# NIM: Neuro-symbolic Ideographic Metalanguage for Inclusive Communication

**Prawaal Sharma**
Infosys
Pune, Maharashtra, India
prawaal_sharma@infosys.com

**Poonam Goyal**
BITS Pilani
Pilani, Rajasthan, India
poonam@pilani.bits-pilani.ac.in

**Navneet Goyal**
BITS Pilani
Pilani, Rajasthan, India
goel@pilani.bits-pilani.ac.in

**Vidisha Sharma**
BITS Pilani
Goa, India
vidishasharma@gmail.com

## Abstract

Digital communication has become the cornerstone of modern interaction, enabling rapid, accessible, and interactive exchanges. However, individuals with lower academic literacy often face significant barriers, exacerbating the "digital divide". In this work, we introduce a novel, universal ideographic metalanguage designed as an innovative communication framework that transcends academic, linguistic, and cultural boundaries. Our approach leverages principles of Neuro-symbolic AI, combining neural-based large language models (LLMs) enriched with world knowledge and symbolic knowledge heuristics grounded in the linguistic theory of Natural Semantic Metalanguage (NSM). This enables the semantic decomposition of complex ideas into simpler, atomic concepts. Adopting a human-centric, collaborative methodology, we engaged over 200 semi-literate participants in defining the problem, selecting ideographs, and validating the system. With over 80% semantic comprehensibility, an accessible learning curve, and universal adaptability, our system effectively serves underprivileged populations with limited formal education.

## 1 Introduction

Communication, in its many forms, has always been the bedrock of human progress, shaping societies and driving innovation. In the modern interconnected world, digital communication has emerged as a pivotal force, revolutionizing how we communicate. We propose an ideographic communication metalanguage for semi-literates, promoting socio-digital inclusion, and a more interconnected society. We define semi-literates as individuals with limited formal education typically only up to the primary or early secondary level who face significant challenges in navigating digital communication platforms due to constrained reading, writing, and digital literacy skills.

Visual communication is intuitive, easy to comprehend, language agnostic, containing spatial and directional attributes (Lodding, 1983; Van Dam, 1984; Maguire, 1985; Tatti, 2016). While this can help in the facilitation of human-computer communication specially for people with limited academic education, however having a set of visual ideographs only lacks morphology and syntax, which is ingrained in orthographic forms of communication thus making pure visual ideographic set prone to misinterpretation. Earlier studies on the use of visual communication for academically challenged individuals observe challenges with learnability (due to large inventory of ideographs), extensibility (non-universal design) along with ambiguity (lack of grammar) (Sharma et al., 2023).

Figure 1 presents the conceptual foundation of our work, embodying the neuro-symbolic AI paradigm through a synthesis of symbolic reasoning and neural parametric learning. By augmenting and combining the strength of Large Language Models (LLMs) enabling generation of new concepts, along with structured ontology as guiding principles for semantic decomposition of complex ideas into simpler and atomic concepts, we design a novel ideographic metalanguage for semi-literates. The pipeline goes through processing of input sentence $S_{in}$ into distinct parts of picturable ($W_p$) and textual (non picturable, $W_t$) sections. While $W_p$ gets broken down into ideographic elementary components ($I_e$) which comprises of Semantic Classes, Templates, Variables and Molecules ($SC/ST/(sv, sm)$), the non picturable sections are preserved as plain text coming together in an interactive user interface customized for semi-literate users. Consequently, we call our method as NIM (Neuro-symbolic Ideographic Metalanguage).



*Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 10322–10340

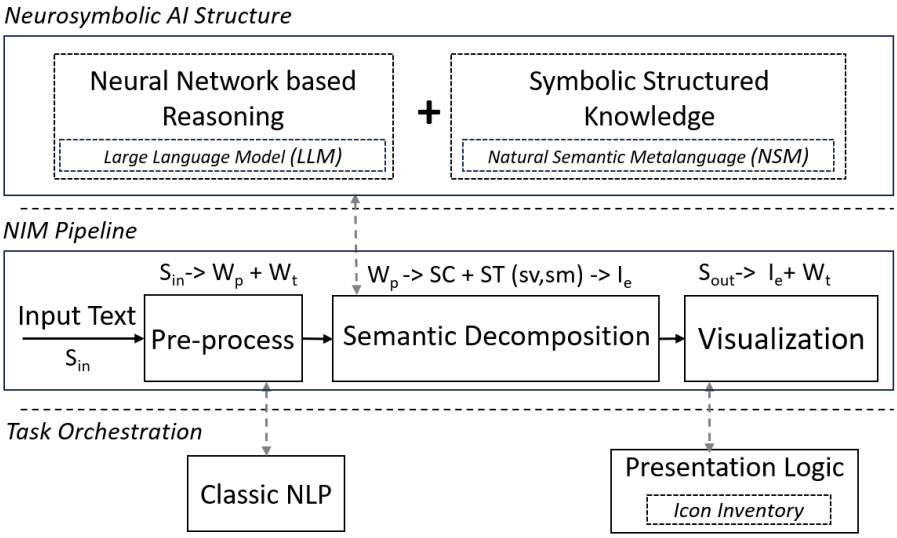



Figure 1: High Level NIM architecture, comprising of Neuro-symbolic AI, Pipeline and Orchestration layers.

NIM leverages the linguistic theory of Natural Semantic Metalanguage (NSM) (Wierzbicka, 1996; Goddard, 2006) for symbolic reasoning module achieving semantic simplification of complex concepts into more elementary elements. This symbolic reasoning is integrated with Generative Artificial Intelligence (Gen-AI) and construct the design of a visual metalanguage (including bare-minimum text, we refer as *binding text*) for effective communication in semi-literate settings. We further enhance our work by making the binding text multilingual thereby making it universal for use.

We adopt a collaborative Human-Centered Design (HCD) methodology that actively involves semi-literate individuals across the development cycle, facilitating authentic problem discovery, participatory design, and rigorous empirical validation. The validation spans multiple dimensions, including comprehensibility, user engagement, ideographic effectiveness, and extensibility. Beyond semi-literate populations, our approach has potential applicability to diverse user groups such as individuals with intellectual disabilities, multilingual teams, and children with dyslexia.

In summary our contribution, is three-fold, (i) We build a novel metalanguage for semi-literates, enabling them for digital communication; (ii) Our work facilitates seamless integration with digital ecosystem, for easy and intuitive human-computer interaction; and (iii) Our work is universal, extensible for multiple domains, geographies and cultures.

It can be argued that the speech-based systems and vision-language models (VLMs) offer promising modalities for accessibility and can be used to solve the problem of digital communication for semi-literate groups. However, despite their promise, these techniques exhibit important gaps that limit their effectiveness in the real-world, semi-literate contexts.

The modality of speech is promising for accessibility, however for communication across communities with significant linguistic diversity (no common standard spoken language), sending voice messages does not help. For instance, in India, where languages and dialects change every few hundred kilometers, voice-based communication cannot reliably support broader accessibility or facilitate interaction across linguistically diverse communities.

Similarly VLMs can possibly generate intermediate images, but it has been observed in earlier "text to image" systems that the choice of the icon for a particular situation/entity is not consistent and the number of icons and their interplay is also non deterministic. VLMs typically produce monolithic visual outputs rather than structured representations. This lack of hierarchical organization limits opportunities for contextual disambiguation and, consequently, hampers the comprehensibility of the final output.

## 2 Related Work

### 2.1 Visual forms of communication

Visual forms of communication encompassing both linguistic and computational approaches include a variety of forms including, (i) *Semantographics*,

having a fixed set of ideographs to communicate high level semantic ideas (Sharma et al., 2023) (ii) *Emojis and emoticons*, convey emotions, tone and context (Seargeant, 2019) (iii) *Way-finding boards*, help navigate at public places including airports, etc. with direction cues (Canter, 1996) (iv) *Infographics*, in forms of charts and graphs conveying insights and relationships in data (Siricharoen, 2013) (v) *Maps*, for geographical relationships (Shirreffs, 1992) and (vi) *Gestures and body language*, for communication of emotion and intentions (Imai, 2005). All these visual forms of communication play a crucial role in human interaction, facilitating effective transmission of information across and work across language barriers.

To the best of our knowledge, none of the existing methods of visual communication incorporate hierarchical simplification on semantic complexity keeping the ideographic inventory smaller in number, leverage the interactive capabilities of digital platforms and are not universal for use.

### 2.2 Neuro-symbolic AI

Neuro-symbolic AI refers to AI systems that seek to integrate neural network-based methods with symbolic knowledge based approaches (Hitzler and Sarker, 2022; Sheth et al., 2023; Hitzler et al., 2024). Embodying intelligent behavior in an AI system must involve both cognition using guiding principles (coming via Symbolic Structures) along with generative and pattern processing capabilities (coming from Neural Structures). Combining both systems enables verifiable reasoning, planning, generalization and abstraction. We use Natural Semantic Metalanguage (NSM) for building Symbolic principles and Generative AI for scale and generation of new concepts.

Natural Semantic Metalanguage (NSM) advocates that all languages in the world irrespective of their origin and timeline contain a semantic core of universals (referred to as semantic primes and molecules in NSM) and this core forms the backbone of human communication (Wierzbicka, 1996; Goddard, 2006; Goddard and Goddard, 2018).

Generative AI in the recent years, have become the backbone for most NLP tasks primarily text generation/ understanding/ inferencing (de Salvo Braz et al., 2006). LLMs encapsulate world's knowledge in form of parametric memory and can be very instrumental in referencing new concepts not seen in model trainings (Radford et al., 2018; Touvron et al., 2023).

## 3 Problem Discovery

To the best of our knowledge, no prior work directly aligns with our approach of combining minimal textual elements for syntax with symbolic representations as the primary carriers of semantics. Existing studies largely fall into three categories: purely symbolic systems, text-based approaches, or the use of emoticons and emojis, which are typically employed as paralinguistic cues (e.g., to signal emotions or attitudes). Consequently, there is no directly comparable baseline in the literature, and we therefore adopt a first-principles approach to problem formulation, system design, and solution development.

### 3.1 User Needs Assessment and Benchmarking

We partner with a mix of urban and rural semi-literate subjects from Indian demography to identify their challenges and expectations from digital communication platforms. A voluntary group of 20 semi-literates (academic education less than twelve years and face challenges with digital communication) participants ($P = \{p_j | j = 1, 2, ...20\}$) were engaged, to discover problems, preferences, and challenges. We illustrate further details about the users engaged in Appendix B.

We request the participants to use existing ideographic methods (BETA, SCLERA and ARASAAC) (Sevens et al., 2015; Hervás et al., 2020) and interpret a set of twenty five sample messages ($M = \{m_i | i = 1, 2, ...25\}$) over a period of five days (which means five ideographic messages each day). The participants were made familiar with these scripts (not an exhaustive training though), at the start of the workshop and feedback provided on each day based on how they responded. Each day, we collect interpretations of the participants (as text) for each message $m_i^j$ ($i$ is message id and $j$ is participant id) and compare these with the actual message text ($a_i$) used to generate the initial message $m_i$. We use comprehensibility (to measure usability) and learnability (to measure learning potential) as essential metrics for benchmarking current methods, and we use the same method (plus additional metrics on user experience, and effectiveness) later when we validate our method. The results are illustrated in Table 1, and further in Appendix A.

Comprehensibility measures how easily users can understand the language, which can be mea-

| | Comprehensibility | Learnability |
|---|---|---|
| **SCLERA** | 0.42 - 0.46 | 0.049 |
| **BETA** | 0.38 - 0.41 | 0.036 |
| **ARASAAC** | 0.43 - 0.46 | 0.036 |

Table 1: Benchmarking existing pictographic methods. (Values for comprehensibility are day (1 - 5) numbers)

sured by evaluating semantic comprehension of sample messages. using Meteor $(M)$ as a metric for comprehensibility, which compares common words considering synonyms, stemming, and word order (Banerjee and Lavie, 2005). For each participant $p_j$, for each day we compare the message interpretations with actual message text $\{\frac{1}{5}M(m_i^j, a_i)|i = 1, 2, ...5\}$ and compute average comprehensibility for each participant $(c_j)$. Finally, we average the scores for all 20 participants and calculate the final comprehensibility $(c)$ for each day.

Learnability measures how quickly users can learn to use the language proficiently. We also consider the phenomenon of *Plateau Effect* in learning, which occurs because initial gains in proficiency are typically easier and faster compared to mastering finer details or nuances (Grossman et al., 2009). We use normalized learnability metric calculated as the mean of normalized difference of comprehensibility on day 5 $(c_5)$ and day 1 $(c_1)$ on comprehensibility along with penalty to accommodate plateau effect $((c_5 - c_1)/c_1)\dot{W}_t$ where $W_t = 1 - (|c_5 - T|/T)$, and T is the learning efficiency threshold, at which the rate of learning slows down (Grange and Mulla, 2015) (we use 90% as threshold value).

### 3.2 Dataset Identification

While user assessment helps to identify existing issues and benchmark current methods, it is very important to identify an appropriate and relevant dataset which can provide meaningful insights in the semi-literate texting behavior enabling better generalization to real-world scenario. We use an existing dataset (3000+ messages; 300+ respondents) from rural and urban Indian demographics, manually collected by talking to semi-literates (Sharma et al., 2021).

### 3.3 Defining the Ideographic Scope

In order to achieve high comprehensibility and learnability, it is important to balance semantic simplification (for complex words) and syntactic preservation (keeping some text as is). We partition the words in the sentence into two categories, the first category contains complex words that needs to be simplified and ideographed, while the second category comprises easy to understand text that acts as the binding element, maintaining sentence structure and flow.

It has been observed that readability scores give an indication of the ease of comprehension primarily on lexical complexity, familiarity, legibility and typography. We use a variety of readability scores to evaluate the complexity of various words in the identified dataset (Farr et al., 1951; Williams, 1972; Gunning, 1969; Kincaid et al., 1975) and conclude that nouns are most complex concepts in the sentence, followed by verbs. We also observe that majority of the dataset volume (66% in this corpus) is composed of nouns and verbs. It is also noted that the length of text message exchanged is limited to 10 words at the most (Mean 7.2, Median 7, Mode 6). We also cross validate our observations on another corpus (Sketch Engine containing a corpus of 36 billion words) and confirm similar findings (Kilgarriff et al., 2004; Kunilovskaya and Koviazina, 2017). Hence, we scope our work for small short text messages ($\leq$20 words) along with enabling semantic simplification for all nouns and complex verbs (preserving the rest of text as binding element).

## 4 System Design

As illustrated in Figure 1, we use Neuro-symbolic AI architecture having two separate components for defining guiding principles (Symbolic Structure) and pattern processing abilities (Neural Structure).

### 4.1 Design of Symbolic Structure

We overlay the architecture of our symbolic structure on the linguistic theory of NSM so as to establish the hierarchy of foundational semantic concepts (referred as Semantic Universals from here on-wards).

NSM recommends semantic universals to follow a three level hierarchy, (i) The first level designates semantic concepts in the same broader category referred to as *Semantic Classes (SC)* and consists of categories like human class, event class, etc.; (ii) The second level indicates more specific patterns or finer distinctions within SC and can be described with same constituents. This is referred as *Semantic Templates (ST)*, including concepts like human

relationships etc.; and (iii) The third level consists of elementary concepts as (key, value) pairs referred to as *Semantic Variable (sv) and semantic Molecule (sm)* tuples. Concepts within each ST can be described by same (sv, st) tuples set.

In order to represent these guiding principles in a closed form, we formulate a *Mathematical Foundational Model* for NSM.

Consider

*Semantic Class (SC)* = $\{sc_i \mid i \in [1,n1]\}$
*Semantic Template (ST)* = $\{st_i \mid i \in [1,n2]\}$
*Relationship* $(R^{sc/st})$ = $x, \{Y\} \mid x \in SC, Y \subset ST$

Concepts within each ST, can be represented with same elementary concepts. The entailment of ST into a set of SV (as keys) and SM (as values) tuples is performed manually for the initial set.

*Semantic Variable (SV)* = $\{sv_i \mid i \in [1, n3]\}$
*Semantic Molecule (SM)* = $\{sm_i \mid i \in [1, n4]\}$
*Entailment (E)* = $\{(e_i \mid i \in [1, n5]\}$
*where* $e_i$ = $\{(sv_i, sm_j) \mid i \in [1, n3]$ & $j \in [1, n4]\}$
$\forall\ sv_i \in SV$ & $\forall\ sm_j \subset SM$
*Relationship* $(R^{st/e})$ = $x, \{y\} \mid x \in ST$ & $y \in E$

Since the scope of ideographic conversion in our work is limited to nouns and complex verbs (which excludes linking verbs like *'is'*, *'am'* etc.), we build an end-to-end pipeline and tokenize the sentence into its components and conduct parts of speech (POS) classification to simplify them into semantic universals as follows.

*Sentence (S)* : $\{\ w_i^{all} \mid i \in [1, n]\ \}$
*Ideographic Words (IW)* : $\{w_j^{id} \mid j \in [1, n']\}$
*IW* $\subseteq S$ ; $\forall\ POS(w_j^{id}) \in \{Noun, Verb\}$
$\forall\ (w_j^{id}) \in (sc_k, st_l)$
$\forall\ (st_l) \in \{(sv_1, sm_1), (sv_2, sm_2), ..., (sv_m, sm_m)\}$

We illustrate some examples, to further clarify the process of semantic simplification within *Humans* (SC) and *Human Relationships* (ST).

$Mother \equiv (Path, P)(Gender, F)$
$Nephew \equiv (Path, S_i, C)(Gender, M)$
$Grandfather \equiv (Path, P; P)(Gender, M)$
*(M-Male, F-Female, P-Parent, $S_i$-Sibling, C-Child)*

Having the mathematical foundational model in place, the next steps is to build an initial ontology for concepts (nouns and verbs, present in our initial dataset) into hierarchical semantic structure.

**Building Initial Ontology**

It has been observed that *Semantic Relatedness (SR)* can be most appropriately evaluated via dense vector representations (Taieb et al., 2020). We have evaluated shallow, context independent distributed vector representations Word2Vec and FastText along with deep transformer based context dependent embedding BERT and ELMO (Xue et al., 2021; Devlin et al., 2018; Peters et al., 2018). We use these embeddings to group together semantic clusters, applying a flat (non hierarchical) clustering algorithms (centroid based and density based) along with hierarchical algorithms (agglomerative and graph based) (Zhang et al., 1997; Baker and McCallum, 1998; Comito et al., 2019).

As illustrated in Table 2, we use combinations of vector representations and two-level clustering techniques to discover semantic clusters in our initial dataset. We initially find semantic classes (SC) as first-level semantic concepts, and then within each SC we explore semantic templates as next-level semantic structures. We evaluate the results by human validation and observe that using a combination of BERT (for embedding) and BIRCH (as clustering approach) give the most appropriate results.

Further breakdown of semantic templates into pairs of semantic variables and molecules is accomplished by careful linguistic analysis of each concept manually. We have collaborated with the domain experts for this work, with inputs from our end users. (Appendix C for details)

### 4.2 Design of Neural Structure

While our initial ontology contains approximately 1550 concepts (1100 noun entities and 450 verb entities), we design a neural method to systematically entail the OOV concepts into the prescribed mathematical model to ensure that our model is universal in use.

The recent evolution of conversational Artificial Intelligence (AI) specifically LLMs, it is ob-

| | W2V | FT | ELMO | BERT |
|---|---|---|---|---|
| **K-Means** | 0.78 | 0.84 | 0.89 | 0.92 |
| **DBSCAN** | 0.79 | 0.78 | 0.90 | 0.91 |
| **BIRCH** | 0.81 | 0.79 | 0.93 | **0.94** |
| **Agglomerative** | 0.88 | 0.89 | 0.72 | 0.93 |

Table 2: Semantic clustering approaches and findings. Human evaluation is normalised on 0-1 scale with 1 being best score. (W2V - Word2Vec, FT- FastText)

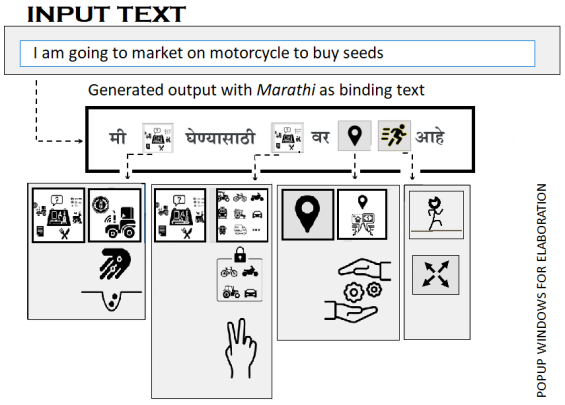



Figure 2: Sample output (binding text in Marathi).

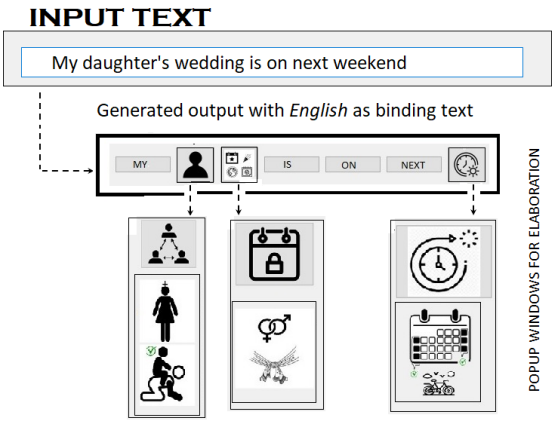



Figure 3: Sample output (binding text in English).

served that LLMs become very instrumental in reasoning and inferencing use cases. LLMs require natural language stimuli referred to as prompt as input to pass instructions to LLM to get respective output (Liu and Chilton, 2022; Reynolds and McDonell, 2021; Bach et al., 2022). Ideally the prompt (P) should contain a Context (C) such as background information of the task, the instructions (I) such as directions, constraints and process to follow and examples (E) as the few shot learning methodology to teach the objective of the exercise $(Prompt(P) \equiv LLM\ (C, I, E))$.

The context is the most crucial component in the prompt, to achieve appropriate output from LLM encapsulating intent along with the environment in consideration. We supply the context in the form of conversational sequence using in-context learning method of Tree Of Thoughts (ToT) (Yao et al., 2024). Instructions are provided in forms of rules and deterministic process which needs to be followed, along with the format (and considerations) of output. The examples provided within the prompt strengthens the inferencing behavior. (Appendix E for details)

### 4.3 Design for User Experience

The final purpose of our work is to assist digital enablement of semi-literates, using intuitive user interface hence UI design becomes very critical for the success of our work. We focus on simplicity of interface elements including input controls (enabling local language support as binding text) and navigational components (including clickable features to illustrate hierarchy).

We refer to *The Noun Project (NP)*[1] for the identification/design of ideographs for our work. NP is a crowd-sourced collection of 3 million icons created by designers from 120+ countries. Each element within semantic class, semantic template and semantic molecule maps to an unique ideograph. Semantic variables are not displayed and kept implicit (based on user feedback). For ideographic selection, we take feedback from our participants to help with the choice of most appropriate icon based on their social and cultural ecosystem (further illustrated in Appendix B).

[1]https://thenounproject.com

## 5 Implementation

We conduct our experiment using Python 3.10.1, PyQt 5 5.15.1, BERT, GPT 3.5 Turbo 16k and Langchain (Devlin et al., 2018; Mavroudis, 2024).

### 5.1 Pre-Processing

We perform text cleaning involving removing irrelevant or unwanted elements from the text, such as special characters, punctuation, etc. followed by lower casing, isolating numbers, Part-of-Speech Tagging and lemmatization. We use *nltk package* for most pre-processing tasks which leverages a predefined set of grammatical rules, a dictionary of words/ punctuation/ identifiers, and for the POS tags, it uses Penn Treebank POS tag set. We perform complexity analysis (using readability scores as described in methodology) on nouns and verbs and identify complex words to be ideographed in downstream processes.

### 5.2 Semantic Decomposition

Each complex token identified is simplified by breaking up of complex concepts into elementary concepts (SC, ST and (SV, SM) tuples). As described in the methodology, we leverage the initial ontology to identify the tokens and use the prebuilt hierarchical semantic decomposition for these

concepts.

For concepts not in initial ontology (which we observe only for rare scenarios, within the same user group) we leverage the power of Gen-AI to entail the concepts into elementary concepts as already illustrated in earlier sections.

Benchmarking results indicate that GPT surpasses other LLMs in inference tasks (Guha et al., 2024; Liu et al., 2023). Once the new concepts decompose into hierarchical semantic schema as required, we also add the same into our ontology for future reference. The binding text is translated and rearranged to native language so that the final message uses the semantics and syntax of end user's native language. We use off the shelf "Google Translate" (via *googletrans* library in python) for translation to native languages, hence the support for native languages is restricted as supported by Google Translate. (See Appendix F for details)

### 5.3 Presentation Framework

NIM is displayed as a sequence of ideographic semantic classes ($sc_i$) with binding text in the initial view. Semantic decomposition into semantic template and further into semantic variable, semantic molecule pairs appear when the corresponding $sc$ icon is clicked. The purpose of keeping semantic details ($st$, $sm$, $sv$) information clickable is to keep the final multimodal sentence in a single view (to avoid distraction by not passing too much of information together) and display hierarchical breakdown of information as and when requested.

Figure 2 and 3 illustrates sample output from our system using Marathi and English as binding text. The syntax of sentence (and hence the location of ideographs) changes with the choice of binding text. The boxes marked with incoming arrows, appear as popups when the corresponding icons are clicked.

## 6 Results and Analysis

To ensure a robust real-world validation of NIM, we collaborate with a separate cohort of 200 participants, distinct from those involved in the design process to minimize bias and provided compensation for their participation. The testing procedure spawned over a period of one week where each day the participants interpret a set of sentences, and we survey their response on multiple parameters.

There is actually no similar work which combines bare minimum elementary text (for syntax) with symbols doing heavy lifting for semantics. Most existing work are either all symbols or text or emoticons (for emotional emphasis). Some latest VLM work is on the other side very non deterministic and does not contain to the boundaries of culture and semantic hierarchy.

Our evaluation approach broadly encompasses three dimensions: **cognitive accessibility and user engagement**, assessing comprehensibility, learnability, and user appeal; **semantic robustness**, evaluating ideograph effectiveness, generalization to out of vocabulary concepts, and the adequacy of concept inventory; **design rationale**, using ablation studies to validate the contribution of individual components. These dimensions provide critical insights into the system's capabilities, detailed in the following sections.

### 6.1 Comprehensibility and Learnability

We use the method illustrated in §3.1 for empirically validating semantic comprehensibility and learnability, but with extended participants and metrics.

We consider 50 sentences stochastically selected from the dataset (Sharma et al., 2021) and translate them to NIM. We distribute these transformed sentences into 5 sets (10 sentences each) and share each set every day for a period of 5 days, with our participants. We ask the participants to share their interpretation every day (verbally or written), and candid feedback is given to them on their response. Not to mention that that every participant is also run through a basic enablement session on NIM before they start.

We use statistical harmonic mean metric METEOR (Banerjee and Lavie, 2005), along with sentence transformers (S-BERT (Reimers and Gurevych, 2019), MPNet (Song et al., 2020) and all-MiniLM-L6 (Face, 2022)) for semantic textual similarity (STS) of human interpretation (with baseline sentences). For learnability, we use normalized learning rate with reward for reaching the threshold faster, reflecting their potential for practical use and impact.

Table 3 summarizes the results (mean values) on all metrics over a period of 5 days and Table 4 illustrates the learning rates which is the differential comprehensibility rate over 5 days period (explained in §3.1).

The empirical results as illustrated in Table 3 and Table 4 indicate that, comprehensibility improves with time. We also observe that NIM reaches a decent value of more than 80% by day 5, reflect-

| D | Meteor | | SBERT | | MPNet | | MiniLM | |
|---|---|---|---|---|---|---|---|---|
| | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ |
| #1 | 0.57 | 0.02 | 0.62 | 0.03 | 0.62 | 0.03 | 0.63 | 0.03 |
| #2 | 0.62 | 0.03 | 0.67 | 0.01 | 0.67 | 0.01 | 0.72 | 0.04 |
| #3 | 0.65 | 0.03 | 0.74 | 0.03 | 0.71 | 0.03 | 0.74 | 0.03 |
| #4 | 0.72 | 0.03 | 0.84 | 0.02 | 0.81 | 0.03 | 0.81 | 0.03 |
| #5 | 0.81 | 0.03 | 0.84 | 0.01 | 0.86 | 0.04 | 0.82 | 0.05 |

Table 3: Empirical Evaluation (0-1 scale) on Meteor, S-BERT, MPNet and miniLM over a period of 5 days(D).

ing good learning rates for our end users. Comparing with benchmarking studies we observe an improvement of 1.9 X in comprehensibility and 8.5 X in learnability. While it may not be appropriate to directly compare with SOTA due to variety of end users, and purpose of other systems, we have shown these results to illustrate the value of NIM in semi-literate context.

### 6.2 User Satisfaction

We have already discovered that the biggest challenge for semi-literates not to use digital communication methods is the lack of engaging content along with non-intuitive user experience. We conduct a qualitative survey each day on 4 parameters namely (i) Expressiveness: ability to conceptualize the semantic characteristics of the inherent concept, (ii) User Experience (UX): ease and convenience for users to interact and understand, (iii) Intention to Reuse: measure of user satisfaction and perceived usefulness and (iv) Interest: measure the value, relevance and appeal to the audience. As illustrated in Table 5, the results (mean values) clearly indicate that the engagement shows a consistent improvement across all measured metrics.

### 6.3 Ideograph Effectiveness

For ideographic effectiveness, we make use of Multiple Index Approach (MIA) as prescribed by European Telecommunications Standard Institute (ETSI) on the principles of CCITT Recommendations (https://www.itu.int/rec/T REC-E.121, 1996; Böcker, 1996). MIA recommends the design of questionnaire on five parameters namely, Valid associations/ Hit-Rate (HR), Invalid associations/ False Alarm Rate (FAR), Missing Associations (MA), Confidence of association/ Subjective Certainty (SC) and Relevance of association/ Subjective Suitability (SS). We conduct test(s) using a questionnaire on the last day of the evaluation when the participants are familiar with the system and can evaluate based on their overall experience over last 5 days. As illustrated in Table 6 the results clearly indicate that the ideographs used have high degree of certainty, relevance, confidence and lower misinterpretation.

| | Meteor | SBERT | MPNet | MiniLM |
|---|---|---|---|---|
| **LCR** ↑ | 0.381 | 0.331 | 0.369 | 0.273 |

Table 4: Learning Curve Rate (LCR) evaluation on various learning metric (similar to Table 1)

| D | Exp | | EOU | | ITR | | Int | |
|---|---|---|---|---|---|---|---|---|
| | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ |
| #1 | 6.30 | 1.56 | 6.22 | 1.40 | 6.36 | 1.32 | 6.35 | 2.10 |
| #2 | 6.35 | 1.29 | 6.42 | 1.32 | 6.43 | 1.26 | 6.79 | 2.35 |
| #3 | 7.67 | 1.28 | 7.63 | 1.30 | 7.69 | 1.30 | 7.48 | 1.42 |
| #4 | 8.25 | 1.48 | 8.08 | 1.19 | 8.08 | 1.57 | 8.36 | 1.45 |
| #5 | 8.21 | 1.45 | 8.35 | 1.45 | 8.37 | 1.43 | 8.40 | 1.38 |

Table 5: Empirical evaluation (0-10 scale) on Expressiveness (Exp), Ease of Use (EOU), Intention to re-use (ITR) and being of interest (Int) over a period of 5 days.

### 6.4 Effects of Generative Methods on New Concepts

While we observe that size and volume of concepts and ideographs follow Zipf's law with a logarithmic growth trajectory and becomes stable, however our design enables handling OOV concepts using LLMs. Our prompts consist of Context (C) which contains the brief summary of the symbolic structure (design philosophy), instructions (I) to keep the response precise and categorically respond without elaboration, along with a set of Examples (E) applicable for the particular context. We use Tree-Of-Thoughts (TOT) as an inferencing style for prompt design, in order to structures the reasoning process of LLM into a tree-like hierarchy of thoughts, thus incorporating multistep reasoning, creative exploration, and optimization.

We leverage GPT 3.5 and use the LLMs incrementally in three steps, first we prompt find Semantic Class, next we use the response and further extract semantic template followed by semantic decomposition into semantic variables and molecules. At each step the prompt ingredients (C, I and E) are carefully articulated and generalized (iteratively) to get the most optimized response. We validate our method, on a set of 200 concepts (from our initial ontology, 150 nouns and 50 verbs; we don't use

| HR ↑ | FAR ↓ | MA ↓ | SC ↑ | SS ↑ |
|---|---|---|---|---|
| 0.89 | 0.07 | 0.05 | 0.86 | 0.83 |

Table 6: Ideographic effectiveness using MIA.

| Meteor | | SBERT | | MPNet | | MiniLM | |
|---|---|---|---|---|---|---|---|
| $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ | $\mu$ | $\sigma$ |
| 0.62 | 0.08 | 0.64 | 0.07 | 0.63 | 0.07 | 0.62 | 0.09 |

Table 7: Ablation study for impact of binding text.

these 200 concepts as examples in prompts), and observe an accuracy of 96% for identification of SC and ST. For identification of SV, SM tuples we achieve 90% accuracy. We also tried with other LLM models (Gemini, Claude 3.5, Llama) and frameworks like DSPy (Khattab et al., 2023), and most models worked in close range.

Given that the Natural Semantic Metalanguage (NSM) theory has been empirically validated as a universal framework and that LLMs excel in the semantic decomposition of novel concepts across diverse cultures, languages, domains, and geographies, it is reasonable to conclude that the proposed metalanguage (NIM) can also demonstrate near-universal applicability.

### 6.5 Inventory Adequacy

Appendix D illustrates the hierchy and size of our ontology to fewer than 500 ideographs to enhance learnability and facilitate efficient comprehension. We also observe that languages often follow power law distribution, i.e., small number of words are used very frequently, while the majority of words are used rarely. We illustrate this phenomena in Figure 4 using a logarithmic scale on the Y-axis to capture the exponential growth of picturable words during early exploration while highlighting the stabilization phase.

### 6.6 Ablation Studies

To validate our design choices and evaluate the contributions of multimodal components (combining ideographs with binding text), we conduct ablation study to assess the impact of removing the binding text from the final output, thereby transitioning to a purely ideographic form of communication. These tests were performed on Day 6 of the testing phase with the same participant group. The results, presented in Table 7, reveal a significant drop in comprehensibility (around 24%) accompanied by an increase in standard deviation (w.r.t. Table 3 Day 5 numbers) . This decline is attributed to the universal arrangement of icons, which lacked context from the participants' native languages. Notably, these observations were made after participants had gained familiarity with the script and developed a reasonable level of competency. The findings underscore the critical role of binding text in reducing ambiguity and enhancing comprehensibility.

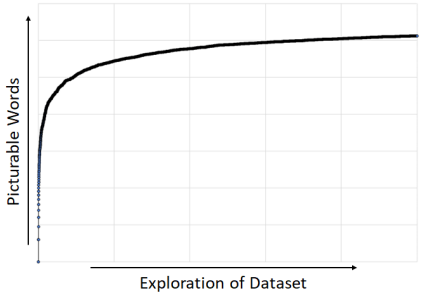



Figure 4: Growth of picturable words with exploration of dataset.

## 7 Conclusion and Future Work

"NIM" facilitates digital literacy for semi-literate population across linguistic and geographic boundaries. It embodies the principles of Neuro-symbolic AI combining the generalization strengths of neural networks (LLM's) with structured reasoning from symbolic logic (leveraging the linguistic theory of NSM). Our design of being multimodal, helps with simpler, highly comprehensible, easy to learn, expressive and engaging output on the interactive digital platforms. Our design is extensible for new domains and languages without much incremental effort.

Our work adopts a participatory, stakeholder-driven design grounded in real-world data from semi-literate users. Using short message datasets and human-in-the-loop evaluation, we co-develop a concept library and communication system that addresses practical digital barriers with contextual ideographs, ensuring both societal relevance and technical robustness.

We envision iterative co-design with semi-literate participants across multiple geographies as a critical direction for future work. Our long-term objective is to open-source this framework, thereby enabling culturally relevant adaptations and fostering broader community-driven innovation. As acknowledged in our Limitations section, a universal set of ideographs may not adequately address the needs of diverse communities; hence, we aim for our methodology to support the creation of localized variants tailored to specific semi-literate or non-literate populations. Furthermore, we plan to develop a stable, production-ready, and robust version of the platform that can be openly shared to encourage collaboration, scalability, and adaptation across diverse regions.

## Limitations

This study is inevitably shaped by the human centric collaborative methodology with a particular social, cultural, and professional backgrounds of both participants and researchers. This would have inevitably influenced the processes of problem framing, ontology construction, and ideograph selection. We do acknowledge the evolving nature of visual literacy and the risks of misinterpretation. We acknowledge that ideographs carry culturally specific resonances and lose its nuance if designed for cultural neutrality. While our approach sought to mitigate these risks through contextual disambiguation, by rarely showing a symbol in isolation and use compositional layouts and semantic groupings to reinforce meaning and reduce misinterpretation. However we cannot ignore the fact that misinterpretation remains an inherent characteristic of pictographic communication.

It is also important to emphasize that our system is not conceived as a replacement for speech or text, but rather as a complementary modality, designed to augment communication in contexts where conventional forms are inaccessible or inadequate.

## Ethical Considerations

Our work encompasses human-centered NLP, and we have strongly collaborated with a group of participants in the entire process. The selection of participants was random and we started with a list of voluntary people, from which final group was picked randomly. We have followed the distribution across demography (urban/rural, age and gender) from earlier work which worked on similar problem (Sharma et al., 2021). Every participant was paid for the job (at par with local rates).

The study adhered to stringent ethical guidelines to ensure participant privacy and data security. Participation was voluntary and participants were made aware of the overall process, along with how the data will be used. Informed consent was obtained in writing from all participants, with no collection, storage, or processing of personally identifiable information (PII) or clinical data. The research complied with the "Guidelines for Ethical Considerations in Social Research & Evaluation in India" (CMS, 2020), and a self-administered Ethics Sensitivity Test based on these guidelines yielded a "Very Good" grade. Safety measures included conducting all tests in closed-door settings to safeguard participant anonymity.

## References


Stephen H Bach, Victor Sanh, Zheng-Xin Yong, Albert Webson, Colin Raffel, Nihal V Nayak, Abheesht Sharma, Taewoon Kim, M Saiful Bari, Thibault Fevry, and 1 others. 2022. Promptsource: An integrated development environment and repository for natural language prompts. *arXiv preprint arXiv:2202.01279*.

L Douglas Baker and Andrew Kachites McCallum. 1998. Distributional clustering of words for text classification. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 96–103.

Satanjeev Banerjee and Alon Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*.

Susana Bautista, Raquel Hervás, Agustín Hernández-Gil, Carlos Martínez-Díaz, Sergio Pascua, and Pablo Gervás. 2017. Aratraductor: text to pictogram translation using natural language processing techniques. In *Proceedings of the XVIII International Conference on Human Computer Interaction*.

Martin Böcker. 1996. A multiple index approach for the evaluation of pictograms and icons. *Computer Standards & Interfaces*.

David V Canter. 1996. Way-finding and signposting: Penance or prosthesis? Dartmouth Publishing Company.

Carmela Comito, Agostino Forestiero, and Clara Pizzuti. 2019. Word embedding based clustering to detect topics in social media. In *IEEE/WIC/ACM International Conference on Web Intelligence*, pages 192–199.

Shakila Dada, Alice Huguet, and Juan Bornman. 2013. The iconicity of picture communication symbols for children with english additional language and mild intellectual disability. *Augmentative and Alternative Communication*, 29(4):360–373.

Rodrigo de Salvo Braz, Roxana Girju, Vasin Punyakanok, Dan Roth, and Mark Sammons. 2006. An inference model for semantic entailment in natural language. In *Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment: First PASCAL Machine Learning Challenges Workshop, MLCW 2005, Southampton, UK, April 11-13, 2005, Revised Selected Papers*, pages 261–286. Springer.

Ioannis Deliyannis, Christina Simpsiri, and Plateia Tsirigoti. 2008. Interactive multimedia learning for children with communication difficulties using the makaton method. In *International Conference on Information Communication Technologies in Education*, pages 10–12.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Hugging Face. 2022. Sentence-transformers.

James N Farr, James J Jenkins, and Donald G Paterson. 1951. Simplification of flesch reading ease formula. *Journal of applied psychology*, 35(5):333.

Cliff Goddard. 2006. Natural semantic metalanguage.

Cliff Goddard and Goddard. 2018. *Minimal English for a global world*.

Philippe Grange and Mubashir Mulla. 2015. Learning the "learning curve". *Surgery*, 157(1):8–9.

Tovi Grossman, George Fitzmaurice, and Ramtin Attar. 2009. A survey of software learnability: metrics, methodologies and guidelines. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 649–658.

Neel Guha, Julian Nyarko, Daniel Ho, Christopher Ré, Adam Chilton, Alex Chohlas-Wood, Austin Peters, Brandon Waldon, Daniel Rockmore, Diego Zambrano, and 1 others. 2024. Legalbench: A collaboratively built benchmark for measuring legal reasoning in large language models. *Advances in Neural Information Processing Systems*, 36.

Robert Gunning. 1969. The fog index after twenty years. *Journal of Business Communication*.

Raquel Hervás, Susana Bautista, Gonzalo Méndez, Paloma Galván, and Pablo Gervás. 2020. Predictive composition of pictogram messages for users with autism. *Journal of Ambient Intelligence and Humanized Computing*, 11:5649–5664.

Pascal Hitzler, Monireh Ebrahimi, Md Kamruzzaman Sarker, and Daria Stepanova. 2024. Neuro-symbolic ai and the semantic web.

Pascal Hitzler and Md Kamruzzaman Sarker. 2022. Neuro-symbolic artificial intelligence: The state of the art.

https://www.itu.int/rec/T REC-E.121. 1996. Ccitt recomendations e.121.

Gary Imai. 2005. Gestures: Body language and nonverbal communication. *Retrieved Oct*.

Omar Khattab, Arnav Singhvi, Paridhi Maheshwari, Zhiyuan Zhang, Keshav Santhanam, Sri Vardhamanan, Saiful Haq, Ashutosh Sharma, Thomas T Joshi, Hanna Moazam, and 1 others. 2023. Dspy: Compiling declarative language model calls into self-improving pipelines. *arXiv preprint arXiv:2310.03714*.

Adam Kilgarriff, Pavel Rychly, Pavel Smrz, and David Tugwell. 2004. Itri-04-08 the sketch engine. *Information Technology*, 105(116):105–116.

J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel.

Maria Kunilovskaya and Marina Koviazina. 2017. Sketch engine: A toolbox for linguistic discovery. *Jazykovedny Casopis*, 68(3):503.

Vivian Liu and Lydia B Chilton. 2022. Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–23.

Xiao Liu, Yanan Zheng, Zhengxiao Du, Ming Ding, Yujie Qian, Zhilin Yang, and Jie Tang. 2023. Gpt understands, too. *AI Open*.

Kenneth N Lodding. 1983. Iconic interfacing. *IEEE Computer graphics and applications*.

Martin C Maguire. 1985. A review of human factors guidelines and techniques for the design of graphical human-computer interfaces. *Computers & graphics*.

Vasilios Mavroudis. 2024. Langchain.

Matthew E Peters, Mark Neumann, Luke Zettlemoyer, and Wen-tau Yih. 2018. Dissecting contextual word embeddings: Architecture and representation. *arXiv preprint arXiv:1808.08949*.

Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, and 1 others. 2018. Improving language understanding by generative pre-training.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Laria Reynolds and Kyle McDonell. 2021. Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–7.

Philip Seargeant. 2019. *The Emoji Revolution: How technology is shaping the future of communication*. Cambridge University Press.

Leen Sevens, Vincent Vandeghinste, Ineke Schuurman, and Frank Van Eynde. 2015. Natural language generation from pictographs. In *Proceedings of the 15th European Workshop on Natural Language Generation (ENLG)*, pages 71–75.

Prawaal Sharma, Navneet Goyal, and Poonam Goyal. 2023. Multimodal semantographic metalanguage (msm): A novel methodology for digital enablement of semi-literates. In *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing*, pages 844–851.

Prawaal Sharma, Navneet Goyal, and MR Vinay. 2021. Semi-literate texting (slt): Survey based text message dataset from digitally semi-literate users in india. *Data in Brief*.

Amit Sheth, Kaushik Roy, and Manas Gaur. 2023. Neurosymbolic artificial intelligence (why, what, and how). *IEEE Intelligent Systems*, 38(3):56–62.

WS Shirreffs. 1992. Maps as communication graphics. *The Cartographic Journal*, 29(1):35–41.

Waralak V Siricharoen. 2013. Infographics: the new communication tools in digital age. In *The international conference on e-technologies and business on the web (ebw2013)*, volume 169174.

Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2020. Mpnet: Masked and permuted pre-training for language understanding. *Advances in Neural Information Processing Systems*, 33:16857–16867.

Mohamed Ali Hadj Taieb, Torsten Zesch, and Mohamed Ben Aouicha. 2020. A survey of semantic relatedness evaluation datasets and procedures. *Artificial Intelligence Review*.

Kristen Tatti. 2016. New iconji language for the symbol-minded-bizwest. *BizWest*.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, and 1 others. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Andries Van Dam. 1984. Computer software for graphics. *Scientific American*.

Anna Wierzbicka. 1996. *Semantics: Primes and universals: Primes and universals*. Oxford University Press, UK.

Robert T Williams. 1972. A table for rapid determination of revised dale-chall readability scores. *The Reading Teacher*.

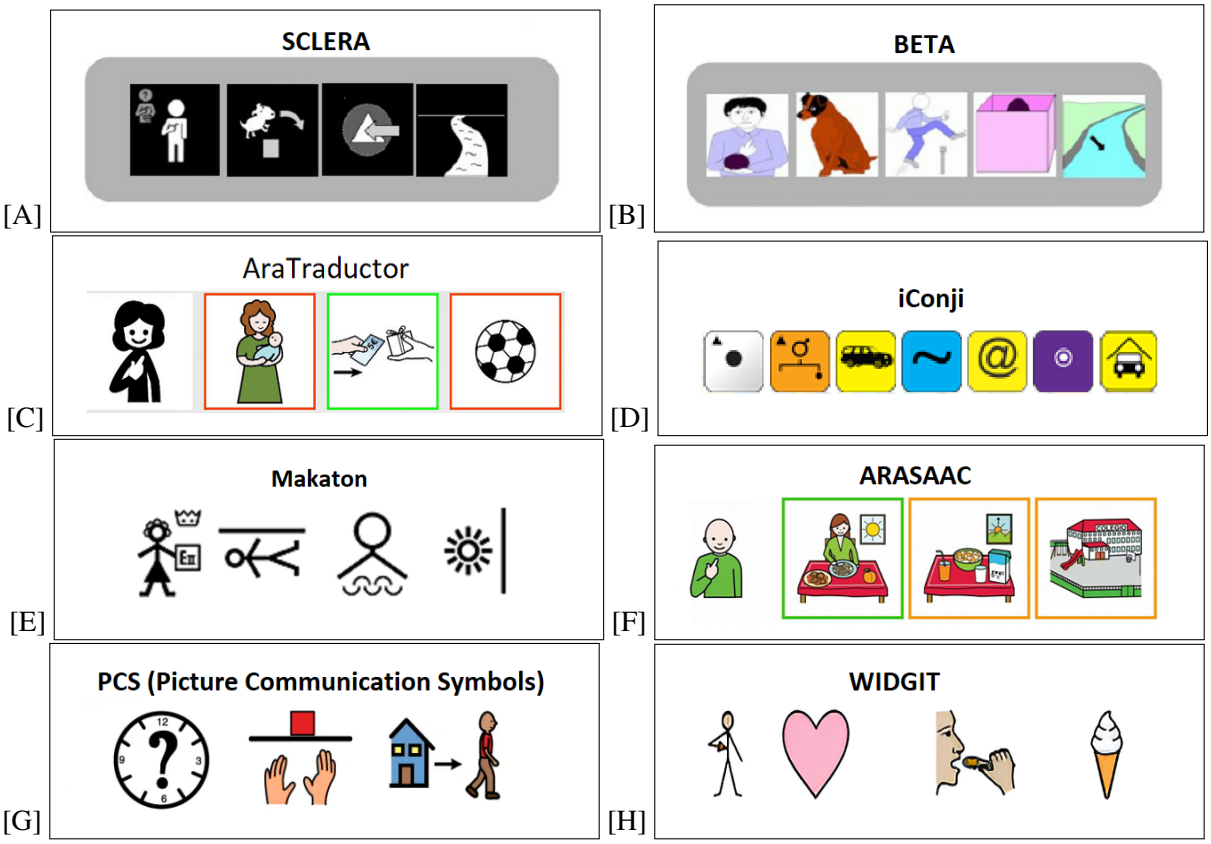
Xingsi Xue, Haolin Wang, Jie Zhang, Yikun Huang, Mengting Li, and Hai Zhu. 2021. Matching transportation ontologies with word2vec and alignment extraction algorithm. *Journal of Advanced Transportation*.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. 2024. Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*, 36.

Tian Zhang, Raghu Ramakrishnan, and Miron Livny. 1997. Birch: A new data clustering algorithm and its applications. *Data Mining and Knowledge Discovery*.

## A Other Methods

Illustrative examples of existing visual methods of communication




| | Method | Illustrated Example | Approximate ideograph count |
|---|---|---|---|
| A | SCLERA | My dog jumped in the river. | 3,500 |
| B | BETA | My dog jumped in the river. | 11,500 |
| C | AraTraductor | My mum bought bought the soccer ball. | 8,500 |
| D | iConji | My father's SUV is at the garage. | 11,00 |
| E | Makaton | My queen died peacefully yesterday. | 11,000 |
| F | ARASAAC | I eat breakfast at school. | 12,000 |
| G | PCS | When do you want to go. | 37,000 |
| H | Widget | I like to eat ice cream. | 15,000 |

(Sevens et al., 2015; Bautista et al., 2017; Tatti, 2016; Deliyannis et al., 2008; Hervás et al., 2020; Dada et al., 2013)

## B Human Engagement

Metadata (on multiple parameters) for the users engaged during assessment and testing phases.

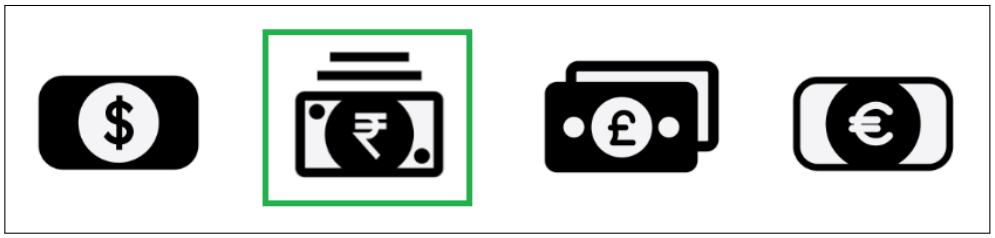
| | Parameter | Initial assessment | Final assessment |
|---|---|---|---|
| 1 | Total participants | 20 | 200 |
| 2 | Demography | Rural - 11; Urban - 9 | Rural - 103; Urban - 97 |
| 3 | Age group | 40-50 years | 40-50 years |
| 4 | Gender | Male - 12; Female - 8 | Male - 123; Female - 77 |
| 5 | Duration | User needs assessment: 1 week<br>Ideographic selection: 1 week | 1 week |

**Selection criteria**

The dataset used in our work (Sharma et al., 2021) is a collection of raw data (using door to door survey) for a very similar problem. Our choice for selecting participants on parameters illustrated in the table above is based on this dataset.
[Age - Mean 36, Mode - 37, Max 65]
[Demography - 48% Urban, 52% Rural]
[Gender - Male 58%, Female 42%]

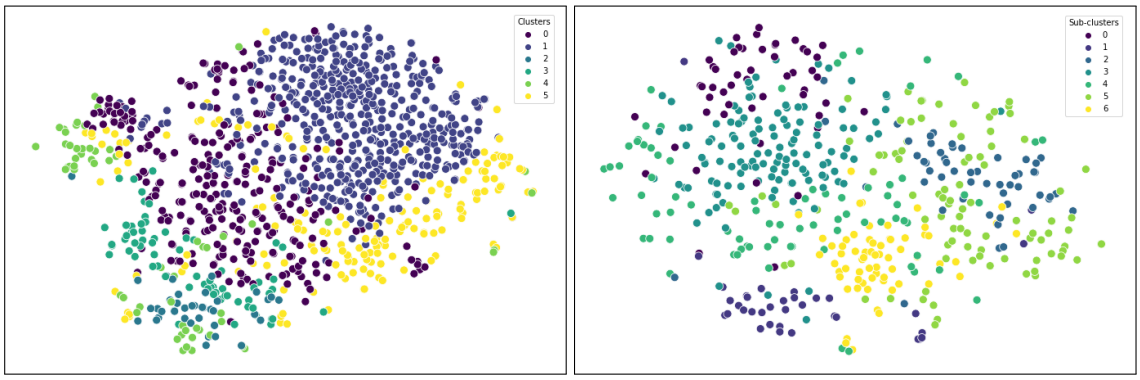
**Ideographic selection criteria based on voting (example)**

*(Candidate ideographs for representing the concept of "currency". The icon in green box is selected as final ideograph based on majority voting.)*



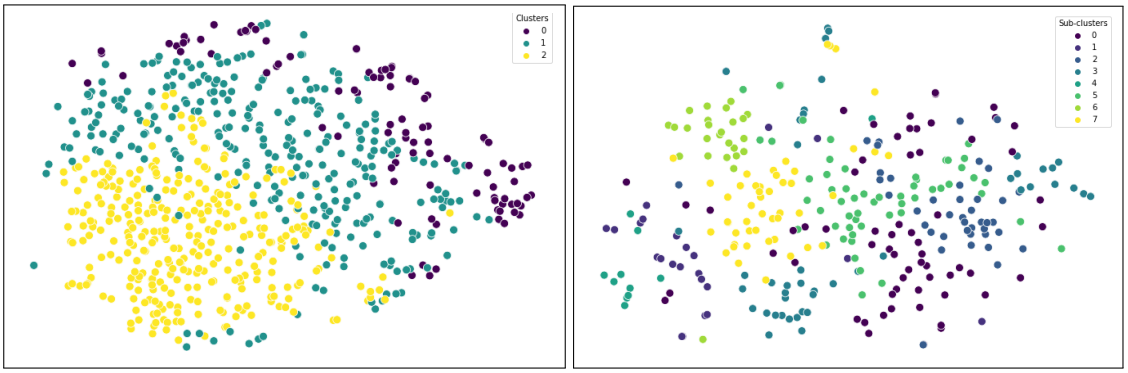
## C Semantic Discovery

**Discovery of Semantic Classes, Semantic Templates, Semantic Variable and Semantic Molecules from the dataset.**
**Step 1:** Initially we split the dataset into nouns and verbs via POS analysis and analyse the concepts separately.
**Step 2:** For nouns using BERT embeddings and BIRCH clustering approach we find 6 distinctive groups (Semantic Classes) at first level. The image below on the left is an visual representation (reduced dimensions) for the same.
**Step 3A:** Within each Semantic Class, we further identify sub clusters which are referred as Semantic Templates in our work. The image on the right is an illustrative example for Semantic Templates within the same Semantic Class.




Semantic Classes (Left) and Semantic Templates for Class #2 (Right) for nouns. Illustrative representation.

**Step 3B:** We repeat the same process for Verbs, and identify 3 Semantic classes and further semantic templates for each semantic class. The figure below is a visual representation for verbs. The image on the right is semantic template identification for cluster 2 on the left.



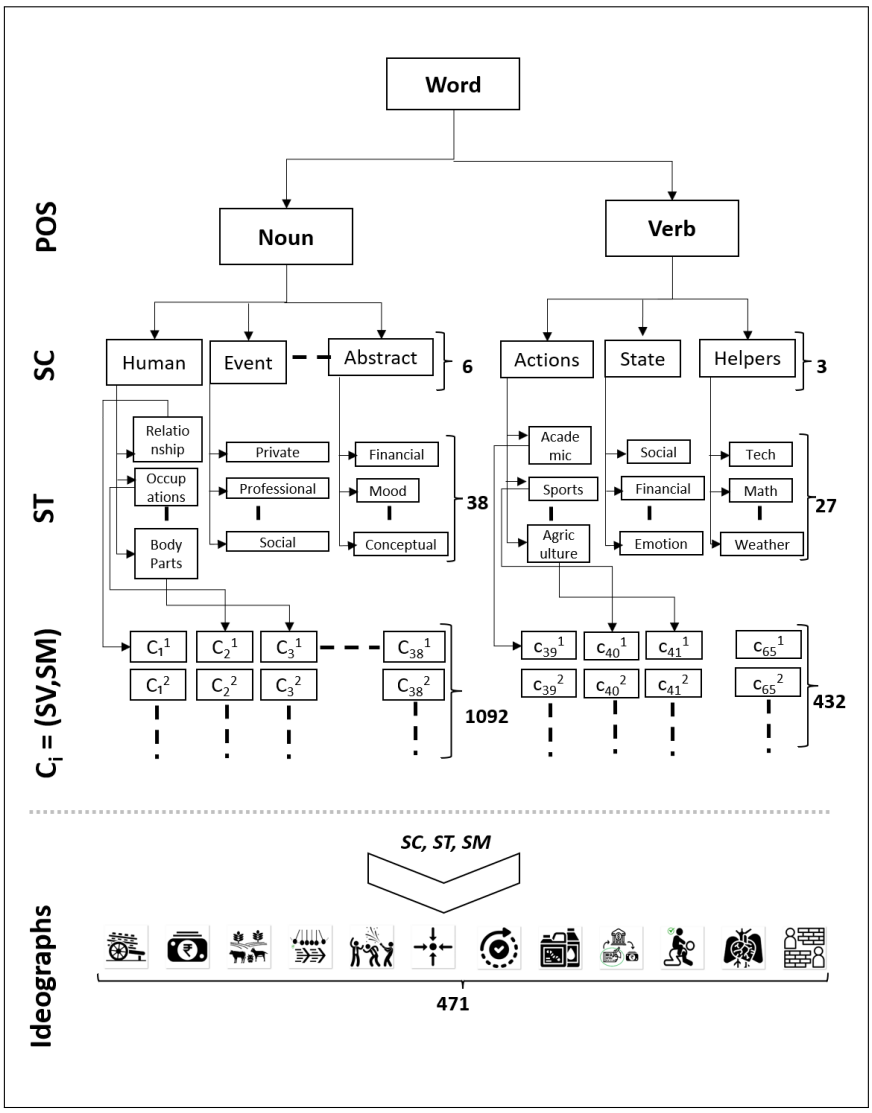

Semantic Classes (Left) and Semantic Templates for Class #2 (Right) for verbs. Illustrative representation.

**Step 4:** Further breakdown into semantic variables and molecules is done manually by careful examination of concepts within the semantic templates. This step is divided into 2 parts (i) Initially the template is generalised/defined as a set of Semantic Variable (SV) and Semantic Molecule (SM) tuples, and (ii) Each concept within the same semantic template is described using the identified (SV, SM) tuple set.

*The entire illustrative count and hierarchy of ontology is illustrated in Appendix D*

## D Ontology

### Hierarchical view




### Tabular View (summary)

| | SC | ST | (sm, sv) | Ideographs |
|---|---|---|---|---|
| Nouns | 6 | 38 | 1092 | 471 |
| Verbs | 3 | 27 | 432 | |

*$C_i$ issued to represent a unique combination of (SV,SM) tuple set for explication of individual concepts.*

# E Prompt Engineering

**Illustrative examples for OOV concepts using Gen-AI**
Consider the following sentence as an illustrative example.
Input Sentence: “I am going to mandi on motorcycle to buy seeds”.

Complex-Words identified - CW1:mandi, CW2:motorcycle, CW3:seeds.

In the example illustrated, CW2 and CW3 are present in the ontology, while CW1 is Out Of Vocabulary (OOV). Hence for CW2 and CW3, we refer ontology and break them down into semantic universals as below.

*CW2 - motorcycle*
*SC : Things*
*ST : Automobile*
*(SV, SM tuples): (Category, Private Transport) (Wheels, Two)*

*CW3 - seeds*
*SC : Things*
*ST : Agro*
*(SV, SM tuples): (Category, Germinate)*

For complex words not in ontology (“Mandi”/CW1 in this example), we use a sequence of prompts using TOT and few-shot learning, where relevant parts of the pre-compiled ontology is passed as context to a LLM and get the recommended response.
Input Prompt (P1) - “Imagine the human annotator has been given the task to hierarchically break down a word into 2 levels. Level 1 considers of broad category of word and is referred as SC, Level 2 considers of narrow category of word and is referred as ST. Now considering the examples as illustrated *«Examples with concepts and categorization of SC and ST come here »*, use your inferencing to find the SC and ST for the word *«Mandi»*”

Leveraging response parsing (json parser) and GPT 3.5 as the LLM the final response is illustrated below.

Output json = “SC”:“Location”,“ST”:“Commercial”

Once the Semantic Template (ST) is identified (“Commercial” in this case), we share multiple examples of other concepts in the same ST, within the prompt and leverage LLM to break this down into SV, SM tuples as illustrated below.

Input Prompt (P2) -“Imagine the human annotator has been given the task to explain the semantic meaning of the word «Mandi», using Key-Value pairs. Keys to be considered should be from the predefined set of values *«add all SV elements for the ST ‘Commercial’ »*. Values considers should be of a predefined set of values *«add all SM elements for the ST ‘Commercial’»*. Each semantic variable can only take limited values from semantic molecule sets as illustrated *«For ‘Commercial’ ST add all semantic variable and molecule combinations here»*. Now considering the examples and constraints as illustrated use your inferencing to find the Key-Value pairs for the word *«Mandi»*. When there are multiple Key-Value pairs, use Key1,Value1, Key2,Value2 in your response.”

The response received from this prompt appears as:
Output json = “Key1”:“Purpose”,“Value1”:“Business”

Putting all of this together the OOV concept *mandi* is represented in the hierarchical structure as below.

*CW1 - mandi*
*SC : Location*
*ST : Commercial*
*(SV, SM) tuples: (Purpose, Business)*

We also tried with other LLM models (Gemini, Claude 3.5, Llama) and frameworks like DSPy (Khattab et al., 2023), and most models worked in close range. Since LLMs are evolving fast we plan to re-validate these models in our future work.

## F Illustrative NIM execution

**Step by step process of input text to multimodal output**

**Text Message (input) in English language for a Marathi end user**:
I am going to market on motorcycle to buy seeds

---

**Complex Word identification and Translation (to Marathi)**:

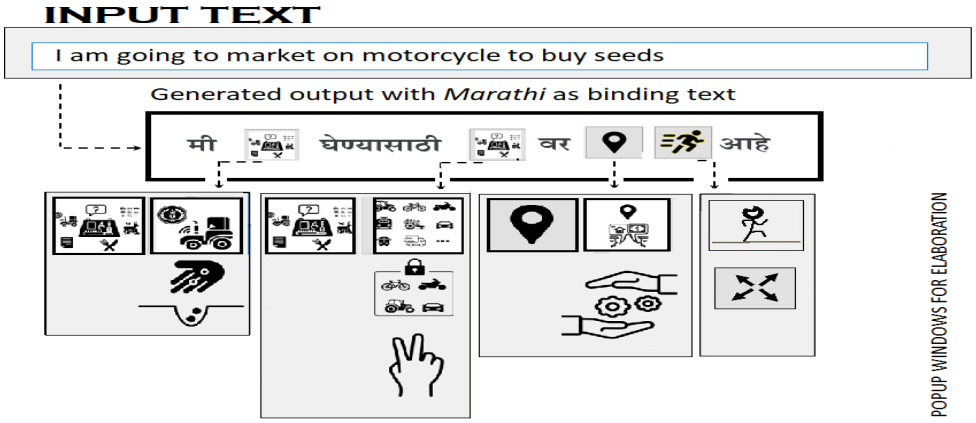
मी SEEDS घेण्यासाठी MOTORCYCLE वर MARKET जात आहे

---

**Entailment of complex words into hierarchical semantic universals**:
<elementalization >
-<cw >SEEDS </ cw >
— <sc >things </ sc >
— <st >agro </ st >
—— <sv >category </ sv ><sm >germinate </ sm >
- <cw >MOTORCYCLE </ cw >
— <sc >things </ sc >
— <st >automobile </ st >
—— <sv >category </ sv ><sm >private transport </ sm >
—— <sv >wheels </ sv ><sm >two </ sm >
- <cw >MARKET </ cw >
— <sc >location </ sc >
— <st >commercial </ st >
—— <sv >purpose </ sv ><sm >business </ sm >
</ elementalization >

---

**Final multimodal message output displayed to end user**:



**Text Message (input) in English language for a Nepali end user**:
There may be a typhoon tomorrow

---

**Complex Word identification and Translation (to Nepali)**:

TOMORROW यहाँ TYPHOON आउन सक्छ

---

**Entailment of complex words into hierarchical semantic universals**:
<elementalization >
- <cw >TOMORROW </ cw >
— <sc >time </ sc >
— <st >temporal </ st >
—— <sv >marker </ sv ><sm >day(+1) </ sm >
- <cw >TYPHOON </ cw >
— <sc >event </ sc >
— <st >event climate </ st >
—— <sv >category </ sv ><sm >wind </ sm >
—— <sv >intensity </ sv ><sm >high </ sm >
</ elementalization >

---

**Final multimodal message output displayed to end user**: