Understanding User Behavior and Platform Usage in Jargon: A Chrome Extension for Language Learning

2025-04-30

Introduction

Background and Motivation

Jargon is an innovative Chrome extension (Chrome Web Store (https://chromewebstore.google.com/detail/jargon/gghkanaadhldgmknmgggdgfaonhpppoj), Official Website (https://www.jargonlearn.com/)) created by my friend that transforms English web content into learning opportunities using generative AI technology. Launched in June 2024, Jargon offers two types of learning experiences: foreign language learning (Spanish, Chinese, etc.) and English style adaptation (GRE vocabulary, TikTok slang, etc.).

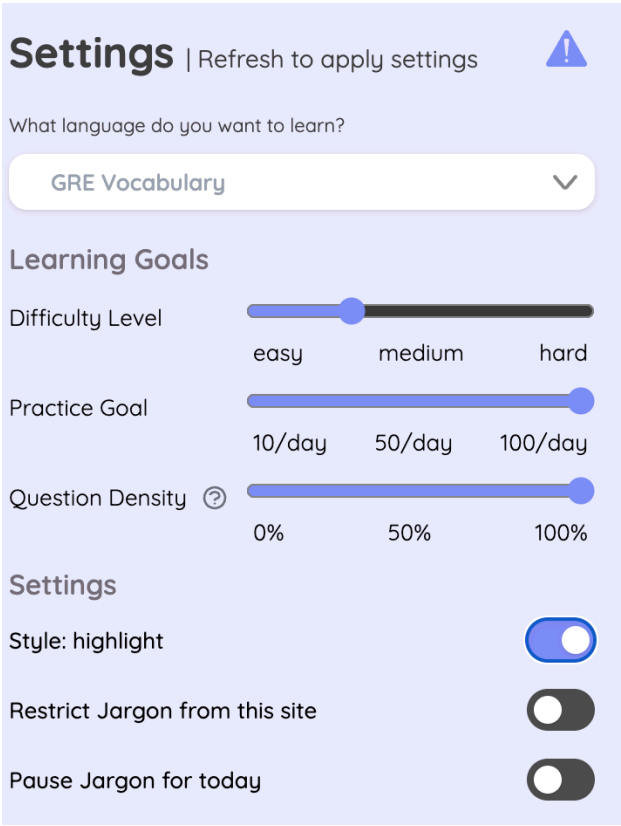## How Jargon Works

Customization Options

Key Features

Language Selection

All types, from foreign languages like Spanish and Chinese to English variations such as TikTok Slang

Learning Goals

• Difficulty: Easy-Hard (1-10)
• Daily Target: 10-100 questions

Question Density

Controls percentage of eligible sentences (0-100%) highlighted for practice on each webpage

Display Settings

• Text Style: Highlight or underline
• Site Controls: Enable/disable per website or temporarily

Figure 1: User Settings Interface showing customization options

Text Selection Methods

especially in Broca's and Wernicke's areas. Humans acquire language through social interaction in early childhood, and children generally speak fluently by approximately three years old. Language and culture are codependent. Therefore, in addition to its strictly communicative uses, language has social uses such as

especially in Broca's and Wernicke's areas. Humans acquire language through social interaction in early childhood, and children generally speak fluently by approximately three years old. Language and culture are codependent. Therefore, in addition to its strictly communicative uses, language has social uses such as

Figure 2a: Highlight Style - Text appears with background color emphasis

Figure 2b: Underline Style - Text appears with underline emphasis

Underline style was the default option initially provided by Jargon. Based on user feedback, the highlight style was introduced, offering an alternative that users can switch to for a better UI experience.
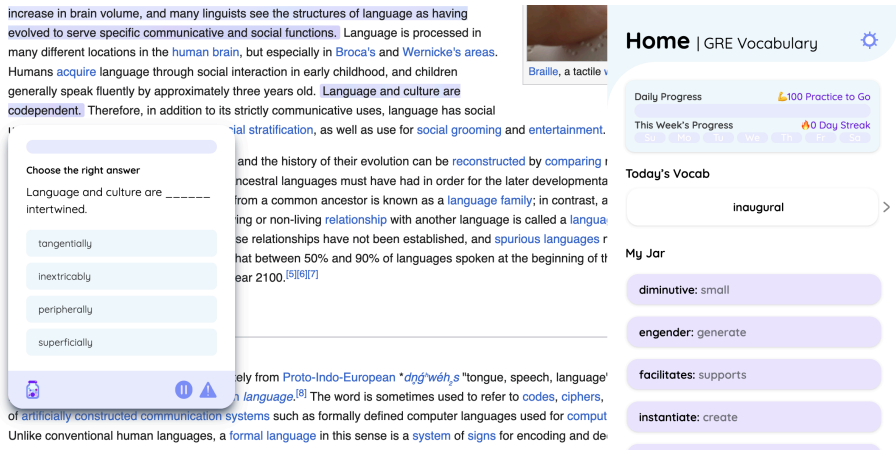
Language Transformation Examples

Figure 3: Question Generation Process - Users select text from any webpage to create practice questions
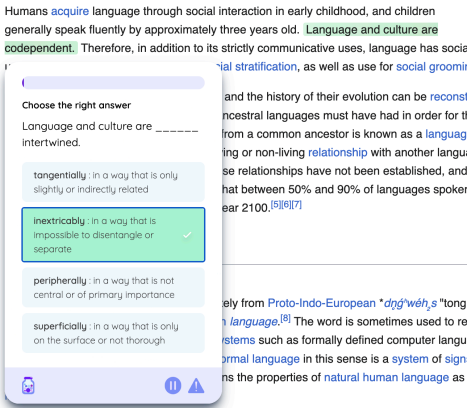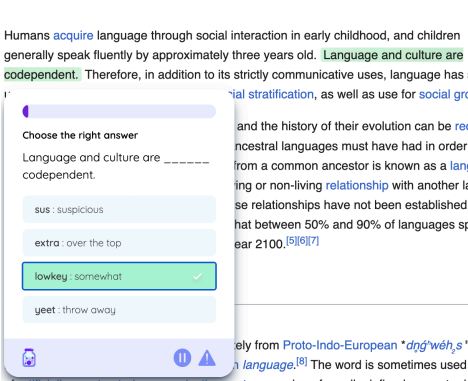
Figure 4a: GRE Mode - Advanced vocabulary transformation

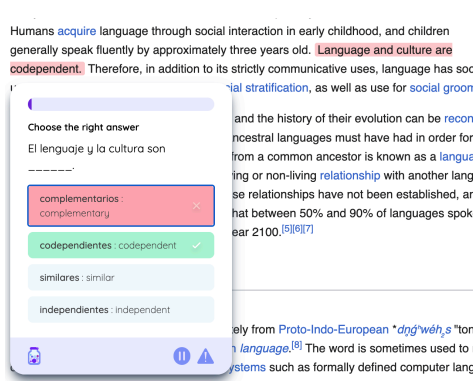*Figure 4b: TikTok Style - Contemporary social media language*

Figure 4c: Spanish Mode - English to Spanish translation

The GRE mode enhances vocabulary learning by replacing common words with their more sophisticated alternatives (e.g., "good" becomes "exemplary"), while TikTok style transforms formal English into contemporary social media expressions (e.g., "That's cool" becomes "That's bussin fr fr"). These AI-powered transformations maintain the original meaning while adapting to different language registers.

Research Questions and Hypotheses

After 10 months of operation and 92 users, this analysis investigates two key aspects of user behavior:

1. **Usage Context and Platform Patterns**

- Research Question: "What are the common contexts and platforms where users engage with Jargon?"
- Hypothesis: Users primarily engage with Jargon on social media or entertainment sites and banned academic sites.
- Rationale: Understanding where users naturally integrate Jargon into their browsing can inform platform-specific optimization and marketing strategies.

2. **Feature Adoption and User Success**
   - Research Question: "What features and settings distinguish active users from occasional users?"
   - Hypothesis: Active users utilize more customization options (density settings, highlight styles) and has achievable daily goals.
   - Rationale: Identifying the features that correlate with sustained engagement can guide onboarding improvements and feature prioritization.

Methods

Data Collection

The data for this analysis was collected from Jargon's Supabase database, covering user interactions from the extension's launch in June 2024 through March 16, 2025. The dataset comprises five main tables:

Dataset	Records	Description
Profiles	92	User profiles and settings
Questions	2442	Generated practice questions
Words	1594	Vocabulary entries and translations
Levels	117	User progression through difficulty levels
Websites	27	Websites where extension was disabled

Table 1: Overview of Dataset Components

Dataset Descriptions

1. Profiles Dataset

Variable	Type	Description	Notes
user_id	Primary Key	Unique identifier for each user	Anonymized identifier
level	Integer	Current proficiency level	Range: 1-10
paused	Boolean	Extension status on Chrome	TRUE/FALSE (Default: TRUE)
chrome_notifs	Boolean	Notification preferences	TRUE/FALSE
language	String	Current selected language mode	e.g., 'GRE Vocabulary', 'TikTok Slang'
last_question_time	DateTime	Timestamp of most recent question	UTC timezone
daily_goal	Integer	Target questions per day	User-set goal
density	Integer	Frequency of questions	Percentage of eligible sentences shown (0-100)
highlightStyle	String	Text selection preference	'highlight' or 'underline'

Table 2: Key Variables in Profiles Dataset

2. Questions Dataset

Variable	Type	Description	Notes
question_id	Primary Key	Unique question identifier	
user_id	Foreign Key	Associated user	References profiles
created_at	DateTime	Question generation time	UTC timezone
sentence	Text	Original selected text	English source content
word	String	Target word for learning	
language	String	Transformation mode	Selected language mode
original_sentence	Text	Source text	Pre-transformation content
options_array	Array of String	Multiple choice options	Even indices: options in target language; Odd indices: English translations
answered_at	DateTime	Completion timestamp	NULL if unanswered
chosen_option	String	User's answer	NULL if unanswered
user_rating	Integer	Question quality rating	Feature not yet implemented

Table 3: Key Variables in Questions Dataset

3. Words Dataset

Variable	Type	Description	Notes
created_at	DateTime	Word entry timestamp	UTC timezone
word	String	Target vocabulary	
language	String	Language mode	
user_id	Foreign Key	Associated user	References profiles
translation	Text	English translation	AI-generated translation
status	String	Learning status	Currently all set to 'learning'

Table 4: Key Variables in Words Dataset

4. Levels Dataset

Variable	Type	Description	Notes
user_id	Foreign Key	Associated user	References profiles
language	String	Language mode	
level	Integer	Difficulty level	Range: 1-10

Table 5: Key Variables in Levels Dataset

5. Websites Dataset

Variable	Type	Description	Notes
user_id	Foreign Key	Associated user	References profiles

website	String	Blocked URL	Sites where Jargon is disabled

Table 6: Key Variables in Websites Dataset

Data Processing

To prepare the data for analysis, we enhanced the profiles dataset by aggregating key user activity metrics from various tables. Through a series of table joins, we incorporated several derived variables as shown below.

Variable	Calculation	Purpose
generated_questions	Count of questions per user	Measure overall engagement
answered_questions	Count of questions with answers	Measure learning completion
blocked_sites	Count of blocked websites	Understand avoidance patterns
levels_attempted	Count of unique combination of languages and difficulty levels	Track learning progression

Table 7: Overview of Derived Variables

For consistency in our analysis, we handled missing values by replacing NAs with zeros in all count-based metrics. This enhanced profiles dataset serves as the foundation for our subsequent analyses of user engagement patterns. ## Data Exploration

Our exploratory data analysis examines patterns that inform both research questions about usage context and feature adoption. We organize our exploration into four main categories:
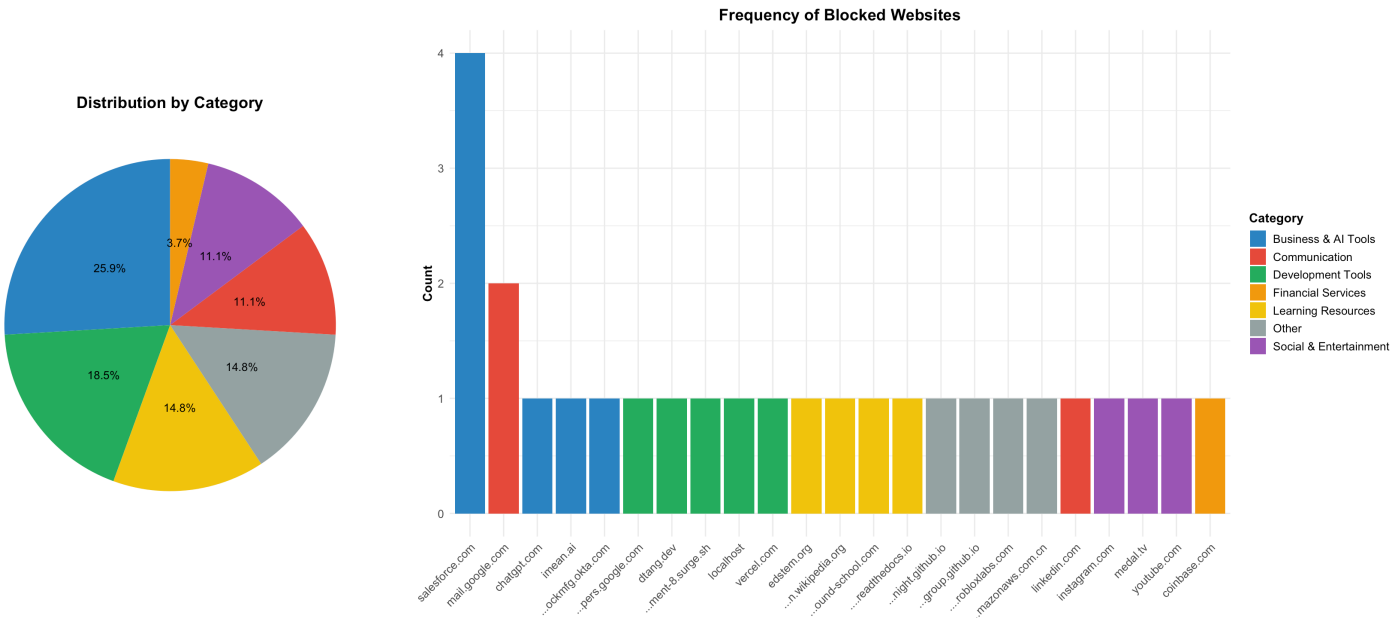
1. Blocked Website Patterns

Figure 5: Website Usage Analysis - Distribution of blocked websites by category (left) and frequency of individual websites (right)

The analysis of blocked websites reveals distinct patterns in how users interact with the Jargon extension. Professional tools—particularly Salesforce and AI platforms—are the most frequently blocked, suggesting that users tend to avoid using Jargon during work-related activities. The presence of development environment blocks indicates that some users are technical professionals, though this group represents only a modest portion of the overall user base. Educational content also features prominently among blocked websites, with users often disabling the extension on documentation sites and learning platforms, possibly to maintain focus during concentrated study sessions.

However, it is important to note that there are only 27 blocked sites across 92 users. This limited usage suggests that the blocking feature is not widely utilized, and the current data may not be conclusive. Caution should be exercised when generalizing these findings, as they may not fully represent the broader user population.
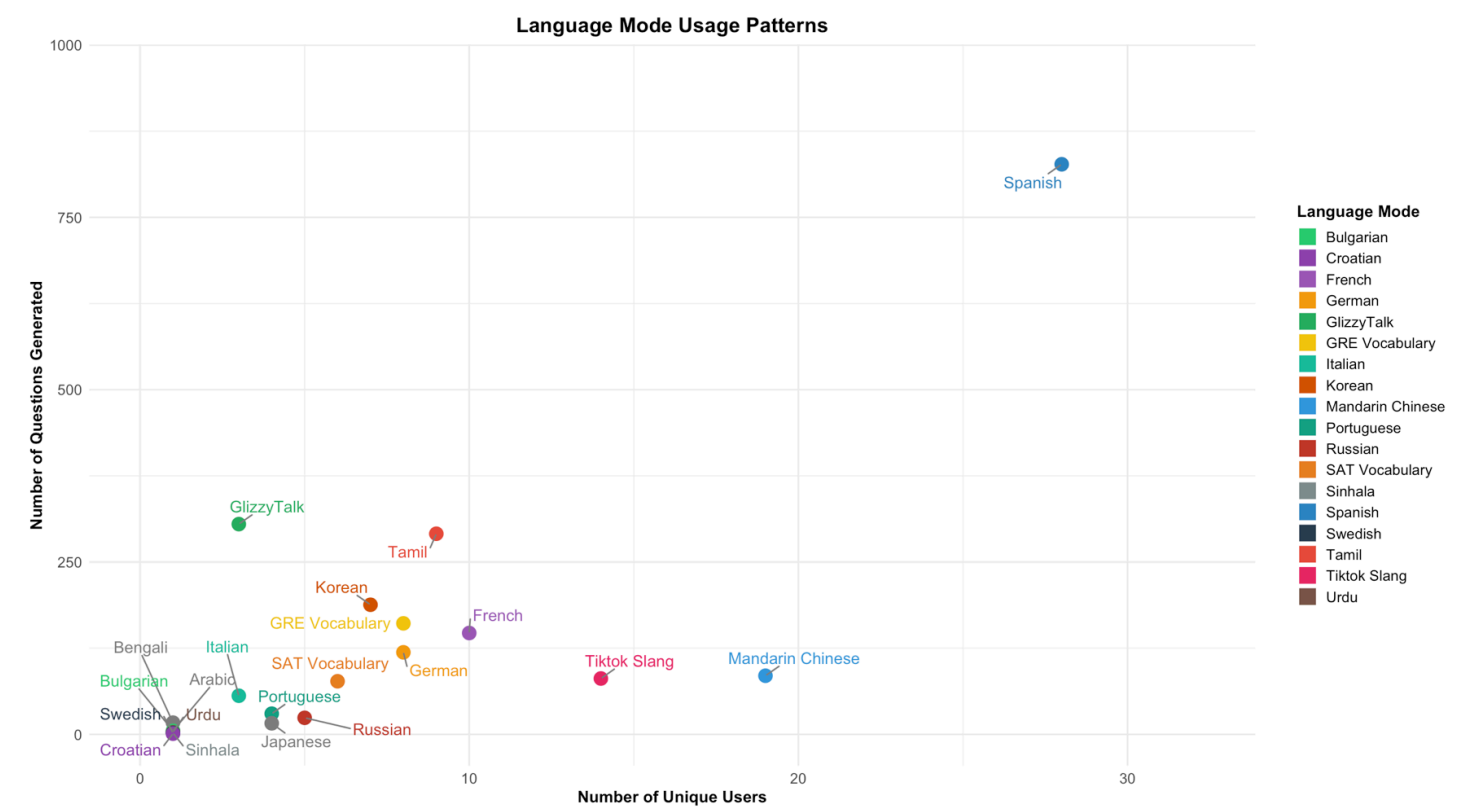
2. Language Mode and Feature Usage

Figure 6: Scatter plot showing the relationship between user adoption and question generation across different language modes

The scatter plot highlights key patterns in language mode usage:

- Spanish is the most active mode, with the highest number of questions (~800) and users (~30).
- GlizzyTalk and Tamil show moderate engagement (~300 questions each).
- Korean and GRE Vocabulary form a middle tier (~200 questions).
- Most other languages have low adoption, with fewer users and questions.
- Some modes (e.g., Tamil) have high question counts despite fewer users, indicating intensive use by dedicated learners.

Overall, while usage intensity and adoption vary widely across languages, traditional language learning modes drive most activity.
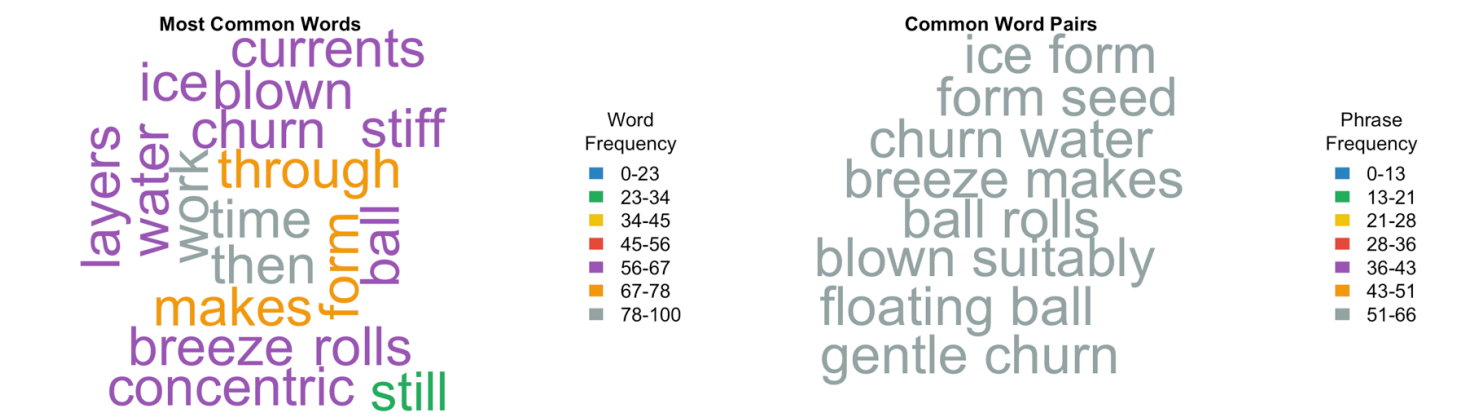
Figure 7: Word frequency analysis showing common words (left) and word pairs (right) in learning content. Colors indicate frequency of occurrence.

Insights from Word and Phrase Frequency Analysis (based on the English original sentences selected for content generation):

- The most common words and word pairs (e.g., "currents," "ice," "churn," "concentric," "ice form," "churn water") suggest that users frequently select technical or scientific content for practice, possibly from educational or informational sources. Descriptive and Process-Oriented Language:
- Many frequent terms describe physical processes or states (e.g., "breeze," "rolls," "floating ball," "gentle churn"), indicating an emphasis on dynamic or descriptive language in the learning material.
- The recurrence of similar words and phrases (e.g., "form," "water") implies that certain concepts or topics are repeatedly practiced, which may reflect user interests or the nature of the source material.

Overall, the word frequency analysis reveals that users are engaging most with scientific and descriptive content, focusing on process-oriented vocabulary and recurring technical terms.
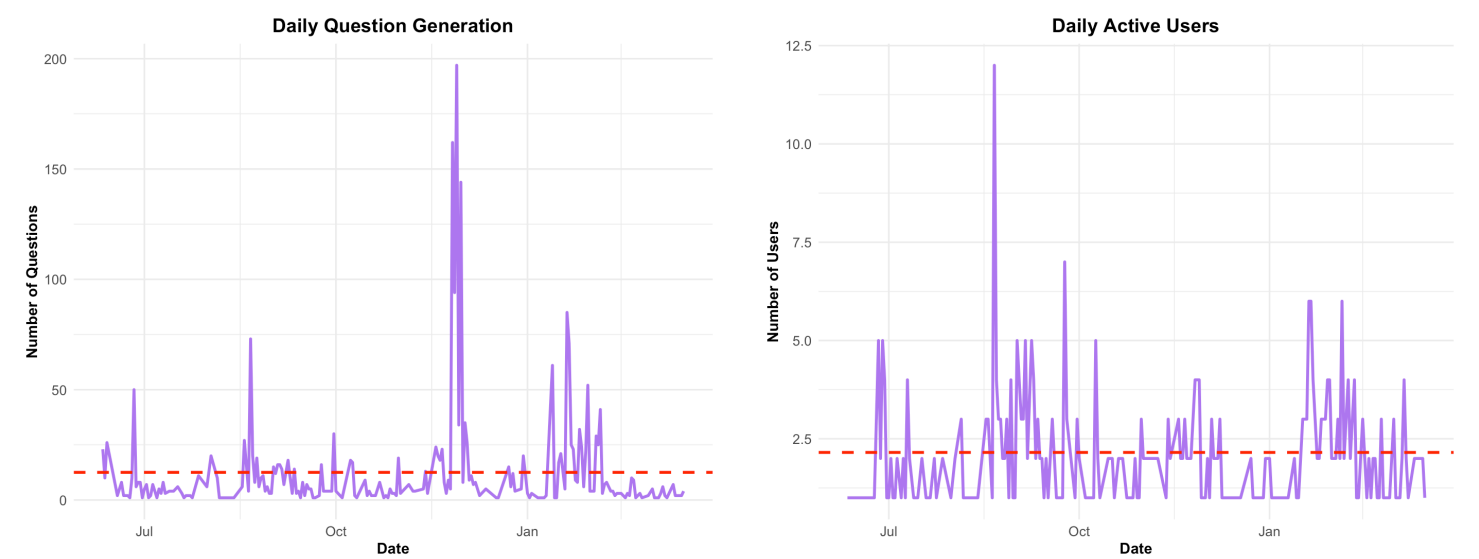
3. Temporal and Engagement Patterns

Figure 8: Daily activity patterns showing question generation and active users with their respective averages (dashed lines) over the observation period, based on UTC timezone. Questions average: 12.5 per day; Users average: 2.2 per day.
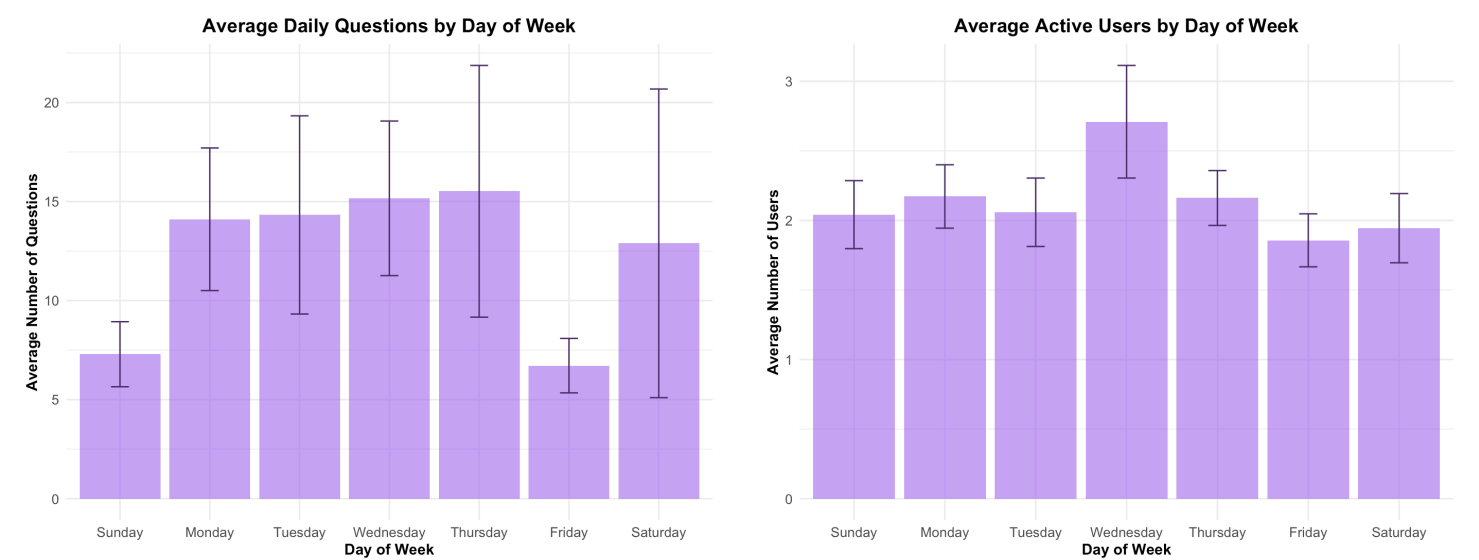
Figure 9: Weekly activity patterns showing average questions generated and active users by day of week (UTC timezone), with error bars indicating standard error.

The temporal analysis reveals several key patterns in user engagement, based on both daily and weekly activity (all timestamps in UTC):

- **Daily Trends::** Question generation and active user counts fluctuate considerably day-to-day, with occasional spikes (up to 200 questions or 12 users), but most days remain below the average (12.5 questions, 2.2 users).This indicates a small but steady user base, with 1–5 active users on most days.

- **Weekly Trends:** Question generation is highest on Mondays, Tuesdays, and Wednesdays, then tapers off toward the weekend,suggesting users are more engaged during the workweek. There is substantial variability across days, as shown by the error bars.

Together, these patterns indicate that Jargon's usage is characterized by low but regular engagement, with activity peaking midweek and significant day-to-day variability. This suggests a core group of users who interact with the platform most during the workweek.

4. User Engagement Distribution

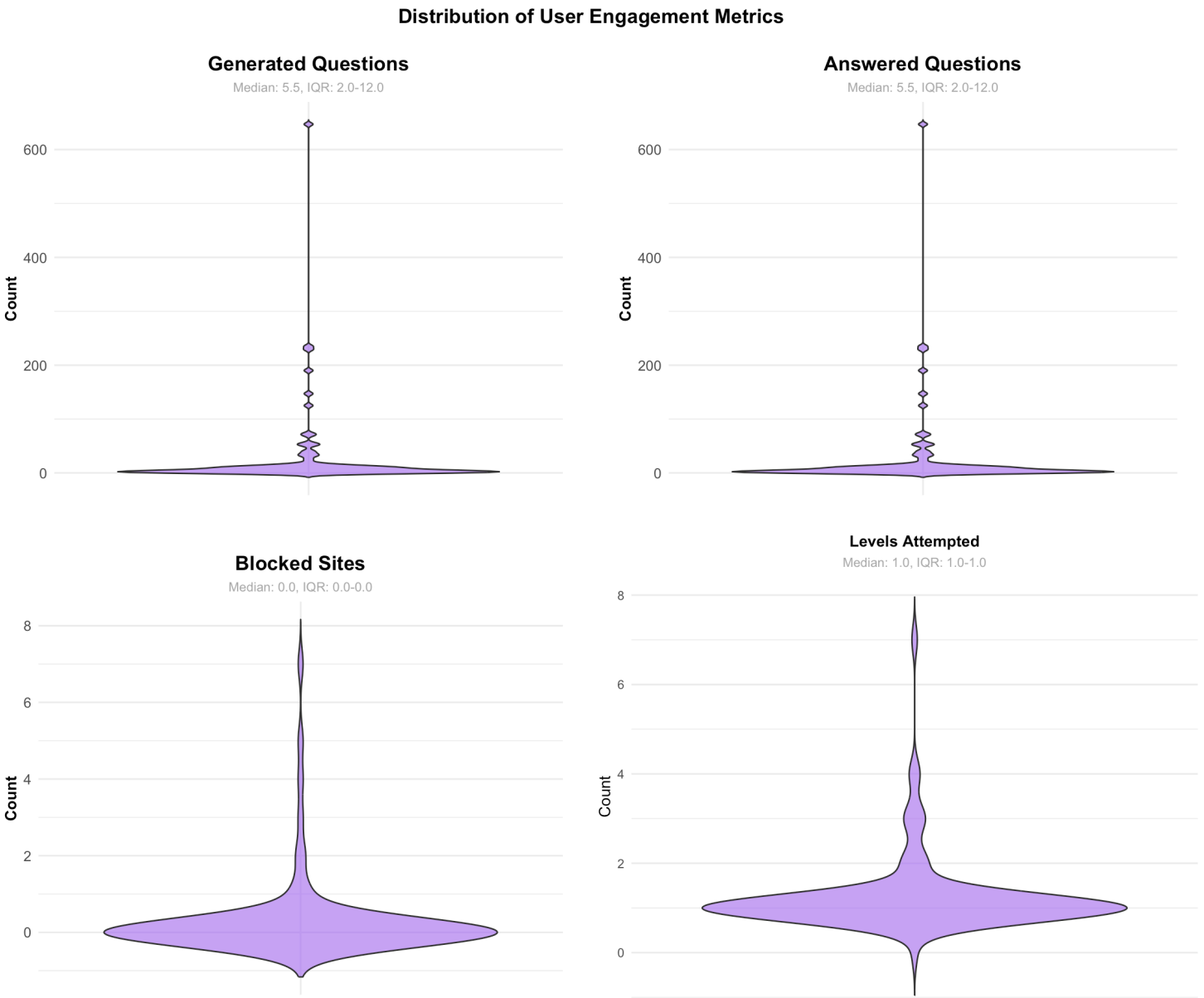**Distribution of User Engagement Metrics**

Figure 10: Distribution of key engagement metrics across users, showing individual violin plots for each metric with median and interquartile range (IQR) statistics. Each plot uses a distinct color and includes summary statistics.

The violin plots provide a clearer view of the distribution of user engagement metrics:

- **Generated Questions & Answered Questions:** Most users generate and answer only a small number of questions, as shown by the wide base near zero. A few users are highly active, producing a long tail of outliers with much higher counts.
- **Blocked Sites:** The vast majority of users do not block any sites (distribution concentrated at zero), with only a handful blocking more than one site.
- **Levels Attempted:** Most users attempt only one level, with very few exploring multiple levels. The distribution is sharply peaked at one, with a small tail for higher values.

Overall, the violin plots highlight that engagement is highly skewed: most users interact minimally, while a small subset are much more active or exploratory. This pattern is consistent across all four metrics.

Further Analysis

Research Question 1: Usage Context and Platform Patterns

Sentiment Analysis of User-Selected Content

To further address our first research question—"What are the common contexts and platforms where users engage with Jargon?"—we performed sentiment analysis on the English original sentences that users selected for learning. Using the `syuzhet` package in R, each sentence was assigned a sentiment score, where positive values indicate positive sentiment, negative values indicate negative sentiment, and values near zero indicate neutral sentiment. This approach allows us to quantitatively assess the emotional tone of the content users choose to engage with.

Topic Modeling of User-Selected Content (LDA)

To further explore the contexts in which users engage with Jargon, we applied Latent Dirichlet Allocation (LDA) topic modeling to the English original sentences selected by users. In addition to standard stopwords, we removed a custom list of common or uninformative words to improve topic quality. This method uncovers the main themes or topics present in the content users choose to learn from.

Research Question 2: Feature Adoption and User Success

Correlation Analysis

To explore the relationships between user features and engagement metrics, we computed a correlation matrix for key variables in the enhanced_profiles dataset. This helps identify which features are associated with higher engagement or other usage patterns.

K-Means Clustering to Define Active vs. Occasional Users

To provide a data-driven segmentation of user engagement, we applied k-means clustering to the key engagement metrics: generated questions, answered questions, blocked sites, and levels attempted. This approach groups users into clusters based on their overall activity patterns, rather than relying on arbitrary thresholds or quantiles.
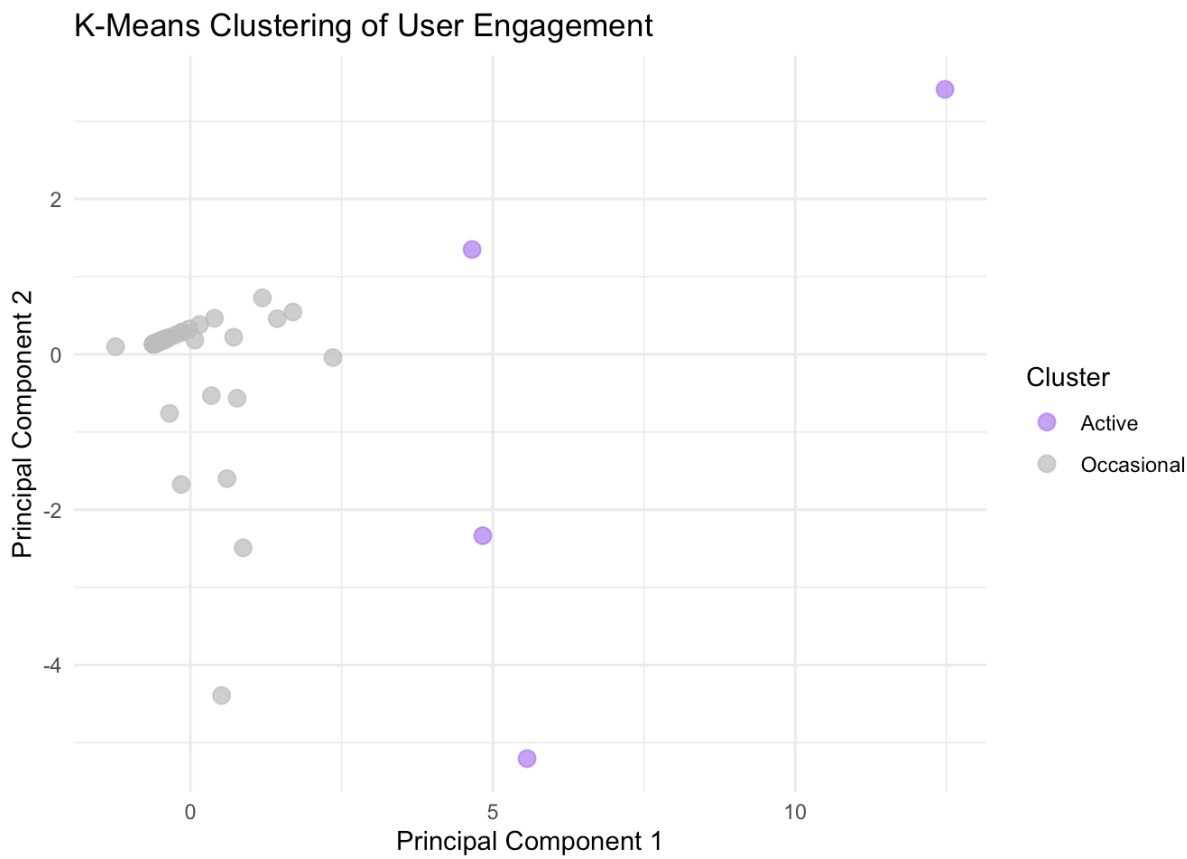
Figure 11: K-means clustering of users based on engagement metrics. Each point represents a user, colored by cluster (Active or Occasional) in PCA space.

Due to the high dimensionality of the data, PCA was used to reduce the feature dimensions from 4 to 2 for data visualization. The PCA plot of K-means clustering reveals a clear separation between two user groups: 'Active' and 'Occasional'. The majority of users are clustered closely together in the 'Occasional' group, indicating similar and relatively low engagement across key metrics. In contrast, only a few users are classified as 'Active', and these are well separated from the main cluster, highlighting their much higher engagement levels. This pattern suggests that while most users interact with the platform at a modest level, a small subset of users are highly engaged, driving much of the activity. The presence of only a few very active users is typical in many online platforms, where a minority of users contribute disproportionately to overall engagement. This finding aligns with Figure 10 from our Data Exploration.

Summary Statistics by K-Means Group

kmeans_group	n	mean_generated_questions	mean_answered_questions	mean_blocked_sites	mean_levels_attempted
Active	4	325.25000	325.25000	2.7500000	4.500000
Occasional	88	12.96591	12.96591	0.1818182	1.113636

Table 8: Summary statistics for k-means-defined active and occasional users

The summary statistics in Table 8 show a clear gradient in engagement: 'Active' users (identified by k-means) have much higher activity across all metrics, but are few in number. To provide a more meaningful middle group, the 'Occasional' category (top 25% by generated questions among non-active users) was added, since k-means alone identified only a small number of highly engaged users. This three-group segmentation allows for a more nuanced comparison of user behaviors.

User Segmentation: Very Active, Active, and Regular Users

To provide a more nuanced segmentation, we define three user groups:

- **Very Active:** Users in the k-means 'Active' cluster (highest engagement across all metrics)

- **Active:** Among the remaining users, those in the top 25% by generated_questions
- **Regular:** All other users

user_group3	n	mean_generated_questions	mean_answered_questions	mean_blocked_sites	mean_levels_attempted
Active	23	38.826087	38.826087	0.2608696	1.304348
Regular	65	3.815385	3.815385	0.1538462	1.046154
Very Active	4	325.250000	325.250000	2.7500000	4.500000

Table 9: Summary statistics for Very Active, Active, and Regular users

The summary statistics in Table 9 show a clear gradient in engagement: 'Very Active' users (identified by k-means) have much higher activity across all metrics, but are few in number. To provide a more meaningful middle group, the 'Active' category (top 25% by generated questions among non-very-active users) was added, since k-means alone identified only a small number of highly engaged users. This three-group segmentation allows for a more nuanced comparison of user behaviors.

After user segmentation, we plan to use an ordinal regression model, a statistical technique for modeling ordered categorical outcomes, to identify important features and their significance. This approach is chosen because it effectively assesses the impact of variables on ranked user activity, providing insights crucial for enhancing targeted strategies.
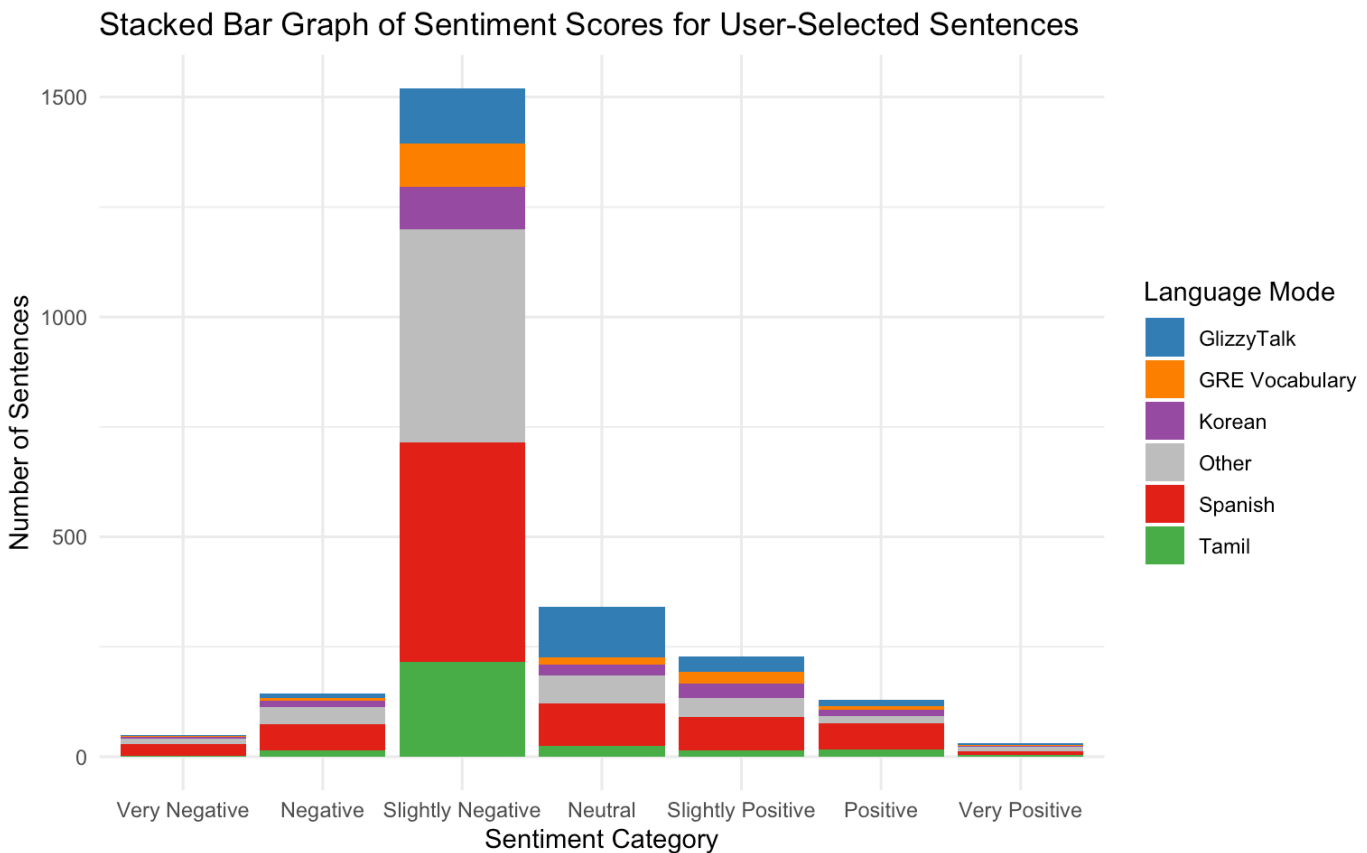
Results

Sentiment Analysis

Figure 12: Stacked bar graph showing the frequency of user-selected sentences in each sentiment category, stacked by language mode (top 5 languages shown in color, all others in grey).

The stacked bar graph displays the overall distribution of sentiment categories, highlighting the top five language modes in color and grouping all other languages in grey. This visualization underscores the predominance of neutral and slightly negative content, with a similar sentiment distribution observed across each language.
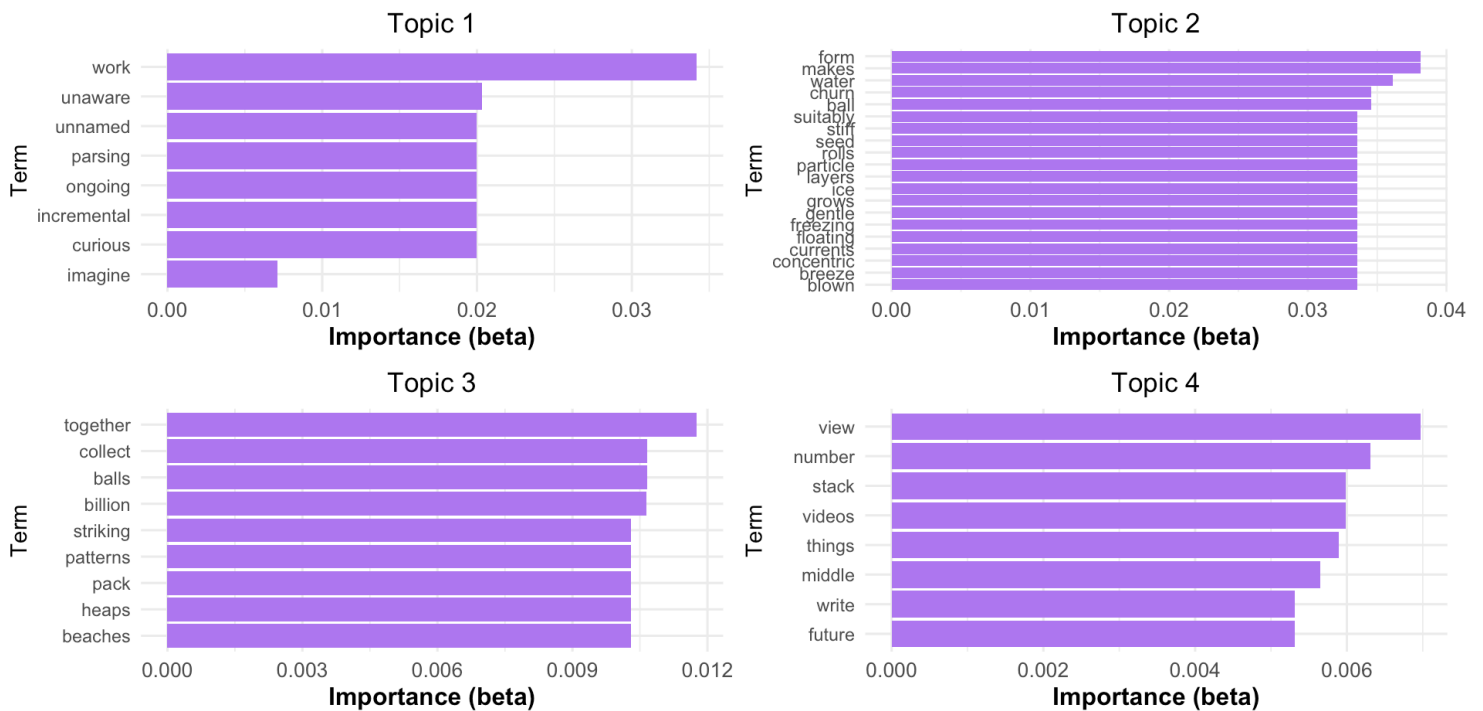
LDA Analysis

Figure 13: Top terms for each topic identified by LDA topic modeling of user-selected sentences. Each panel shows the most important words for one topic, with x-axis numbering visible for all.

The LDA topic modeling did not yield strong or actionable insights about the contexts or platforms where users engage with Jargon. The "Importance (beta)" values are all quite low (well below 0.05), which is typical for LDA on short texts or small datasets, but it also means that no single word dominates any topic. The topics identified are diffuse, with mostly generic or process-oriented terms. This suggests that either the user-selected content is too varied or generic for topic modeling to be effective, or that the dataset is not large or rich enough for LDA to find meaningful structure. This is a valid finding: not all analyses reveal clear patterns, and reporting this transparently demonstrates scientific rigor. It may also indicate that user engagement with Jargon is broad and not easily categorized, or that more data is needed for deeper insights.

Despite the weak themes, a tentative interpretation of the topics is as follows:

- **Topic 1:** May relate to work processes or technical tasks (e.g., "work," "parsing," "incremental," "curious").
- **Topic 2:** Appears to focus on scientific or physical phenomena, especially related to water and movement (e.g., "form," "water," "ice," "currents," "breeze").
- **Topic 3:** Suggests group actions or collective activities (e.g., "together," "collect," "balls," "patterns").
- **Topic 4:** Includes terms that could relate to data, viewing, or content creation (e.g., "view," "number," "stack," "videos," "write").

However, these interpretations are tentative due to the low importance values and the generic nature of the terms.
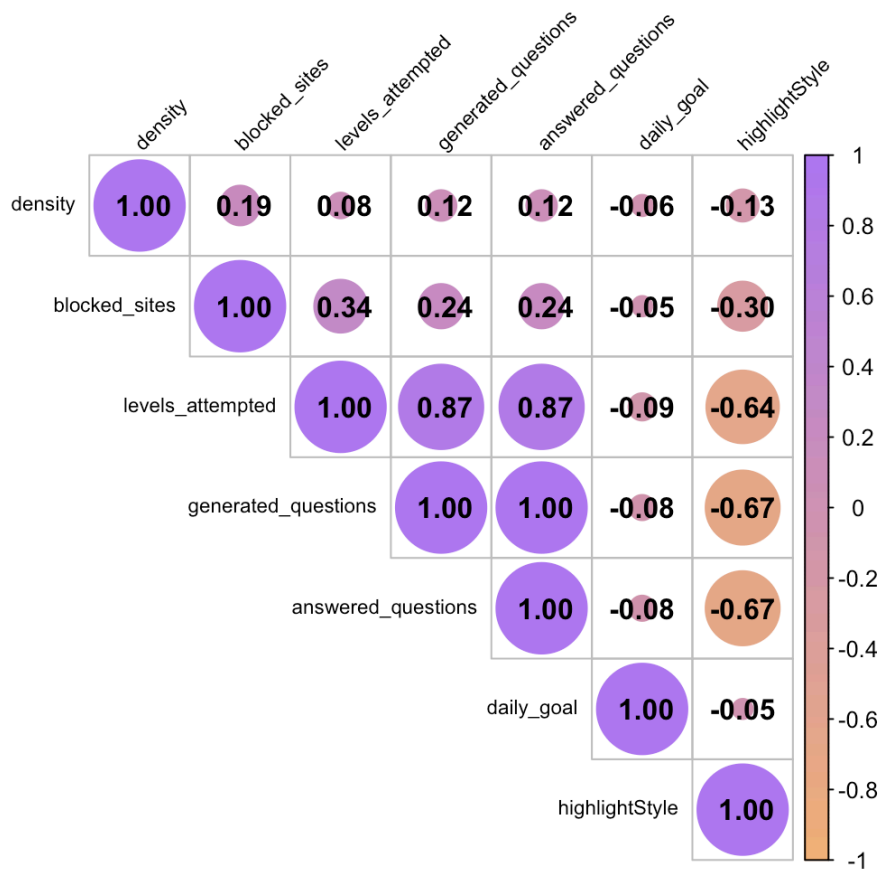
Corelation Matrix

Figure 14: Correlation matrix of user features and engagement metrics. Circle size and color indicate the strength and direction of the correlation (purple = positive, orange = negative).

The correlation matrix reveals several notable relationships:

- **Levels Attempted, Generated Questions, and Answered Questions** are all very strongly and positively correlated ($r \approx 0.87$–1.00), indicating that users who attempt more levels also generate and answer more questions—these are the most engaged users.
- **Highlight Style** shows a strong negative correlation with both Generated and Answered Questions ($r \approx -0.67$), suggesting that users who prefer a particular highlight style (as encoded numerically) tend to engage less.
- **Blocked Sites** is moderately positively correlated with Levels Attempted and Engagement ($r \approx 0.24$–0.34), implying that more engaged users are also more likely to block sites, possibly to focus their learning.
- **Density** and **Daily Goal** show weak or negligible correlations with engagement metrics, suggesting these settings do not strongly predict user activity in this dataset.

Overall, the strongest signals are that higher engagement (more questions and levels) tends to cluster together, and that highlight style preference is inversely related to engagement. Most other relationships are weak, indicating a diversity of user behaviors and settings.

Feature Comparison by User Group

term	estimate	std.error	statistic	coef.type	p.value
daily_goal	-0.031	0.012	-2.607	coefficient	0.009
density	0.014	0.010	1.374	coefficient	0.169
highlightStyleunderline	-4.570	1.277	-3.579	coefficient	0.000
Regular/Active	-3.909	1.442	-2.710	scale	0.007

Active/Very Active	-0.924	1.297	-0.712	scale	0.476

Table 10: Ordinal regression coefficients for user group score (2 = Very Active, 1 = Active, 0 = Regular)

Table 10 presents the results of the ordinal regression model comparing user features across the three user groups. The threshold terms (e.g., 'Regular/Active' and 'Active/Very Active') represent the estimated boundaries on the latent engagement scale that separate the user groups; they are not feature effects but model cutpoints.

Among the features, two are statistically significant predictors of user group ($p < 0.05$):

- **Daily Goal**: The negative coefficient for daily_goal (estimate = -0.031, $p = 0.009$) indicates that users who set higher daily goals are less likely to be in a more active group. This suggests that overly ambitious targets may discourage sustained engagement, possibly due to unrealistic expectations or burnout.
- **Highlight Style (Underline)**: The large negative coefficient for highlightStyleunderline (estimate = -4.57, $p < 0.001$) shows that users who prefer the underline style are much less likely to be highly engaged. This may reflect a preference for a less prominent or less motivating interface.

The density setting is not a significant predictor ($p = 0.169$), indicating it does not have a clear association with user activity level in this dataset. Overall, these results highlight that both goal-setting behavior and interface preferences are important factors in distinguishing more and less engaged users.

Conclusions and Summary

This study provides a comprehensive analysis of user engagement and feature adoption on the Jargon platform, with a focus on understanding usage contexts and identifying the characteristics of highly engaged users.

1. Usage Context and Platform Patterns

- **Findings:**
  Analysis of blocked website data and language mode usage reveals several key patterns:
  - Spanish is the most popular language mode
  - The platform sees consistent usage across different times of day and days of the week, with slight variations in activity patterns
  - Users demonstrate diverse preferences in how they interact with the platform, from casual browsing to focused study sessions
- **Implications:**
  These insights suggest multiple opportunities for platform enhancement:
  - Marketing Strategy: Focus on highlighting Jargon's versatility across different contexts:
    - Language Learning: Emphasize Spanish and other popular language modes
    - Academic Enhancement: Showcase GRE vocabulary and professional content features
    - Social Media Integration: Demonstrate how Jargon enhances casual browsing experiences
  - User Experience: Consider context-specific optimizations:
    - Educational Sites: Enhance features for academic content comprehension
    - Professional Platforms: Develop specialized vocabulary sets for different industries
    - Social Media: Optimize for quick, engaging interactions

2. Feature Adoption and User Success

- **Findings:**
  The segmentation and regression analyses reveal important insights about user engagement:
  - Goal Setting: Users with moderate, achievable daily goals (lower targets) show higher sustained engagement
  - Interface Preferences: The highlight style significantly impacts user activity:
    - Users preferring the highlight style show higher engagement levels
    - The underline style correlates with lower activity levels
  - Feature Usage: While some settings (like density) don't significantly predict engagement, the combination of

appropriate goals and preferred visual style appears crucial
- User Segments: Clear distinctions emerge between very active, active, and regular users, with each group showing different patterns of feature utilization

- **Implications:**

These findings suggest several actionable improvements:
- Onboarding Optimization:
  - Implement smart goal-setting guidance with recommended ranges based on user patterns
  - Offer A/B testing of highlight styles during initial setup
  - Provide interactive tutorials for optimal feature configuration
- Interface Refinement:
  - Consider making the highlight style the default option
  - Develop new visual styles based on user feedback
  - Implement adaptive settings that adjust based on usage patterns
- Engagement Strategy:
  - Create personalized engagement paths for different user segments
  - Develop feature discovery prompts based on user activity levels
  - Implement achievement milestones aligned with realistic usage patterns

Recommendations for Data Collection Improvement

- **Feature Tracking:**
  Enhance data collection by recording the website URL for each user interaction, enabling more granular analysis of usage context.
- **Time Zone Awareness:**
  Shift from tracking all activity in UTC to using users' local time zones. This would allow for personalized recommendations, such as suggesting study breaks or optimal engagement times.

Overall, Jargon users are most engaged when the platform aligns with their study habits and offers achievable, motivating features. By leveraging these insights, Jargon can further enhance user retention and satisfaction, supporting both language learning and productivity.