Toward a Psycholinguistic Signature of Autism: Everyday Narratives as a Window onto Autistic Experience

Yuanze Liu

December 12, 2025

Abstract

Autistic language is often characterized in terms of isolated deficits (e.g., reduced social or pragmatic skill), yet we know relatively little about how autistic people spontaneously narrate their everyday lives. This paper uses large-scale computational text analysis to identify a psycholinguistic "signature" of autism in first-person narratives. We analyze 1,502 short texts written by clinically diagnosed autistic adults ($n_{\text{ASD}} = 216$) and matched neurotypical adults ($n_{\text{NT}} = 1286$), each freely describing whatever was currently on their mind. In Analysis 1, sentence embeddings from a pre-trained transformer model support a strong supervised ASD–NT semantic dimension: a linear discriminant trained on embeddings achieves cross-validated AUC = .91 and yields a continuous axis along which autistic and neurotypical texts are separable. In Analysis 2, we regress this axis and ASD diagnosis on interpretable features derived from lexical norms (sensory strength, concreteness, valence–arousal– dominance) and GPT-based discourse ratings (e.g., complexity, personal focus, positivity). A small set of features—greater personal focus, higher narrative complexity, stronger auditory imagery, and higher concreteness, together with lower dominance, interoceptive focus, and positivity—explain about 19% of the variance in the embedding-based dimension and classify texts with AUC = .83. In Analysis 3, GPT-assisted extraction and clustering of social actors reveals densely connected, largely overlapping networks of social categories in both groups, with autistic narratives more strongly centered on the self, close relations, and healthcare professionals. Together, these findings suggest that autistic language reflects a distinct but richly social way of organizing experience, and illustrate how interpretable NLP tools can illuminate autistic experience without reducing it to opaque "risk signals."

1 Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental condition defined behaviorally by differences in social communication, restricted interests, and repetitive behaviors, and is associated with elevated rates of anxiety and depression in adulthood (Lord et al., 2020; Hollocks et al., 2019). Classic theories have explained these differences in terms of individual deficits, such as impaired theory of mind (Baron-Cohen, 1995), weak central coherence (Frith, 1989), atypical sensory processing (Robertson and Baron-Cohen, 2017), or reduced social motivation (Chevallier et al., 2012), and have sometimes been framed in essentialist terms such as the "extreme male brain" hypothesis (Baron-Cohen, 2002). More recent accounts, however, argue that difficulties often arise from *mismatched* expectations and communicative styles between autistic and non-autistic people, rather than from a one-sided deficit. The "double empathy problem" (Milton, 2012) and its recent elaboration (Bollen and van Grunsven, 2025) emphasize that both groups may struggle to understand one another's perspectives and that autistic social cognition must be interpreted in the context of this mutual misalignment.

Much of the evidence informing these debates comes from structured tasks, standardized questionnaires, and brief clinical interviews. These approaches have been invaluable for diagnosis and theory-building (e.g., Lord et al., 2020; Williams et al., 2008), but they capture only a small slice of autistic experience. We still know relatively little about how autistic and neurotypical adults *spontaneously* describe their everyday lives, what kinds of events and social actors they foreground, and how they position themselves in these narratives. Existing work on language in autism has often focused on narrow features—such as pronoun use, local/global coherence, echolalia, or emotion words—measured on relatively small samples of speech or writing (e.g., Gernsbacher et al., 2016; Williams et al., 2008; Sterponi et al., 2015; Schaeffer et al., 2023). This literature has revealed important differences, but the field lacks an integrated picture of how such features combine into broader patterns of meaning and narrative stance that might speak directly to theories of autistic experience.

In parallel, researchers in psychology and the social sciences have developed a suite of tools for automated text analysis. Early work relied on dictionary-based methods such as LIWC (Pennebaker et al., 2001; Tausczik and Pennebaker, 2010), and more recent frameworks use machine learning to derive construct-relevant features from large corpora (Humphreys and Wang, 2018; Berger et al., 2020; Atari and Henrich, 2023). Advances in natural language processing (NLP) and large language models (LLMs) have further expanded what is possible: pretrained embeddings provide rich semantic representations of text, and LLMs can act as flexible annotators or hypothesis generators (Rathje et al., 2023; Ludwig et al., 2025; Zhou et al., 2024). These methods open up new ways to summarize complex narratives and derive high-level discourse features, but they also raise ethical concerns if they are deployed as opaque "autism detectors" or screening tools. A central methodological challenge is therefore to harness these tools in ways that illuminate autistic experience and public narratives about autism, without reifying deficit-based assumptions or turning linguistic differences into decontextualized "risk signals."

In this paper we use these methods to study how autistic and neurotypical adults write about whatever is currently on their mind. We analyze a corpus of short first-person texts written by clinically diagnosed autistic adults and matched neurotypical adults, all asked to freely describe their ongoing thoughts and experiences (Sarkar, 2025). Our goal is to

characterize the *structure* of autistic versus neurotypical narratives and to relate any group differences to ongoing theoretical debates about autistic social experience and stigma (Milton, 2012; Bollen and van Grunsven, 2025). By working with first-person language, we aim to complement lab-based measures with a broader window onto how autistic people themselves narrate social life.

We organize our investigation into three analyses that proceed from embedding-level representations to interpretable features and social representations. In **Analysis 1**, we ask whether sentence-level text embeddings contain a robust ASD–NT discriminant: can a simple linear classifier trained on embeddings distinguish autistic from neurotypical texts, and does this define a continuous semantic dimension along which narratives vary? Establishing such a dimension would show that everyday language carries graded information about autistic versus neurotypical styles. In **Analysis 2**, we link this embedding-based dimension to interpretable features by regressing it, and ASD diagnosis itself, on a set of lexical norms (e.g., concreteness, sensory strength, valence–arousal– dominance (Lynott et al., 2020; Brysbaert et al., 2014; Warriner et al., 2013; Mohammad, 2025)) and GPT-based discourse ratings that capture global properties such as complexity, personal focus, and positivity. This step asks to what extent the embedding-based ASD–NT contrast can be reconstructed from psychologically meaningful properties of language rather than opaque model dimensions. In **Analysis 3**, we move to more complex social representations, using GPT-assisted extraction and clustering of social actors to construct co-occurrence networks of social categories in autistic and neurotypical narratives. Here we ask whether autistic and neurotypical writers talk about different people, or organize a largely shared cast of social actors in different ways.

Together, these analyses aim to answer three questions with both theoretical and ethical stakes: (1) Do everyday narratives contain a reliable, continuous ASD–NT semantic dimension? (2) To what extent can this dimension, and ASD diagnosis itself, be reconstructed from interpretable properties of language—such as sensory focus, emotional tone, and narrative stance—rather than from black-box embedding features? and (3) How similar or different are the social worlds that autistic and neurotypical adults describe when talking about their lives, and what might these similarities and differences reveal about stigma, social positioning, and mutual misunderstanding? By connecting embedding-based discriminants to concrete linguistic features and social networks, we seek to understand how autistic language reflects a uniquely organized way of narrating social life.

2 Analysis 1: Deriving an ASD–NT semantic dimension from text embeddings

In our first analysis we asked a basic but foundational supervised learning question: do everyday first-person texts written by autistic and neurotypical adults contain enough semantic signal for a simple linear classifier to distinguish between groups? By training a low-capacity supervised model on sentence embeddings, we can test whether ASD and NT narratives are separable in a high-dimensional semantic space and, if so, extract the resulting linear discriminant as a continuous ASD–NT semantic dimension. This supervised embedding-based axis then serves as the backbone for the interpretive analyses that follow.

2.1 Methods

2.1.1 Data and sentence embeddings

For this analysis we used the full set of texts written by autistic and neurotypical participants in the daily narrative field study (total $N = 1502$ texts; $n_{\text{ASD}} = 216$, $n_{\text{NT}} = 1286$), as described in Sarkar (2025). Each text was labeled according to the writer's diagnostic group (ASD = 1, NT = 0). We represented each text with a sentence-level semantic embedding using the `all-MiniLM-L6-v2` model from the `sentence-transformers` library. The model maps each text into a 384-dimensional vector in a shared semantic space. We stacked these vectors into a 1502×384 matrix and standardized each embedding dimension across texts (zero mean, unit variance) before further analyses.

2.1.2 Exploratory PCA of the embedding space

As an unsupervised baseline, we first characterized the overall variance structure of the embedding space using principal component analysis (PCA). We fit PCA to the standardized embeddings and inspected (a) the cumulative proportion of variance explained as a function of the number of principal components (PCs), and (b) a two-dimensional projection of texts onto the first two PCs (PC1 and PC2), colored by diagnostic group.

To quantify how well the leading PCs alone distinguished autistic from neurotypical texts, we used the PC scores as features in simple classifiers. Specifically, we fit logistic regression models predicting ASD (1) vs. NT (0) using either PC1 alone or the pair (PC1, PC2) as predictors. We evaluated each model with 5-fold stratified cross-validation and recorded the mean area under the ROC curve (AUC) across folds. These PC-based models served as unsupervised baselines against which to compare the supervised discriminant.

2.1.3 Linear discriminant analysis (LDA)

We then trained a Linear Discriminant Analysis (LDA) classifier directly on the standardized embeddings to derive a supervised ASD–NT semantic dimension. LDA finds a single linear combination of embedding dimensions (LD1) that maximizes the ratio of between-group to within-group variance in the projected space.

We again used 5-fold stratified cross-validation to evaluate classification performance. In each fold, we fit LDA on 80% of the texts and computed the ROC curve and AUC on the held-out 20%. We averaged the AUC across folds to obtain a robust estimate of performance.

After cross-validation, we refit LDA on the full dataset to obtain final LD1 scores for every text. These scores define a continuous ASD–NT semantic dimension: higher values indicate more NT-like language and lower values more ASD-like language. We visualized the distribution of LD1 scores separately for ASD and NT texts.

2.2 Results

2.2.1 Unsupervised structure is diffuse and only weakly separates groups

The PCA scree plot (Figure 1A) showed a gradual, almost linear increase in cumulative explained variance with the number of components. No small set of principal components

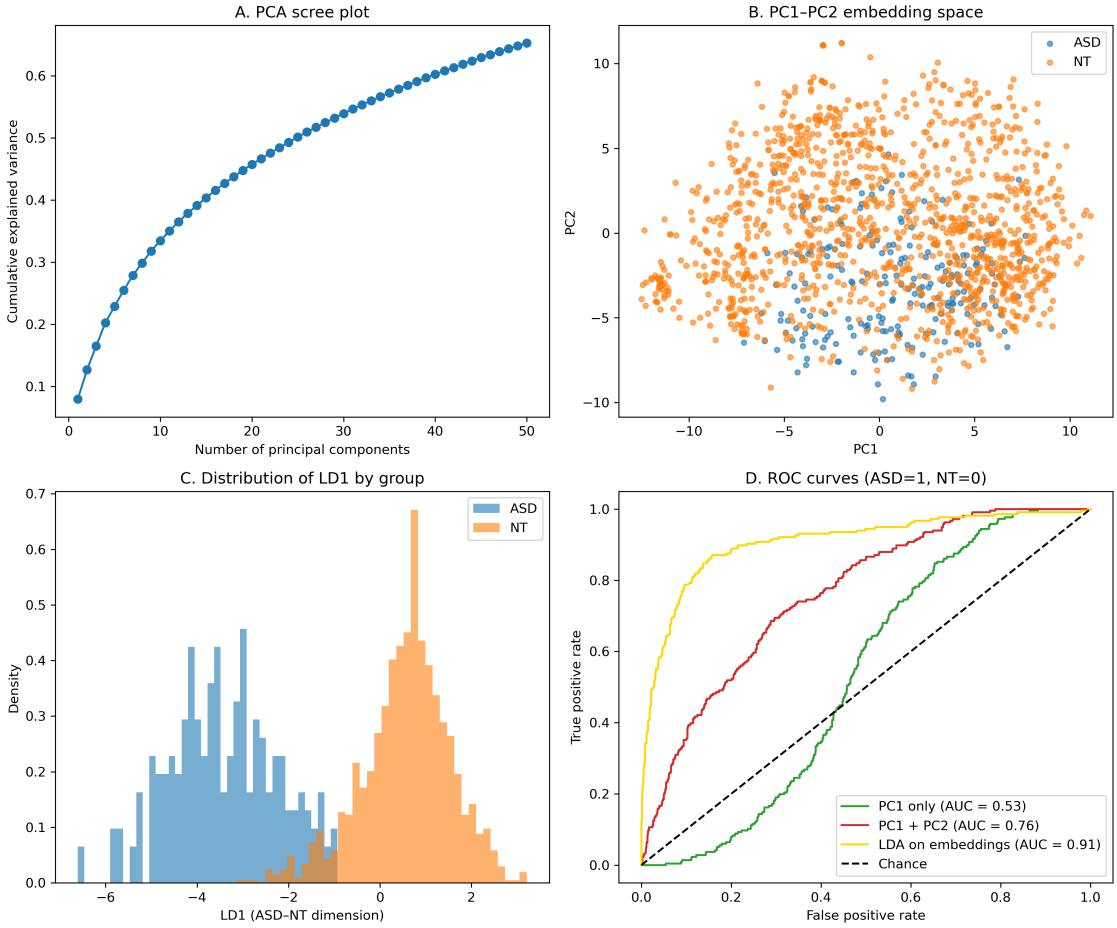

Figure 1: **Analysis 1: PCA and LDA on MiniLM embeddings.** **A**. PCA scree plot showing the cumulative proportion of variance explained as a function of the number of principal components. **B**. Projection of texts onto PC1 and PC2, colored by diagnostic group (ASD vs. NT). **C**. Distribution of LD1 scores (the supervised ASD–NT semantic dimension derived from LDA) for autistic and neurotypical texts. **D**. ROC curves comparing three classifiers: a logistic model using PC1 only, a model using PC1 and PC2, and an LDA classifier trained on the full embedding space.

captured the majority of variance in the embedding space: the first few PCs together explained only a modest fraction of the total variance. This indicates that semantic variation in the essays is high-dimensional and not dominated by a handful of unsupervised directions.

When texts were projected onto PC1 and PC2 (Figure 1B), autistic and neurotypical texts showed only limited visual separation. Points from the two groups overlapped extensively, with only a slight shift in their centroids. Consistent with this impression, a logistic regression using PC1 alone achieved a cross-validated AUC of 0.53, essentially at chance. Using both PC1 and PC2 improved performance to an AUC of 0.76, indicating that the leading unsupervised components do carry some group-related information, but still fall well short of an ideal classifier (Figure 1D).

2.2.2 A supervised discriminant (LD1) captures a strong ASD–NT axis

In contrast, the LDA classifier trained on the full embedding space performed substantially better. Across 5-fold cross-validation, LDA achieved an AUC of 0.91 (Figure 1D), indicating excellent discrimination between autistic and neurotypical texts based solely on their sentence embeddings. The ROC curve for LDA clearly dominated the curves for the PC1-only and PC1+PC2 models.

Refitting LDA on all texts yielded a single discriminant score LD1 for each essay. The distribution of LD1 scores by group (Figure 1C) showed a pronounced shift: autistic texts clustered on one side of the axis and neurotypical texts on the other, with relatively limited overlap. This pattern supports the interpretation of LD1 as a continuous ASD–NT semantic dimension in the embedding space.

Taken together, these results show that while the dominant unsupervised axes of variance (PC1, PC2) provide only limited separation between groups, a supervised linear discriminant trained on the same embeddings captures a strong one-dimensional ASD–NT contrast. In the next analysis, we use interpretable lexical norms and GPT-derived semantic dimensions to unpack what psychological and linguistic properties underlie this embedding-based ASD–NT semantic axis.

3 Analysis 2: Interpretable language features underlying the ASD–NT semantic dimension

The embedding-based discriminant in Analysis 1 showed that autistic and neurotypical texts can be separated along a single semantic axis (LD1) in a high-dimensional sentence-embedding space. In Analysis 2 we ask what this axis represents psychologically and linguistically. Rather than treating LD1 as a black-box direction in the embedding space, we relate it—and ASD diagnosis directly—to a set of interpretable, low-dimensional features derived from lexical norms and GPT-based discourse ratings. This allows us to identify which aspects of emotional tone, sensory language, and narrative style systematically distinguish autistic from neurotypical language.

3.1 Methods

3.1.1 Lexical and norm-based features

To characterize the linguistic content of each text, we constructed a set of interpretable lexical and norm-based features. The underlying norms are defined at the *word* level; we aggregated them to the *text* level by computing the mean rating across all tokens in the text for which norms were available. For every text we computed:

- **Length**: number of whitespace-delimited tokens (`n_tokens`).
- **Sensory strength**: mean modality-specific sensorimotor ratings for words in the text (`sens_auditory`, `sens_gustatory`, `sens_haptic`, `sens_interoceptive`, `sens_olfactory`, `sens_visual`), based on published sensorimotor norms (Lynott et al., 2020).
- **Concreteness**: mean concreteness score across content words (`conc_mean`), using large-scale English concreteness norms (Brysbaert et al., 2014).
- **Valence–arousal–dominance (VAD)**: mean valence (`vad_valence`), arousal (`vad_arousal`), and dominance (`vad_dominance`) ratings based on the NRC VAD Lexicon v2 (Mohammad, 2025).

We merged these text-level features with LD1 scores and group labels at the text level. Because a few texts were missing at least one lexical or norm-based feature, the regressions that included these predictors used the subset of texts with complete data ($n = 1497$).

3.1.2 GPT-based discourse dimensions

In addition to lexical features, we constructed a small set of discourse-level semantic dimensions using large language models. The goal was to obtain interpretable, bipolar scales that summarize how texts differ in style and content.

We first used a maximum-variation sampling strategy (Lu and Lin, 2025) to select a diverse subset of texts: 50 written by autistic participants and 50 by neurotypical participants that were highly different on preliminary lexical and embedding-based features. For each ASD–NT pair in this subset, we prompted a GPT model to compare the two texts and describe the most salient semantic or stylistic differences in free text (for example, which text was more complex, more emotional, more personal, etc.). This procedure yielded a corpus of short natural-language descriptions of cross-group differences.

We then applied BERTopic and hierarchical cluster analysis to these GPT descriptions to group them into a small number of coherent themes. The final solution consisted of eight bipolar dimensions, each defined by two labeled poles and brief verbal descriptions that serve as anchors:

- **Complexity (Complexity vs. Simplicity)**: from detailed, lengthy, and comprehensive narratives to brief, to-the-point statements.
- **Dynamism (Dynamism vs. Static)**: from texts describing activities and change to texts conveying inaction or lack of change.

- **Emotionality (Emotional vs. Practical)**: from focus on feelings and personal experiences to focus on tasks and technical details.
- **Internality (Internal vs. External)**: from introspective texts about personal thoughts and feelings to objective descriptions of external objects and actions.
- **Involvement (Involvement vs. Detachment)**: from highly engaged, reflective narratives to minimally engaged, detached accounts.
- **Personal focus (Personal vs. Impersonal)**: from introspective, self-reflective texts to impersonal texts that lack self-reference.
- **Positivity (Positive vs. Negative)**: from positive, affirming sentiment to criticism and frustration.
- **Temporality (Temporal vs. Timeless)**: from texts that convey a clear sense of time and urgency to texts with no temporal reference.

Using these anchors, we then asked a GPT model to rate each text on each dimension on a 5-point bipolar scale, where higher scores indicated stronger alignment with the first pole of the dimension (e.g., more complex, more personal, more positive) and lower scores indicated stronger alignment with the second pole (e.g., simpler, more impersonal, more negative). Each (text, dimension) rating was obtained from a single GPT call with temperature set to 0, yielding deterministic scores. In the analyses that follow, we refer to these ratings using variable names with a `gpt_` prefix (e.g., `gpt_complexity`, `gpt_personal_focus`) in the code, but we use the shorter labels (Complexity, Personal focus, etc.) when describing the dimensions in the text.

3.1.3 Regression models predicting LD1

To assess how much of the variation in LD1 could be explained by interpretable features, we estimated three multiple regression models with LD1 as a continuous outcome. In all models, predictors were standardized (mean 0, SD 1) so that coefficients are interpretable as the change in LD1 associated with a one-standard- deviation increase in each predictor.

1. **Lexical-only model.** LD1 regressed on the 11 lexical and norm-based features:

$$\text{LD1}_i = \beta_0 + \beta_1\,\text{n\_tokens}_i + \beta_2\,\text{sens\_auditory}_i + \cdots + \beta_{11}\,\text{vad\_dominance}_i + \varepsilon_i.$$

   This model used the $n = 1497$ texts with complete lexical data.

2. **GPT-only model.** LD1 regressed on the eight GPT-based discourse dimensions:

$$\text{LD1}_i = \beta_0 + \beta_1\,\text{gpt\_complexity}_i + \cdots + \beta_8\,\text{gpt\_temporality}_i + \varepsilon_i,$$

   estimated on all $N = 1502$ texts.

3. **Combined model.** LD1 regressed on the full set of 19 predictors (lexical + GPT dimensions), using the $n = 1497$ texts with complete data.

For the combined model, we plotted standardized coefficients and their 95% confidence intervals to visualize which features contributed most strongly to LD1 (Figure 2, panel A).

3.1.4 Logistic models predicting ASD diagnosis

We next asked whether the same interpretable features could directly predict ASD diagnosis (ASD = 1, NT = 0). We fit three logistic regression models with an L2 penalty and class-balanced weights (to account for the smaller ASD group), again using standardized predictors:

1. **Lexical-only model**: ASD vs. NT predicted from the 11 lexical and norm-based features ($n = 1497$).
2. **GPT-only model**: ASD vs. NT predicted from the eight GPT dimensions ($N = 1502$).
3. **Combined model**: ASD vs. NT predicted from all 19 lexical + GPT features ($n = 1497$).

We evaluated each model using 5-fold stratified cross-validation. For every fold, we fit the model on 80% of the data and computed the ROC curve, the area under the curve (AUC), and balanced accuracy on the held-out 20%. We then averaged AUC and balanced accuracy across folds, and plotted ROC curves for the three models (Figure 3). For the combined model, we also visualized standardized logistic coefficients and their 95% confidence intervals (Figure 2, panel B).

3.2 Results

3.2.1 Explaining the embedding-based ASD–NT dimension

Lexical and norm-based features alone explained a modest but non-trivial share of variance in LD1. In the lexical-only regression, the model accounted for $R^2 = 0.144$ of the variance in LD1 (adjusted $R^2 = 0.137$), indicating that text length, sensory language, concreteness, and VAD norms jointly capture some of the systematic differences between more ASD-like and more NT-like language.

The GPT-only model performed similarly, explaining $R^2 = 0.151$ of the variance in LD1 (adjusted $R^2 = 0.147$). This suggests that global discourse dimensions such as complexity, personal focus, and internality are also strongly aligned with the embedding-based ASD–NT semantic axis.

When we combined lexical features and GPT dimensions in a single model, explanatory power increased further: the full model accounted for $R^2 = 0.193$ of the variance in LD1 (adjusted $R^2 = 0.183$). As shown in Figure 2A, a relatively small subset of predictors carried most of the weight. Higher **GPT personal focus** and **GPT complexity**, more tokens, and stronger **auditory imagery** were associated with lower LD1 scores (i.e., more ASD-like language). In contrast, higher ratings on **dominance** and **arousal** (VAD norms), more positive tone, and more temporal references tended to predict higher LD1 scores (more NT-like language). These patterns indicate that the embedding-based ASD–NT dimension is systematically related to a combination of emotional tone, perceived control, sensory language, and narrative style, rather than capturing an opaque or purely statistical contrast.

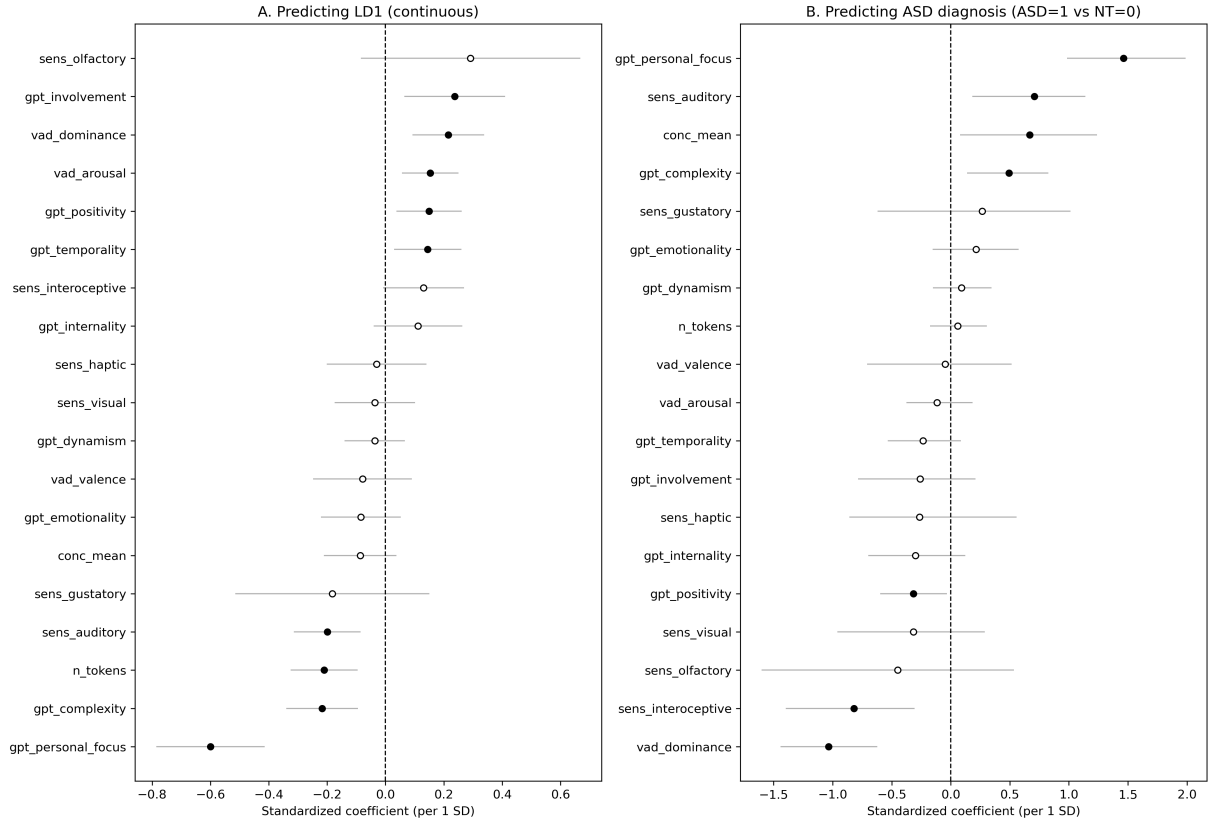

Figure 2: **Analysis 2: Interpretable features underlying the ASD–NT semantic dimension.** **A**. Standardized regression coefficients (with 95% confidence intervals) from the combined linear model predicting LD1 from lexical norms and GPT-based discourse dimensions. Negative coefficients indicate that higher feature values are associated with more ASD-like language (lower LD1). **B**. Standardized logistic regression coefficients (with 95% confidence intervals) from the combined model predicting ASD vs. NT from the same set of features. Positive coefficients indicate that higher feature values increase the log-odds of a text being written by an autistic participant.

3.2.2 Predicting ASD diagnosis from interpretable features

We next examined how well the same features could directly classify texts as autistic vs. neurotypical. The lexical-only logistic model achieved a cross-validated balanced accuracy of 0.725 ± 0.023 and an AUC of 0.795 ± 0.012. The GPT-only model performed similarly, with balanced accuracy 0.716 ± 0.038 and AUC 0.783 ± 0.033.

Combining lexical and GPT features yielded the best performance. The combined model achieved a cross-validated balanced accuracy of 0.769 ± 0.033 and an AUC of 0.834 ± 0.038 (Figure 3), substantially above chance and slightly higher than either feature set alone, although is still somewhat lower than the embedding-based LDA classifier from Analysis 1. The coefficient plot for the combined logistic model (Figure 2B) clarifies which features most

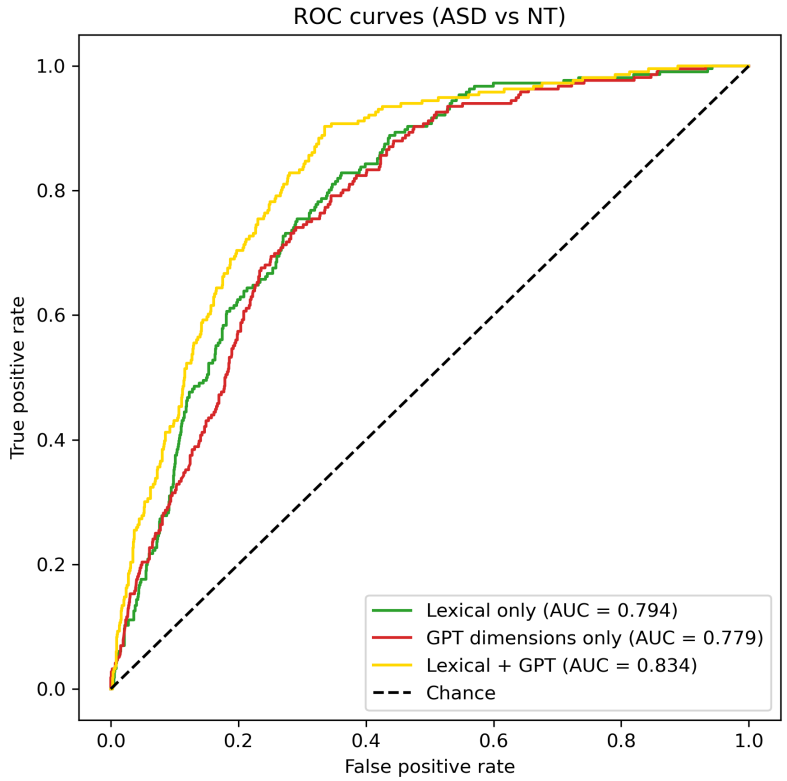

Figure 3: **ROC curves for ASD vs. NT classification using interpretable features.** Lexical-only, GPT-only, and combined (lexical + GPT) models are evaluated with 5-fold stratified cross-validation. The diagonal line indicates chance performance.

strongly distinguish ASD from NT texts when controlling for the others. Higher **GPT personal focus** and **GPT complexity**, stronger **auditory sensory language**, and higher **concreteness** were associated with increased odds that a text was written by an autistic participant. In contrast, higher **dominance**, **interoceptive sensory, and positive language** were associated with *lower* odds of ASD (i.e., more NT-like texts). Other features, such as valence, arousal, and involvement, showed smaller and more uncertain effects.

Taken together, these results show that the ASD–NT semantic dimension identified in Analysis 1 is not merely a black-box artifact of the embedding model. First, approximately

one fifth of the variance in LD1 can be explained by a small set of interpretable lexical norms and GPT-based discourse dimensions, indicating that the discriminant axis is systematically aligned with differences in emotional tone, sensory focus, and narrative perspective. Second, the same features can be used directly to classify texts as autistic vs. neurotypical with substantially above-chance accuracy. Thus, both the existence of a robust embedding-level ASD–NT dimension and its behavioral utility for diagnosis can be partially reconstructed and interpreted in terms of concrete, human-readable properties of autistic and neurotypical language.

4 Analysis 3: Social category networks in ASD and NT narratives

In the final analysis we move from lexical features to the level of social representations. We first used GPT and topic modelling to derive a data-driven set of social categories from the texts themselves, and then asked how autistic and neurotypical participants talk about these categories and how they are organized into co-occurrence networks within each group.

4.1 Methods

4.1.1 Social category coding

We first constructed a set of social categories in a data-driven way. For each text, a GPT model was prompted to identify all social entities mentioned in the narrative (e.g., the self, family members, friends, doctors, coworkers, political figures). We pooled these GPT-extracted social entities across all texts and applied BERTopic to group similar entities into broader categories. We then used GPT to generate concise labels for each cluster (e.g., *Self*, *Family and Close Relations*, *Social Circles*, *Political Authorities*, *Academic Professionals*, *Companion Animals*), yielding a final set of 23 social categories.

Each text was then coded for the presence or absence of each derived social category. A category was treated as present in a text if at least one expression referring to that type of social target appeared anywhere in the text. Coding was applied at the text level separately for autistic (ASD) and neurotypical (NT) participants.

For descriptive analyses we computed, for each group and category, the proportion of texts in which the category was mentioned at least once (Figure 4). These proportions provide a first comparison of which kinds of social targets are more or less salient in ASD versus NT narratives.

4.1.2 Co-occurrence network construction

We then constructed a co-occurrence network for each group. Nodes correspond to the 23 social categories. For each group separately, we added an undirected edge between two categories i and j if they co-occurred in at least one text written by a participant from that group; the edge weight was the number of texts in which the pair co-occurred. For

the network-level metrics reported below, we treated the graphs as unweighted (presence vs. absence of an edge).

For each group-specific network we computed: number of nodes and edges; network density (proportion of possible edges that are present); average degree; average clustering coefficient; number of connected components and average shortest path length within the largest connected component (LCC).

Finally, we computed betweenness centrality for each node within each group-specific network, treating edges as unweighted. Betweenness centrality captures how often a node lies on shortest paths between other nodes and thus indexes how strongly a category acts as a bridge between different parts of the social network. We visualized the networks themselves, as well as group differences in category prevalence and betweenness centrality (Figures 4–6).

4.2 Results

4.2.1 Prevalence and network-level structure

Both groups frequently mentioned a small core of social categories. As shown in Figure 4, *Self* appeared in nearly all texts in both groups, and *Family and Close Relations*, *Social Circles*, and *Romantic Relationships* were mentioned in a substantial minority of texts. Other categories, such as *Healthcare Professionals*, *Companion Animals*, and *Academic Professionals*, appeared less often overall, but were noticeably more common in ASD texts than in NT texts. This pattern is consistent with autistic participants' more frequent contact with clinicians and the fact that they were recruited through healthcare providers and knew their narratives would be used in autism research. More distal or institutional categories (e.g., *Global Entities*, *Survey Stakeholders*, *Media Entities*) were relatively rare in both groups.

The resulting co-occurrence networks were dense and well connected in both groups (Figure 5). The ASD network contained 23 nodes and 136 edges, whereas the NT network contained the same 23 nodes but more edges (163). Correspondingly, the NT network showed higher density and average degree: density was 0.538 for ASD and 0.644 for NT, and average degree increased from 11.83 (ASD) to 14.17 (NT). Both networks formed a single connected component, with high clustering coefficients (0.80 for ASD, 0.81 for NT), indicating that social categories tended to form tightly knit triads. Average shortest path length within the largest component was short in both groups but slightly shorter in the NT network (1.36 vs. 1.46), consistent with the higher density.

Overall, these network-level statistics suggest that autistic and neurotypical participants draw on a largely overlapping set of social categories when narrating their daily experiences. However, the NT network is somewhat more densely interconnected, with more co-occurrences between categories and slightly shorter paths between any two nodes.

4.2.2 Node-level centrality differences

We next examined which specific categories occupy the most central positions in each network. As shown in Figure 6, *Self* had by far the highest betweenness centrality in both groups, reflecting its role as a hub connecting many other social categories. *Family and Close Relations* and *Social Circles* also exhibited high betweenness centrality, indicating that close others and friendship networks sit at the core of both ASD and NT social representations.

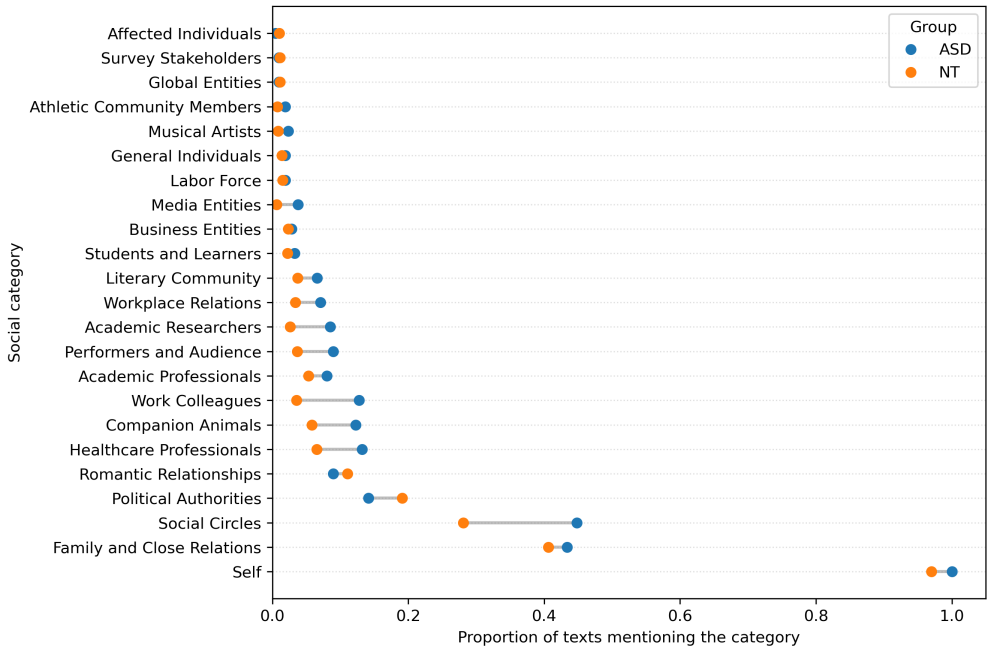

Figure 4: **Prevalence of social categories in ASD and NT texts.** For each category, points indicate the proportion of texts in which the category is mentioned at least once, separately for ASD and NT participants.

Beyond this shared core, there were more subtle differences in how other categories functioned as bridges. In the ASD network, betweenness centrality was somewhat more concentrated on *Self*, *Family and Close Relations*, and *Social Circles*, suggesting that many other categories are connected through the self and immediate social environment. In the NT network, centrality was distributed slightly more evenly.

Taken together, the network analysis indicates that autistic and neurotypical participants embed their experiences in broadly similar social worlds organized around the self, family, and close social circles. At the same time, the denser NT network and its somewhat more institutionally distributed centrality pattern suggest that neurotypical participants more often connect everyday narrative to a wider set of formal roles and societal institutions, whereas autistic participants' networks are comparatively more centered on the self and immediate personal relationships.

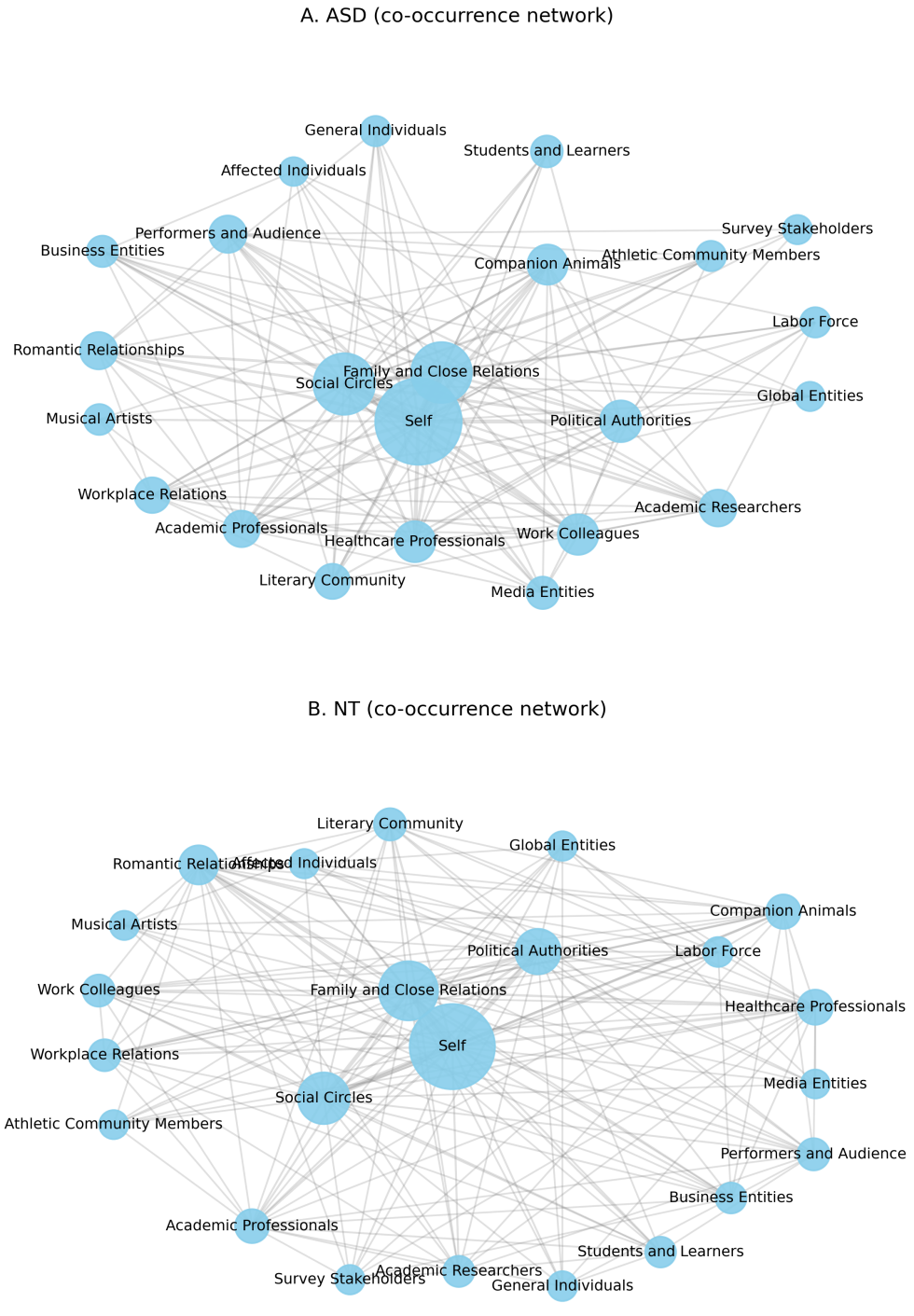

Figure 5: **Co-occurrence networks of social categories in ASD and NT texts.** Nodes represent social categories (node size proportional to degree); edges connect categories that co-occur in at least one text within each group. Panel A: ASD network. Panel B: NT network.

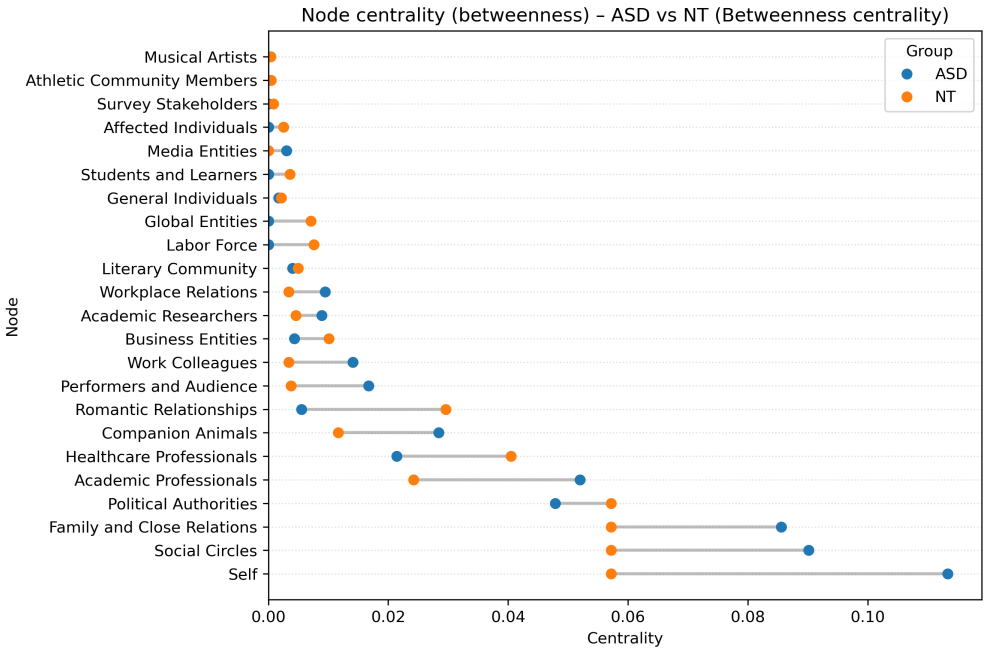

Figure 6: **Betweenness centrality of social categories in ASD and NT networks.** Points indicate node betweenness centrality for each category in the ASD and NT co-occurrence networks.

5 Discussion

Across three analyses, we used large-scale psycholinguistic tools to characterize how autistic and neurotypical adults describe their everyday experiences. Analysis 1 showed that an embedding-based linear discriminant separates ASD and NT texts above chance, implying that first-person narratives contain a reliable ASD–NT semantic dimension. Analysis 2 linked this dimension to interpretable lexical norms and GPT-based discourse ratings and showed that the same features can directly predict ASD diagnosis with substantial accuracy. Analysis 3 moved from word- and discourse-level features to social representations, revealing largely overlapping but differently organized networks of social categories in autistic versus neurotypical narratives. In this section we discuss the implications of these findings for theories of autistic experience, public narratives and stigma, methodological contributions, and future work.

5.1 Implications for theories of autistic experience

Traditional theories of autism have emphasized individual-level social deficits, such as impaired theory of mind, weak central coherence, or reduced social motivation (e.g., Baron-Cohen, 1995; Frith, 1989; Chevallier et al., 2012). More recent accounts (e.g., Milton, 2012; Bollen and van Grunsven, 2025), instead conceptualize autistic difficulties as arising from misaligned expectations and narrative styles between autistic and non-autistic people rather than from a unilateral deficit. Our results speak to both sides of this debate.

First, the existence of a robust ASD–NT semantic dimension in embedding space (Analy-

sis 1) confirms that autistic and neurotypical participants on average write differently about their social lives. The LD1 distributions, however, were still continuous and overlapping to some extent: many autistic narratives were more "NT-like" and vice versa. This graded pattern is more consistent with a spectrum of communicative styles than with a sharp deficit boundary (cf. Lord et al., 2020).

Second, the features that best explained LD1 and predicted ASD diagnosis (Analysis 2) are not simply markers of impoverished language. Autistic texts were more self-focused and personal, higher in narrative complexity, and richer in certain kinds of sensory language, while showing lower dominance and positivity. These patterns suggest an intense, introspective, and sometimes vulnerable stance toward social life rather than a lack of social content per se, consistent with qualitative work on autistic narration and experience (e.g., Williams et al., 2008; Sterponi et al., 2015; Schaeffer et al., 2023). In conjunction with the network results, they support the idea that autistic people inhabit rich social worlds that are structured and narrated differently, not empty or deficient ones.

Third, the social category networks (Analysis 3) revealed substantial overlap between groups: both ASD and NT participants organized their narratives around the self, family, and close social circles, with similar peripheral categories such as romantic partners, colleagues, and institutional actors. The NT network was somewhat denser and distributed centrality more evenly across institutional categories, whereas the ASD network placed especially strong weight on the self and immediate relations and referenced healthcare professionals more often. This pattern aligns with accounts that emphasize differences in social positioning and experiences (e.g., frequent contact with clinicians, heightened salience of close support networks) rather than a fundamental inability to represent social others (e.g., Williams et al., 2008; Robertson and Baron-Cohen, 2017).

Taken together, our findings fit best with theories that treat autistic experience as structurally different but richly social. Autistic and neurotypical participants appear to share many social reference points, yet they organize and narrate these points through partially diverging linguistic and experiential lenses (Milton, 2012; Bollen and van Grunsven, 2025).

5.2 Language, stigma, and public narratives about autism

Because our analyses identify linguistic features that distinguish ASD and NT texts, they could easily be misinterpreted as "objective markers" of deficit or pathology. We therefore see our results as an opportunity to reflect on how language both reflects and shapes public narratives about autism (e.g., Bollen and van Grunsven, 2025).

Some of the features associated with ASD narratives—greater personal focus, lower expressed dominance and positivity, more frequent mention of healthcare professionals—can be read as traces of structural disadvantage: repeated experiences of being assessed, diagnosed, or misunderstood, and the need to explain oneself. If such linguistic signatures are naively treated as individual shortcomings, they risk reifying stigma: autistic people may appear overly self-focused, negative, or "clinical" simply because their lives are more often routed through clinical and evaluative contexts (Milton, 2012; Bollen and van Grunsven, 2025).

At the same time, our results highlight the richness of autistic storytelling. Autistic texts were not shorter, less complex, or devoid of social content; rather, they conveyed social worlds through a different balance of internal reflection, sensory detail, and affective tone.

Recognizing this variety can counter public portrayals of autistic people as asocial or lacking narrative selfhood (Sterponi et al., 2015). Instead, differences in language can be understood as alternative ways of making sense of social life under conditions of marginalization and misunderstanding.

These considerations also underscore the ethical limits of using language-based models for screening or risk prediction. While our combined feature model achieved respectable predictive performance, it is far from perfect and is trained on texts from a specific context. More importantly, any attempt to use such models in clinical or educational settings would need to confront the risk of amplifying existing biases about how "normal" social language should look. We see greater value in using these tools to illuminate patterns of experience and narrative, not to police them (see also Bollen and van Grunsven, 2025; Ludwig et al., 2025).

5.3 Methodological contributions

Methodologically, the project illustrates how recent NLP and LLM tools can be used to study autistic experience in a way that balances prediction with interpretability. Rather than training a large black-box model to classify diagnosis, we first derived a low-dimensional ASD–NT semantic axis from sentence embeddings (Analysis 1), then asked what lexical and discourse-level features explain this axis and how far they go in reproducing its predictive power (Analysis 2). This two-step strategy shows that an embedding-based discriminant can be decomposed into a relatively small set of meaningful features capturing sensory language, emotional tone, and narrative stance, echoing broader calls for interpretable automated text analysis in psychology and the social sciences (Tausczik and Pennebaker, 2010; Humphreys and Wang, 2018; Berger et al., 2020).

Our use of GPT is similarly constrained and interpretable. We used LLMs to construct global discourse dimensions (e.g., Complexity, Personal focus, Positivity) and to extract social actors that were then clustered into social categories (Analysis 3). In both cases, GPT primarily serves as a tool for turning qualitative, high-dimensional judgments into structured variables that can be inspected and related back to theory, rather than as an opaque end-to-end learner. This illustrates a general workflow in which LLMs are used to scaffold human theory-building instead of replacing it (Rathje et al., 2023; Ludwig et al., 2025; Zhou et al., 2024).

More broadly, the combination of embeddings, lexical norms, GPT-based ratings, and network analysis demonstrates how different levels of linguistic representation can be integrated in computational social science (e.g., Atari and Henrich, 2023; Lu and Lin, 2025). Embedding spaces provide sensitive, data-driven discriminants; norm-based features and LLM ratings provide interpretable axes; and social networks summarize how people assemble concrete social categories in narrative. We see this multi-level approach as a template for studying other forms of neurodivergent or marginalized experience.

5.4 Limitations and future directions

Several limitations qualify our conclusions and point to future work. First, our data consist of English-language written narratives collected in a specific research and recruitment context.

Many autistic participants were connected to healthcare providers, and all knew that their texts would be used in a study about autism. This likely shaped which experiences and social actors they chose to highlight (e.g., frequent mention of clinicians). Whether similar linguistic patterns would appear in more naturalistic conversations, different languages, or community samples remains an open question (cf. Lord et al., 2020).

Second, our measures of discourse and social categories rely on GPT-based annotations and clustering choices. Although we used deterministic prompts and interpretable anchors, LLMs can encode biases from their training data and may systematically misinterpret certain styles of autistic writing. Similarly, the choice of embedding model, clustering parameters, and network thresholds all introduce degrees of freedom. Future work could compare alternative annotation pipelines (e.g., human ratings, other LLMs, or multilingual models) and assess the robustness of the inferred dimensions and networks (e.g., Rathje et al., 2023; Ludwig et al., 2025).

Third, our analyses are cross-sectional and descriptive. They show how autistic and neurotypical participants differ on average, but they do not identify within-person trajectories or causal mechanisms. Longitudinal designs could examine how language changes across development, diagnosis, or changes in social environment, and whether shifts in language accompany changes in well-being or social connectedness. Comparative work could also examine how the ASD–NT semantic dimension relates to other forms of neurodivergence or mental health conditions (e.g., Hollocks et al., 2019).

Finally, although we framed our analyses in relation to theoretical debates about autistic social cognition, we focused on summary dimensions rather than fine-grained interactional phenomena. Future studies might combine our approach with conversation analysis, pragmatic coding, or experimental tasks to link global narrative styles to more specific processes such as turn-taking, perspective-taking, or expectation management in mixed autistic–neurotypical interactions (see Sterponi et al., 2015).

6 Conclusion

Using a combination of embeddings, lexical norms, GPT-based discourse ratings, and social network analysis, we identified a psycholinguistic signature of autism in first-person narratives of everyday experiences. Autistic and neurotypical participants wrote in systematically different ways, yet their social worlds were far from disjoint: both groups organized their stories around the self, family, and close social circles, with nuanced differences in stance, affect, and institutional embedding. Our findings support accounts that understand autistic experience as richly social but differently structured, and they illustrate how computational methods can help clarify—and potentially humanize—the linguistic traces of neurodivergent lives. Rather than treating these differences as deficits, we hope this work can contribute to more nuanced, respectful narratives about autism and its place in the social world.

References

Atari, M. and Henrich, J. (2023). Cultural evolution and the measurement of psychological variation. *Current Directions in Psychological Science*, 32(4):290–296.

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press, Cambridge, MA.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences*, 6(6):248–254.

Berger, J., Sherman, G., and Ungar, L. (2020). Textanalyzer: A system for automated text-based measurement of psychological constructs. *Behavior Research Methods*, 52:1–18.

Bollen, C. and van Grunsven, J. (2025). In defense of the double empathy problem hypothesis: An urgently needed alternative to fallacies and stigma. *American Psychologist*. Advance online publication.

Brysbaert, M., Warriner, A. B., and Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3):904–911.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., and Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, 16(4):231–239.

Frith, U. (1989). *Autism: Explaining the Enigma*. Blackwell, Oxford, UK.

Gernsbacher, M. A., Morson, E. M., and Grace, E. J. (2016). Language and speech in autism. *Annual Review of Linguistics*, 2:413–425.

Hollocks, M. J. et al. (2019). Anxiety and depression in adults with autism spectrum disorder: A systematic review. *Psychological Medicine*, 49:559–572.

Humphreys, M. and Wang, T. (2018). Automated text analysis for social science: A review. *Annual Review of Political Science*, 21:415–433.

Lord, C., Elsabbagh, M., Baird, G., and Veenstra-Vanderweele, J. (2020). Autism spectrum disorder. *Nature Reviews Disease Primers*, 6:5.

Lu, J. and Lin, C. (2025). Network models reveal high-dimensional social inferences in naturalistic settings beyond latent construct models. *Communications Psychology*, 3(1):98.

Ludwig, J., Mullainathan, S., and Rambachan, A. (2025). Large language models: An applied econometric framework. Working Paper 33344, National Bureau of Economic Research.

Lynott, D., Connell, L., Brysbaert, M., Brand, J., and Carney, J. (2020). The lancaster sensorimotor norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, 52(3):1271–1291.

Milton, D. E. M. (2012). On the ontological status of autism: The 'double empathy problem'. *Disability & Society*, 27(6):883–887.

Mohammad, S. M. (2025). NRC VAD Lexicon v2: Norms for valence, arousal, and dominance for over 55k English terms. Technical Report arXiv:2503.23547, arXiv. Retrieved from arXiv.

Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). Linguistic inquiry and word count (liwc). *Mahwah, NJ: Lawrence Erlbaum*.

Rathje, S., Mirea, D.-M., Sucholutsky, I., Marjieh, R., Robertson, C., and Van Bavel, J. J. (2023). GPT is an effective tool for multilingual psychological text analysis. *PsyArXiv*. Preprint.

Robertson, C. E. and Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience*, 18:671–684.

Sarkar, A. (2025). A psycholinguistic signature of autism. Unpublished DCF proposal.

Schaeffer, J., Abd El-Raziq, M., Castroviejo, E., Durrleman, S., Ferré, S., Grama, I., Hendriks, P., Kissine, M., Manenti, M., Marinis, T., Meir, N., Novogrodsky, R., Panzeri, F., Silleresi, S., Sukenik, N., Vicente, A., Zebib, R., Prévost, P., and Tuller, L. (2023). Language in autism: Domains, profiles and co-occurring conditions. *Journal of Neural Transmission*, 130(12):1681–1703.

Sterponi, L., de Kirby, K., and Shankey, J. (2015). Rethinking language in autism. *Autism*, 19(5):517–526.

Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54.

Warriner, A. B., Kuperman, V., and Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4):1191–1207.

Williams, D., Botting, N., and Boucher, J. (2008). Language in autism and specific language impairment: Where are the links? *Psychological Bulletin*, 134(6):944–963.

Zhou, Y., Liu, H., Srivastava, T., Mei, H., and Tan, C. (2024). Hypothesis generation with large language models. arXiv preprint arXiv:2404.04326.