarize the NLP pipeline of BERT, and summarize our analysis approach using it. The prior work on TBI motivates some predictions for our analysis. We then share our results and conclude with theoretical implications.

The NLP Pipeline in BERT

The bidirectional encoder representations from transformers (BERT) learns the context that a word occurs in by predicting it from its backward and forward context. After training on lots of data, such as the entirety of Wikipedia, words can be represented as numeric embeddings in BERT's neural layers. These embeddings can be used to compare words and sentences in order to judge semantic similarity and more. BERT quickly became a standard model by 2020 as it exceeded performance of many models on several benchmarks (Rogers et al., 2020). Despite recent advances in these benchmarks, BERT's layers are a familiar basis on which to conduct our proof-of-concept analysis.

What do BERT's many layers do? There is a long history of studying embedding vectors in statistical models as a way to define and measure word meaning, including in prior models that can capture multiple linguistic features (Johns, Jones, & Mewhort, 2019; Jones & Mewhort, 2007; Landauer & Dumais, 1997; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). As a deep neural network, BERT's capacities in this regard are substantial, encoding not just "word" meaning but many other aspects of the text. BERT approximately "recapitulates" a natural language processing (NLP) pipeline across its 12 layers of transformers. For example, de Vries et al. (2020) showed that BERT, when presented with sequences of words, has layers that successively handle distinct aspects of language, like parts of speech, entity identification, and coreference.

Despite BERT's performance, it has been superseded by more recent transformer DNN models. Nevertheless, it remains widely used for various reasons. First, its embeddings have been widely studied and so are somewhat better understood. Second, it is a smaller and so more tractable model, easy to download directly from well-known machine-learning services to be used locally on a single PC. Third, there has been success using BERT in the past to analyze samples of conversation in a manner similar to what is presented here. For example, by adapting BERT's internal neural layers as a measurement scheme, Rosen and Dale (2023) show that semantic similarity can be detected even in relatively small datasets. We adapt this method for the present analysis.[1]

Convergence-Entropy Analysis

The method described in Rosen (2023) works as follows. First, we convert the entirety of an utterance to word embeddings using a transformer-based DNN model. We then calculate a probability of how likely it is that the word embedding for each token in one sentence could be recovered as a function of the context in a second sentence. This is accomplished by taking the lowest cosine error (*CoE*) for the comparison of the token from the first sentence and any other token in the other sentence.[2] For pairs of compared sentences, *CoE* is then converted to a probability by calculating its likelihood using a half-Gaussian distribution with a location parameter $\mu = 0$, and a pre-selected scale parameter σ. This process is

[1] https://pypi.org/project/convergence-entropy-metric

[2] Note that this analysis is asymmetric. One way of understanding this is that sentences may have different numbers of tokens, making the shorter sentences easier to "recover" from longer sentences.

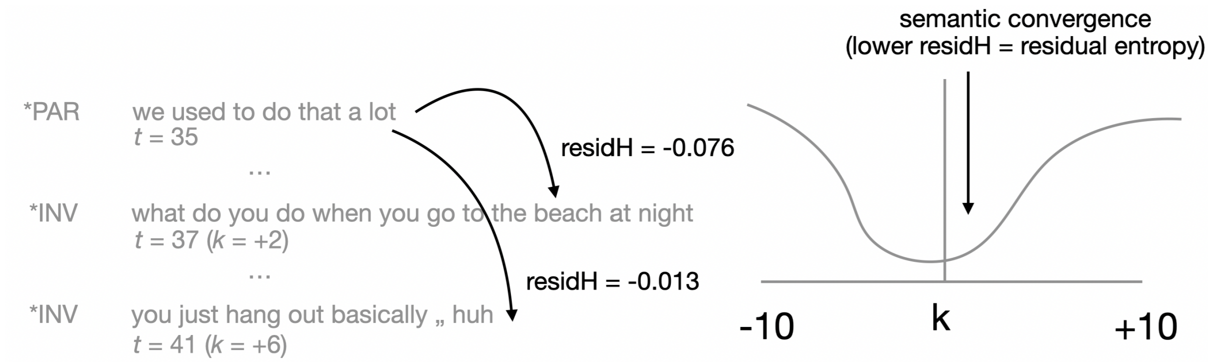

Figure 2: We measured 'residual entropy' ('residH') which is an entropy (or disorder) score that relates utterances – the lower the entropy the more predictable or convergent the comparison. Past work suggests convergence should be strongest near $k = 0$ (proximity of turn time) and example sentences and values shown here.

described in the following equation:

$$P(E_{xi}|E_y) = P_{\mathcal{N}_{[0,\infty]}} \left(\max_j \left(CoE(E_{xi}, E_{yj}) \right) \middle| \mu = 0, \sigma \right) \quad (1)$$

This method then calculates the total entropy – the total degree to which the semantic content, word-by-word, differs between two sentences – with the following:

$$H(x;y) = -\sum_i P(E_{xi}|E_y) \log P(E_{xi}|E_y) \quad (2)$$

When there is evidence of high convergence between two sentences, the "convergence entropy" value returned by Eq. 2 is low. When there is high divergence between two sentences, this value is high.

In this paper, we analyze convergence across many of BERT's layers using a corpus of TBI conversation. This corpus from Coelho et al. (2002) contains conversations with controls and with patients with TBI. In the next section, we summarize the corpus and the analysis approach, offering more technical detail about this convergence method under BERT's layers.

Methods

We used the Coelho Corpus in the TBIBank section of TalkBank (Coelho, Liles, & Duffy, 1991; Coelho, 1995; MacWhinney, 2007). This corpus contains interactions between patients with TBI and an investigator, including a conversational prompt for an informal conversation. The corpus also includes a matched sample of control participants. This informal conversation was extracted from the source transcripts. We analyzed 48 control conversational transcripts and 49 that include patients with TBI. These conversations often involved discussing why participants were at the clinic, their personal lives and work, and so on.[3]

In Fig. 1 we illustrate the structure of a transcript, with a marker for the identity of a speaker for a given turn, and then the sequence of words used in that turn. *INV is the investigator, and *PAR the participant. The numbers at the end of the lines are timestamps. For each such turn, we tracked the identity of the speaker, the line number of the conversation, the particular word sequences used, and removed the timestamps and other fillers using the PyTorch tokenizer. We removed markers for unintelligible ('xxx') and other annotations. We only kept turns that were 6 or more tokens to avoid extreme values in the convergence-entropy estimate.[4] Conversations were about 10 minutes each and the filtered transcripts had between 28 and 175 turns remaining for analysis (mean approx. 100).

We extracted contextual embedding vectors for these transcripts at the world level using BERT.[5] These numeric layers are used as a basis for measuring the convergence between two turns in a conversation. These turns can be analyzed across various points in time. For example, we can take time point t and compare it to the immediate turn at $t+1$. We can also compare that turn to any range within $t \pm k$. For this analysis we chose $k = 10$ as this turn distance would seem to yield diminished convergence. Convergence is therefore a measure of how similar the conversational contributions are across different points in time.

Convergence allows us to assess the extent to which participants are aligning semantically with their conversation partner. By comparing this over time, as described in Turkstra et al. (2006), we can also assess the dynamic shape of this convergence. How quickly does it dissipate across $t+k$ comparisons? We can also compare controls to participants with TBI, and determine if those with TBI diverge or converge in

[3] https://tbi.talkbank.org/access/English/Coelho.html. NB: Some conversational preambles were removed from raw transcripts prior to analysis to ensure that data only included the main conversational prompting.

[4] We acknowledge this may limit access to some structure of conversation, which is often managed in short turns. We make this choice in our initial analysis to stabilize average entropy measurements across turn comparisons, and recognize the need to improve this in future work.

[5] We used `bert-base-cased` from HuggingFace. Because this model was trained with unpublished books and Wikipedia, it is unlikely that the Coelho Corpus was in its training input.

different ways from their conversation partner. By analyzing this convergence across different layers of BERT, we can examine which linguistic properties, such as relative semantic depth, may best characterize divergence from control.

Results

After filtering, the dataset of $k \pm 10$ comparisons yields $748{,}824$ pairwise semantic comparisons for control participants and $742{,}531$ for TBI participants across all 12 BERT layers. We also visually assessed the distribution of residual entropy H after controlling for Levenshtein (surface string) distance and distributions across all BERT layers are approximately normal.[6]

We first tested whether participants in the TBI conditions exhibit any aggregate difference from controls across the entire dataset. We ran a series of multilevel models with `lmer` in `R` in the following way: a **base** model that specifies only conversational distance k and a random intercept, a **TBI-**condition model that includes an interaction term between k and condition (TBI vs. controls), a 'self vs. other' model that factors in whether the convergence is within or across conversation partners, and then a **TBI * self * BERT layer** model that includes layer depth (1-12). To index 'self,' we added a dichotomous variable called 'self' set to 1 when two turns are compared from the participant themselves, and 0 if the turns are compared across both conversation partners. Importantly, BERT layer numbers were included as factors not continuous variables.

The models were defined in the following way, with a simple intercept random effect structure for individual participants that yielded convergence:

Models, par count, and AIC

base :	$\text{residH} \sim k$	4	-2628043
TBI :	$\text{residH} \sim \text{TBI} * k$	6	-2628041
TBI + self :	$\text{residH} \sim \text{TBI} * k * \text{self}$	10	-2631419
full model :	$\text{residH} \sim \text{TBI} * k * \text{self} * L$	98	-3907251

Here, 'residH' is a dependent variable that reflects average entropy rate with surface string similarity between the turns residualized out (Levenshtein distance). To test the contributions of condition, distance, layers and their interactions, we compared the models with `anova` and assessed AIC and statistical significance. We compared each model to its more complex variant. The model with full complexity, containing a factor representing which BERT layer the convergence was measured with, results in the lowest AIC. This full model yields a very large array of significant coefficients. We summarize a few observations from the model. First, the dominant main effects are the factors from the BERT model itself (all p's $< .00001$ for layers 2-12 relative to input layer). Numerous interaction effects were observed in this model, especially between layers and the 'self' and 'TBI' variables (most two-way interactions between layer and these two variables significant p's $< .001$). This suggests, as we predicted, that TBI may represent a particular disruption measurable from the structure in embeddings. We resolve these interactions further below and show that participants in the TBI conditions are showing more pronounced 'self' convergence (they tend to stay close to the semantics contributed on their own to the conversations relative to controls).

[6]For materials, see: https://github.com/racdale/tbi-dnn

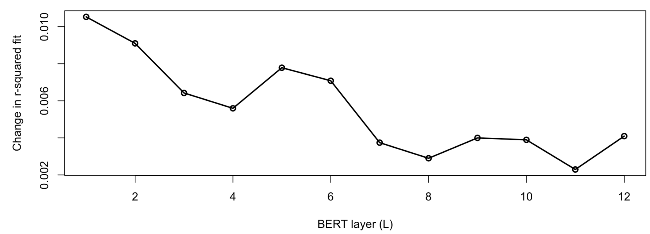

Figure 3: L is a BERT layer model comparing the base (k only) to the full model with k, 'self' and 'TBI' predictors. The change to model fit is assessed by correlating predicted and observed semantic convergence scores and comparing their squared values between base and full models.

Importantly though, TBI does not on its own contribute to the linear model. As shown above, the AIC does not drop when adding TBI to the model that includes the pairwise turn-distance variable k. Indeed, we can get a sense of overall fit by correlating predicted convergence entropy with observed entropy for each model and calculating r^2. Distance of comparison k and TBI only account for less than 1% of the observed variance, and so adding TBI to the linear model accounts for only a fraction of a percent. Adding BERT's layers to the model yields a much higher fit, over 50% of variance accounted for ($r^2 = 0.58$). Of course, this model is much more complex, but its relative AIC is considerably lower than all simpler models. BERT's layers and their interaction with other key variables account for much more of the variance observed in convergence-entropy, suggesting that the embeddings at various layers deviate from one another in how they underlie pairwise semantic comparisons.

Analysis by BERT Layers

To investigate how BERT's layers are contributing to this aggregate model, we ran separate multilevel models over each layer. We simply reduced our overall data to smaller subsets of data that included only semantic convergence measure at each layer and across all 12 layers. The results are shown in Fig. 3.

All layers show significant departure from the base model though across a range of effect sizes. The biggest effect is observed in earlier than later layers, suggesting that surface semantics may be key to understanding these effects. This seems to be a consistent result across the BERT layers, where the strongest contribution of TBI and 'self' comes from the earliest embeddings.

Further visualization of our convergence results suggests that 'self' is an important variable for understanding patterns of significance. When we examine these effects across k in plots shown in Fig. 4, we find that TBI participants show more convergence with themselves especially at earlier layers. In other words, they tend to stay 'closer to their semantic space' than controls. This may explain the significant contribution of the 'self' variable including its widespread interactions with layer. We revisit this in discussion below.

Individual Differences

Consistent with results shown in the prior section, the effects of TBI are very small. Weak aggregate effects may be a sign of underlying variation or complexity at the individual level. For example, Gordon and colleagues showed that while TBI participants diverge in conversational word count, the effects are weaker relative to the wide individual dispersion across participants (Gordon et al., 2015). We suspected this may explain the weak effects here, and results may include wide individual differences across TBI participants.

To examine this, we ran similar multilevel models but included the full set of control participants and each TBI participant individually. Such a model's coefficients would measure potential departure of the individual TBI participant from the mass distribution of the controls. After running all of these models across all participants, we indeed find quite a large range of variation across participants, shown in Fig. 5. Consistent with Gordon et al. (2015), TBI participants show dispersion across the range of possibilities in that they reveal both convergence and divergence across BERT layers. We expand upon this observation in our discussion below.

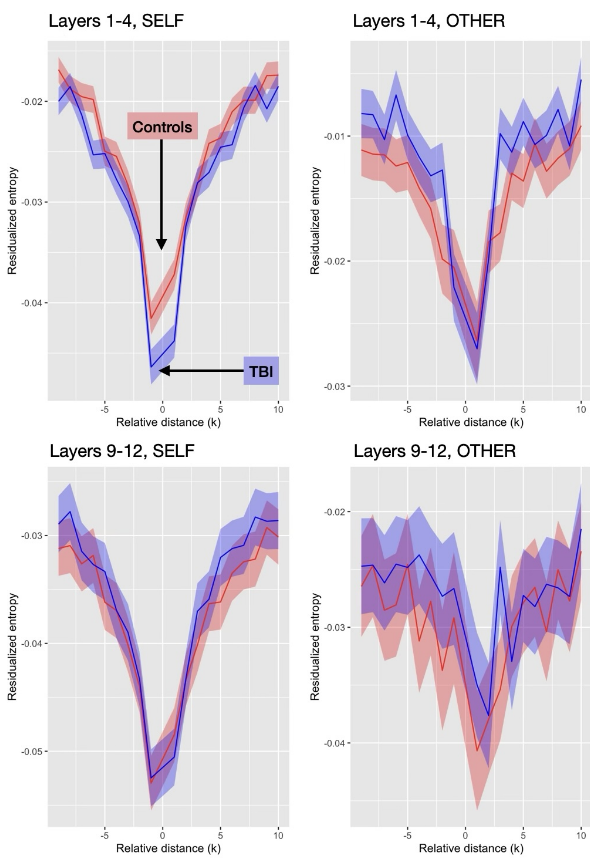

Figure 4: We show entropy convergence (semantic closeness) scores across two sets of earlier and later layers separated by 'self' turn pairs and other. Earlier BERT layers reveal more 'self converence' in TBI than later layers.

General Discussion

Overall, our analysis suggests that embeddings do vary in informative and potentially interesting ways across TBI and controls. However the results are complex and numerous subtle observations frame the foregoing analysis. For example, perhaps curiously, we found that TBI participants converged more with their own contributions to conversation than their partner's. While significant on average, underlying this effect is a wide spread of individual variation, including many TBI participants who appear to show divergence relative to the partner.

There may be a potential distinction here directly relevant to individual differences in clinical contexts. Participants who suffer from the cognitive effects of closed-head injury may have to compensate in context for a fast-changing interaction. There are at least two approaches to this problem. For one, some TBI participants may prefer to converge with their own memory of the conversation, consistent with findings on a general egocentric tendency in social cognition (Epley, Keysar, Van Boven, & Gilovich, 2004). In this case, they may focus on and follow the dominant signal of their own language use. However the opposite may be true for another compensation strategy: TBI participants may compensate for the disruption to their communicative skills by anchoring to their conversation partner. If they struggle with executive control, as suggested in past clinical surveys (Turkstra, Norman, Mutlu, & Duff, 2018), it may help to be guided by and track their interaction partner to the extent they can. This may yield a bimodal distribution, perhaps similar to that observed in the entrainment effects of Gordon et al. (2015).

While this individual level analysis is more speculative, we can suggest that the semantic convergence with BERT yields a wide range of individual patterns across TBI participants, consistent with past work. These wide individual differences could reflect distinctive compensation in response to their injury. Future investigation may couple the transformer-based data mining with targeted content analysis to assess this.

A number of important limitations should be noted. First, we used BERT and there are numerous advances in this semantic modeling since its publication. More recent models may help to fine-tune semantic results, including different training regimes such as next-token prediction and relevant embeddings under those frameworks. Extensions of this work should not rely on a single model but test several architectures.

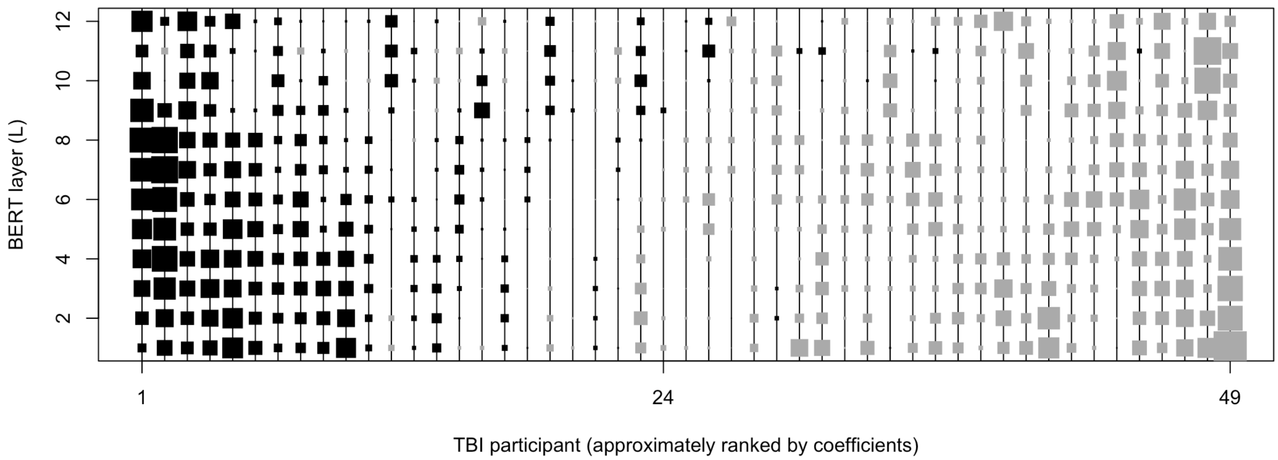

Figure 5: We ran a regession model estimating a coefficient of a TBI participant's status compared to the mass distribution of all controls. These models restricted data to 'self' comparisons nearby with $|k| < 4$. This assesses the local 'self' effect we see in Fig. 4. There is a large tendency for TBI participants to show lower entropy convergence to themselves (dark) but many who show the opposite result (light). The size of the markers is a linear function of the coefficient estimate for plotting.

Another limitation is that we examined a particular sentence length ($n > 5$) and time lag ($k \pm 10$) and future work could investigate how these effects are influenced by such parameters. It may also be helpful to apply hierarchical Bayesian modeling to assist with convergence issues for our inferential models. These would help to improve nested individual-level structures, especially to help gain further insight into individual differences. Finally, as noted, content analysis and more investigation into the layer embeddings would help to understand a finer-grained linguistic locus for these effects – for example, TBIBank encodes a morphosyntactic tier that may be integrated into our analysis to assess structural alignments too.

The approach we frame here, and the potential for future results that examine TBI and other contexts of communication disorders, may represent a fruitful bridging. This bridging is one that could be fostered more, especially in the tools for dynamic analysis and computational linguistics offered by cognitive science. Turkstra et al. (2006) argue for something like this in a discussion of TBI – that more natural communication contexts, especially dynamic and temporally extended ones, may be examined with recent analytic advances. These techniques can robustly integrate large amounts of transcript data while preserving conversational flow. Future such analysis may expand our understanding of language, communication, and their cognitive underpinnings – and how they can be disrupted.

Acknowledgments

Supported in part by NSF grant BCS-2120932 to RD. Research also supported by computational resources made possible by the Mark Allen Itkin Centennial Fund in Communication Honoring Andrea L. Rich.

References

Alviar, C., Kello, C. T., & Dale, R. (2023). Multimodal coordination and pragmatic modes in conversation. *Language Sciences*, *97*, 101524. doi: 10.1016/j.langsci.2022.101524

ASHA. (2023). *Social Communication Disorders: Overview.*

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, *16*(5), 893–900. doi: 10.3758/PBR.16.5.893

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., ... Stone, M. (1994). Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques - SIGGRAPH '94* (pp. 413–420). ACM Press. doi: 10.1145/192161.192272

CDC. (2023). *TBI: Get the Facts | Concussion | Traumatic Brain Injury | CDC Injury Center.*

Coelho, C. A. (1995). Discourse production deficits following traumatic brain injury: A critical review of the recent literature. *Aphasiology*, *9*(5), 409–429.

Coelho, C. A., Liles, B. Z., & Duffy, R. J. (1991). Analysis of conversational discourse in head-injured adults. *Journal of Head Trauma Rehabilitation*, *6*(2), 92–98. doi: 10.1097/00001199-199106000-00011

Coelho, C. A., Youse, K. M., & Le, K. N. (2002). Conversational discourse in closed-head-injured and non-brain-injured adults. *Aphasiology*, *16*(4-6), 659–672.

Dale, R. (2021). GPT-3: What's it good for? *Natural Language Engineering*, *27*(1), 113–118. doi: 10.1017/S1351324920000601

de Vries, W., van Cranenburgh, A., & Nissim, M. (2020). What's so special about BERT's layers? A closer look at the NLP pipeline in monolingual and multilingual models. In *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 4339–4350). (arXiv:2004.06499 [cs]) doi: 10.18653/v1/2020.findings-emnlp.389

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*(3), 327.

Fusaroli, R., Bjørndahl, J. S., Roepstorff, A., & Tylén, K. (2016). A heart for interaction: Shared physiological dynamics and behavioral coordination in a collective, creative construction task. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(9), 1297–1310. doi: 10.1037/xhp0000207

Gordon, R. G., Rigon, A., & Duff, M. C. (2015). Conversational synchrony in the communicative interactions of individuals with traumatic brain injury. *Brain Injury*, *29*(11), 1300–1308.

Hidaka, S., & Yu, C. (2010). Analyzing multimodal time series as dynamical systems. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction.*

Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (2019). Using experiential optimization to build lexical representations. *Psychonomic Bulletin & Review*, *26*(1), 103–126. doi: 10.3758/s13423-018-1501-2

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1–37. doi: 10.1037/0033-295X.114.1.1

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. doi: 10.1037/0033-295X.104.2.211

MacWhinney, B. (2007). The TalkBank project. In J. C. Beal, K. P. Corrigan, & H. L. Moisl (Eds.), *Creating and Digitizing Language Corpora* (pp. 163–180). Springer.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26* (pp. 3111–3119). Curran Associates, Inc.

Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, *20*(3), 336–366. doi: 10.1111/josl.1$_1$2177

Mondada, L. (2019). Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics*, *145*, 47–62. doi: 10.1016/j.pragma.2019.01.016

Perkins, M. R. (2005). Clinical pragmatics: An emergentist perspective. *Clinical Linguistics & Phonetics*, *19*(5), 363–366. doi: 10.1080/02699200400027080

Riordan, M. A., Kreuz, R. J., & Olney, A. M. (2014). Alignment Is a Function of Conversational Dynamics. *Journal of Language and Social Psychology*, *33*(5), 465–481. doi: 10.1177/0261927X13512306

Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A Primer in BERTology: What We Know About How BERT Works. *Transactions of the Association for Computational Linguistics*, *8*, 842–866. doi: 10.1162/tacl$_{a0}$0349

Rosen, Z. P. (2023). A BERT's Eye View: A Big Data Framework for Assessing Language Convergence and Accommodation. *Journal of Language and Social Psychology*, *42*(1), 60–81. doi: 10.1177/0261927X221095865

Rosen, Z. P., & Dale, R. (2023). BERTs of a feather: Studying inter- and intra-group communication via information theory and language models. *Behavior Research Methods*. doi: 10.3758/s13428-023-02267-2

Sarno, M. T., Buonaguro, A., & Levita, E. (1986). Characteristics of verbal impairment in closed head injured patients. *Archives of Physical Medicine and Rehabilitation*, *67*(6), 400–405.

Steel, J., & Togher, L. (2019). Social communication assessment after traumatic brain injury: a narrative review of innovations in pragmatic and discourse assessment methods. *Brain Injury*, *33*(1), 48–61. doi: 10.1080/02699052.2018.1531304

Stivers, T. (2021). Is Conversation Built for Two? The Partitioning of Social Interaction. *Research on Language and Social Interaction*, *54*(1), 1–19. doi: 10.1080/08351813.2020.1864158

Streeck, J., & Jordan, J. S. (2009). Communication as a Dynamical Self-Sustaining System: The Importance of Time-Scales and Nested Context. *Communication Theory*, *19*(4), 445–464. doi: 10.1111/j.1468-2885.2009.01351.x

Turkle, S. (2015). *Reclaiming conversation: the power of talk in a digital age*. New York: Penguin press.

Turkstra, L. S., Brehm, S. E., & Montgomery, E. B. (2006). Analysing Conversational Discourse After Traumatic Brain Injury: Isn't It About Time? *Brain Impairment*, *7*(3), 234–245. doi: 10.1375/brim.7.3.234

Turkstra, L. S., Norman, R. S., Mutlu, B., & Duff, M. C. (2018). Impaired theory of mind in adults with traumatic brain injury: A replication and extension of findings. *Neuropsychologia*, *111*, 117–122. doi: 10.1016/j.neuropsychologia.2018.01.016

Turkstra, L. S., & Politis, A. M. (2017). Traumatic Brain Injury. In L. Cummings (Ed.), *Research in Clinical Pragmatics* (Vol. 11, pp. 291–322). Cham: Springer International Publishing. (Series Title: Perspectives in Pragmatics, Philosophy & Psychology) doi: 10.1007/978-3-319-47489-2$_1$2

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). *Attention Is All You Need.* arXiv. (arXiv:1706.03762 [cs])