On the Optimality of Kinship Naming: an Information-theoretic Approach

Anonymous ACL submission

Abstract

The structure of naming systems in natural languages hinges on a trade-off between high informativeness and low complexity. Focusing on the domain of kinship naming, we analyze such trade-off while addressing simplifying assumptions of prior work, namely: (i) universal communicative need across languages, and (ii) optimal listeners. To that aim, we collect data from four different languages, and analyze how different communicative needs and variations in the listener model influence the informativeness–complexity trade-off. Adopting a referential game setup from emergent communication, we further show that trade-off optimality is not only theoretically achievable but also emerges empirically in learned communication systems.

1 Introduction

Languages vary in how they structure meaning partitions. For instance, unlike in English, Vietnamese offers different words for *aunt*, depending on whether she is the younger or elder sister of one's mother or father. A body of work suggests that languages do not explore the space of possible semantic partitions freely, as evidenced by constrained and recurrent cross-linguistic patterns (Kemp and Regier, 2012; Regier et al., 2015; Zaslavsky et al., 2018; Kemp et al., 2018; Carr et al., 2020; Chaabouni et al., 2021). Instead, the structure of object naming systems appears to reflect pressures for communicative efficiency and learning. In particular, these studies suggest that languages tend to evolve toward object naming systems that approximate an (often near-)optimal trade-off between informativeness and complexity.

This trade-off can be formalized with constructs from information theory (Shannon, 1948). Informativeness is typically quantified by the amount of information preserved in communication, while complexity measures how concisely a language compresses meaning into words. Using these notions, frameworks define optimal trade-off boundaries: curves along which no system can reduce complexity without increasing information loss.

On the domain of kinship, Kemp and Regier (2012) measure complexity based on shortest semantic description length, and demonstrate that kinship systems lie near an optimal trade-off frontier. However, such descriptions are not grounded on empirical data, and a universal communicative need is assumed across languages–even though cultural differences can be reflected in kinship communicative need (Anand and Regier, 2023). A relevant limitation of their framework is that their measure of complexity is not analytically tractable to derive a closed-form expression for the trade-off curve. In contrast, the Information Bottleneck (IB) framework (Tishby et al., 2000) allows for the derivation of theoretical, closed-form *approximations* of the optimal trade-off frontier.

Here, we apply the IB framework to the case of kinship. Our contributions are as follows. First, inspired by multiclass classification, we adapt the information loss in the IB framework. This allows us to (i) apply the framework to discrete domains, and (ii) deduce the optimal trade-off curve directly from it–which prevents the need to add additional parameters to estimate it empirically, avoiding abnormal situations for some values of those parameters. Second, we present a novel dataset which estimates communicative need in four languages, and we use it to analyze how different assumptions (such as different communicative needs or variations in the listener model) influence the informativeness–complexity trade-off in communication. Third, adopting a referential game setup commonly used in emergent communication (Lazaridou and Baroni, 2020; Lazaridou et al., 2017; Havrylov and Titov, 2017; Chaabouni et al., 2022), we show that our notion of optimality is not only theoretically achievable but also emerges empirically in learned

communication systems.

2 Background

2.1 Kinship Naming

Every society uses language to refer to family members, or *kin*, through a system of lexical items that categorize familial roles (e.g., father and sister). A relevant source of cross-linguistic variation lies in how kinship meaning space is partitioned, or in other words, which family members are considered part of the same semantic category. For example, in English, both maternal and paternal grandmothers fall under the same category (grandmother). In contrast, Vietnamese differentiates both lineage and age, employing distinct terms for maternal versus paternal grandparents, and for older versus younger siblings (Van Luong, 1989).

Kemp and Regier (2012) observe that kinship systems appear to reflect a trade-off between informativeness—here, the ability to distinguish between kin members based on kinship names—and complexity. The latter relies on a symbolic rule system to characterize the meaning of kinship names through logical compositions of primitives (e.g., mother would be *PARENT & OLDER & FEMALE*). Complexity is then quantified as the minimal number of logical rules needed to generate the system. Informativeness is measured as the expected Kullback-Leibler (KL) divergence between intended and inferred referents, averaged over a communicative need distribution.

2.2 Information-Bottleneck framework

Zaslavsky et al. (2018) introduce an information-theoretic framework for quantifying the trade-off between informativeness and complexity in lexical systems. Focusing on the continuous domain of color, they propose that color naming systems approximate optimal trade-offs by compressing perceptual meanings into words in a manner consistent with the Information Bottleneck (IB) principle (Tishby et al., 2000), provided that listeners are optimal in the Bayesian sense.

Concretely, in the IB framework for color naming, a Speaker and Listener communicate about colors $u \in \mathcal{U}$, where $\mathcal{U}$ represents a continuous perceptual space. Upon perceiving a color u, the Speaker selects a meaning m, modeled as a distribution over $\mathcal{U}$, and then generates a name w using encoder $q_s(w|m)$. The Listener decodes the message using:

$$\hat{m}_w(u) = \sum_m \tilde{q}_s(m|w)\, m(u),$$

which defines the *Bayesian-optimal listener*.

Assuming a Bayesian-optimal listener, complexity is quantified as mutual information between the Speaker's meaning variable M and the word variable W:

$$I_{q_s}(M;W) = \sum_{m,w} p_s(m)\, q_s(w|m) \log \frac{q_s(w|m)}{p_s(w)},$$

while informativeness is captured by the mutual information $I_{q_s}(W;U)$, measuring how much information the word conveys about the original object.

Following the IB principle, an optimal trade-off between complexity and informativeness is achieved by minimizing the following objective:

$$\mathcal{F}_\beta[q_s(w|m)] = I_{q_s}(M;W) - \beta I_{q_s}(W;U),$$

where $\beta \geq 1$ is a trade-off parameter that balances compression and informativeness.

Recently, the IB framework has been applied to discrete domains, such as artifact and animal naming (Zaslavsky et al., 2019) and personal pronouns (Zaslavsky et al., 2021). In these settings, $\mathcal{U}$ denotes the set of all possible objects (e.g., animals), and the meaning m represents the speaker's uncertainty in recognizing the target. That is, it captures the probability that the speaker confuses the intended object with another similar one (e.g., a dog versus a wolf).

As a result, one can straightforwardly apply the IB framework to kinship naming. However, this approach has three main drawbacks. First, the framework is designed to capture cognitive constraints such as limited memory or confusability between similar objects. While this perspective is appealing when the space of objects is large (e.g., an effectively infinite color spectrum), it is unclear how to estimate and formally encode these cognitive limitations in the object–representation mapping. Existing instantiations (Zaslavsky et al., 2021) often rely on *ad hoc* choices. Second, the informativeness term $I_{q_s}(W;U)$ does not characterize the full communication process, as it leaves out the impact of the listener. In practice, the IB framework typically assumes a Bayesian-optimal listener and a noiseless communication channel, assumptions that are not always realistic. For instance, when the listener has a hearing impairment or when the

signal is corrupted by channel noise. Finally, the IB framework does not account for the influence of the communicative need distribution.

3 Framework

We propose an adaptation of the IB framework to the problem of kinship naming that addresses the three issues pointed out above. First of all, for domains with only a small number of discrete objects, such as kinship, it is difficult to justify the assumption that a human speaker cannot reliably distinguish between two items (e.g., two family members), as would be the case in color naming due to its continuous nature. In such cases, M and U effectively become identical random variables. Therefore, we eliminate M, and consider the following scenario.

There is a pool $\mathcal{U}$ of kinship relations, associated with a communicative need distribution $p(\cdot)$. A Speaker aims to communicate about a kinship relation $u \sim p(u)$ by selecting a message (or name) $w \in \mathcal{W}$ with encoding probability $q_s(w|u)$. The Speaker sends this message to a Listener, who then attempts to infer the intended kinship using a decoder $q_l(u|w)$.

As illustrated in Figure 1, consider an English-speaking Speaker who wishes to refer to $u = \text{elder brother}$. Due to the structure of English kinship terminology, the most specific term available is $w =$ "brother", and the Speaker deterministically selects it, i.e., $q_s(\text{"brother"}|\text{elder brother}) = 1$. Upon receiving the message, an English-speaking Listener infers that the referent could be either elder brother or younger brother, assigning equal probabilities: $q_l(\text{elder brother}|\text{"brother"}) = q_l(\text{younger brother}|\text{"brother"}) = 0.5$.

3.1 Complexity

Like in the IB framework, we quantify complexity as the amount of information the Speaker compresses through its encoder $q_s(w|u)$. This is measured by the mutual information between the object random variable U and the message random variable W:

$$C = I_{q_s}(U;W) = \sum_{u,w} p(u)\, q_s(w|u) \log \frac{q_s(w|u)}{p_s(w)} \tag{1}$$

where $p_s(w) = \sum_{u \in \mathcal{U}} p(u)\, q_s(w|u)$ is the marginal distribution over messages. Intuitively, this quantity captures the average amount of information that a message w conveys about the intended object u. If the same message is used for every meaning, mutual information is zero. Conversely, if each meaning is encoded with a distinct message, mutual information reaches its maximum value—the entropy $H(U)$. Thus, complexity is bounded by $0 \leq C \leq H(U)$.

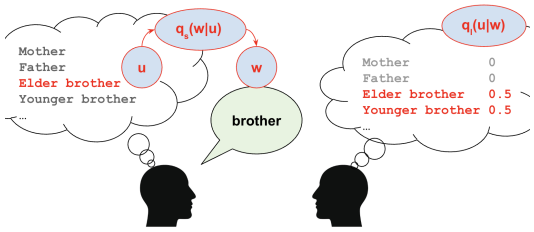

Figure 1: Illustration of two English-speaking agents playing the kinship naming game. The Speaker (left) selects a family member and produces a name. The Listener (right) receives the name and infers which member is being referred to.

3.2 Information Loss

Different from the IB framework, we define information loss to quantify how much information is *not* preserved *throughout the whole communication process*, and is therefore directly related to the errors made by the Listener when picking a referent. We define information loss as the expected cross-entropy between the true referent and the Listener's prediction, i.e., the standard loss function in multi-class classification:

$$L = -\mathbb{E}_{u \sim p}\, \mathbb{E}_{w \sim q_s(\cdot|u)} \log q_l(u|w).$$

Intuitively, the more confident the Listener can be about the intended object u given the message w, the lower the information loss. Conversely, uncertainty in decoding leads to higher loss.

3.3 Optimality

Let $\tilde{q}_s(u|w) \propto q_s(w|u)\, p(u)$ be the Speaker's Bayesian decoder. We prove in Appendix A that:

$$L = H(U) - C + \mathbb{E}_{w \sim p_s}\left[D_{\text{KL}}(\tilde{q}_s \| q_l)\right] \geq H(U) - C \tag{2}$$

A direct result from this identity is that equality holds if and only if $q_l = \tilde{q}_s$, which is also entailed by the IB framework. This implies that, the optimal trade-off is achieved when the Listener's decoder exactly matches the Bayesian decoder of the Speaker. However, it is essential to emphasize that

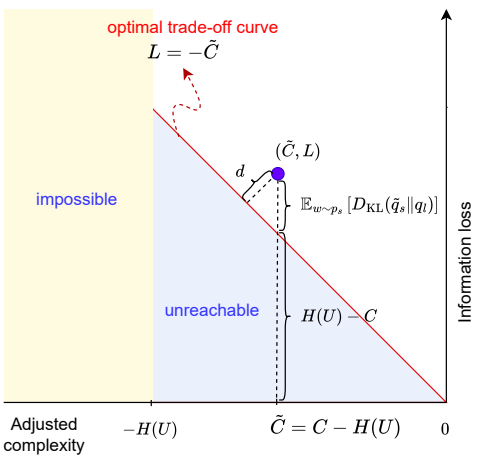

Figure 2: Visualization of the optimal trade-off curve (red line) and the feasible region (white area) encompassing all valid communication systems.

although systems with Bayesian listeners are optimal in this formal sense, they are not necessarily equally effective. Human communication succeeds because it balances informativeness and simplicity, often sacrificing some simplicity to convey sufficient meaning. In contrast, low-information systems can also be considered optimal—due to their high simplicity—but are practically unhelpful because they are too ambiguous. Thus, inline with previous work e.g. Kemp and Regier (2012); Zaslavsky et al. (2018), optimality in this framework must be understood in the context of both information loss and complexity.

Interestingly, Equation 2 allows us to measure Euclidean distance from a given communication system to the curve *analytically* by $d = \frac{L - H(U) + C}{\sqrt{2}}$.

3.4 Need-agnostic Theoretical Optimal Curve

The curve $L = H(U) - C$ described above depends on the communicative need distribution $p(u)$, due to the inclusion of the entropy term $H(U)$. This dependency complicates cross-linguistic comparisons of optimality, as different need distributions induce different optimal curves in the complexity–information loss space To address this issue, prior work—including Kemp and Regier (2012) and Zaslavsky et al. (2018)—assumed a universal, fixed need distribution across languages.

Here we propose a more principled alternative: by defining an adjusted complexity $\tilde{C} = C - H(U) \leq 0$, we transform the optimal curve into the simplified form $L = -\tilde{C}$. This reformulation removes the dependence of the optimal trade-off curve—*but not of information loss or complexity themselves*— on $p(u)$, while preserving Euclidean distance. As a result, it enables clear visualization across languages, regardless of their underlying communicative needs. Notably, the smaller $\tilde{C}$ is, the less complex the system is, and the more *capacity* it has to increase in complexity. When $\tilde{C} = 0$, the system reaches its maximum allowable complexity under the given need distribution.

Figure 2 illustrates the optimal trade-off curve, along with the feasible region in which all valid communication systems must lie.

4 The Emergence of Kinship Naming

In this section, we present kinship emergence from communication, based on the familial structure introduced by Kemp and Regier (2012). This structure includes 33 family members spanning five generations, with a designated *ego* representing the Speaker (Figure 3). Since kinship terms in some languages—such as Korean—depend on the (binarized) gender of the speaker, we consider two ego identities: Alice (female) and Bob (male).

We examine two types of kinship naming terminologies: human (i.e., based on a sample of natural languages) and neural network-based (i.e., emerging from neural-network (NN) agents simulations). In the former case, we assume a simple probabilistic model to formalize encoding and decoding of messages by Speaker and Listener, while in the latter, the encoder and decoder are learned with the same neural network agents, which develop their own kinship terminology while playing a referential game. We refer to these systems as HP (for Human-Probabilistic) and NN, respectively.

4.1 Human (HP) Kinship Systems

We investigate kinship naming across four languages: English, Dutch, Spanish, and Vietnamese. We estimate the communicative need distribution $p(u)$, the Speaker's encoder $q_s(w|u)$, and the corresponding Bayesian decoder $\tilde{q}_s(u|w)$ using frequency counts extracted from text corpora, as in Kemp and Regier (2012). However, unlike this study, we estimate a separate need distribution for each language.

Concretely, for each family member u, we compile a set of commonly used referring expressions $T(u)$, such as, "mother", "mommy", "mom" for mother. For each term $w \in T(u)$, we estimate count(u, w)—the number of times w refers

to u—by searching the corpora using possessive constructions with the first person singular pronoun in each language, such as "*my* mother". If a term is polysemous (e.g., "brother" can refer to either elder brother or younger brother), the count is evenly divided among all plausible referents u.

We then compute the total frequency of each referent u by summing the counts across all its corresponding terms $\text{count}(u) = \sum_{w \in T(u)} \text{count}(u, w)$. Let $\mathcal{F}$ be the set of all family members excluding ego, we finally estimate required distributions for speakers

$$p(u) = \frac{\text{count}(u)}{\sum_{v \in \mathcal{F}} \text{count}(v)} \; ; \quad q_s(w|u) = \frac{\text{count}(u, w)}{\text{count}(u)}$$

$$\tilde{q}_s(u|w) = \frac{q_s(w|u)p(u)}{\sum_{v \in \mathcal{F}} q_s(w|v)p(v)}$$

Estimating the listener decoder $q_l(u \mid w)$ is inherently challenging, as text corpora rarely provide explicit information about how listeners interpret messages. To overcome this limitation, we simulate a large number of listener decoders in our experiments to systematically explore other potential trade-offs in human communication.

Further details regarding the corpora used and the counts are presented in Appendix F.

4.2 Emergent (NN) Kinship Systems

In order to prompt the emergence of a neural-network based kinship system, we frame the model task as a referential game in which the referents are family members (see Figure 1). The NN-Speaker is given access to the full family tree along with a randomly selected target individual. Based on this input, the Speaker generates a message intended to identify the target. Upon receiving the message and observing a candidate set that includes the target, the NN-Listener must infer which candidate the NN-Speaker is referring to.

4.2.1 Input encoding

The tree-like structure of a family motivates the use graphs (Akkerman et al., 2024). A kinship graph consists of 33 nodes, including an *ego* node, which can be either "Bob" (male) or "Alice" (female). The graph is adapted from the used family tree and is visualized in Figure 3. Each node represents an individual family member and is annotated with categorical features that encode key relational distinctions: *Gender* (male or female),[1] *Gender relative to ego* (equal or different), *Age relative to ego* (older or younger), and *Age relative to parent* (older or younger). All features are one-hot encoded and are designed to reflect the compositional kinship rules used in Kemp and Regier (2012).

[1]We consider only binary gender distinctions, consistent with those typically encoded in human kinship systems.

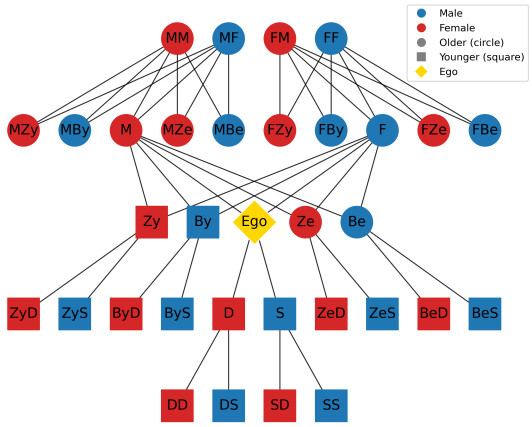

Figure 3: The kinship graph is adapted from the familial structures described by Kemp and Regier (2012). Nodes are labeled using abbreviations, where "F", "M", "B", "Z", "S", "D", "y", and "e" stand for "father", "mother", "brother", "sister", "son", "daughter", "younger", and "elder", respectively. For example, MBe denotes the "mother's elder brother". Each edge in the graph is bidirectional, labeled parent-of when traversing top-down and child-of when traversing bottom-up.

To connect nodes (i.e., family members), we diverge from Kemp and Regier (2012) by using only the two most primitive relationships—parent-of and child-of—which allow bidirectional traversal across generations. For example, the node F (father) connects to Bob (ego) via "F is parent of Bob," and to Be (Bob's elder brother) via "F is parent of Be"; correspondingly, Bob and Be each connect back to F via child-of edges. More complex relationships in the kinship trees of Kemp and Regier (2012), such as sibling-of, must instead be inferred compositionally from the primitive relations. This design encourages agents to discover and exploit relational structure rather than rely on shortcut or explicitly labeled edges.

We observe that the kinship graph contains redundant information. For instance, determining whether two individuals are brothers can be achieved by checking whether they share either the same mother or the same father. To eliminate such redundancy, we apply a pruning procedure that retains only the shortest paths from each node to the ego. This results in a best-first-search tree. We refer to this process as *graph pruning*.

4.2.2 Model architecture

Our agents are implemented as two neural networks. The NN-Speaker encodes the kinship graph G using a graph neural network GNN_s, producing node-level embeddings:

$$[\mathbf{h}_1, \ldots, \mathbf{h}_{33}] = \text{GNN}_s(G)$$

where $\mathbf{h}_i \in \mathbb{R}^d$ is the representation of the i-th node. To generate a message, the Speaker concatenates the embeddings of the ego and target nodes, transforms them via a two-layer network, and applies the Gumbel-Softmax (GS) to sample a *one-token* message w from a fixed vocabulary $\mathcal{V}$:

$$\begin{aligned} \text{score}_s(u) &= \mathbf{W}_{\text{lex}} \cdot \mathbf{W}_{\text{hid}} \cdot \text{cat}(\mathbf{h}_{\text{ego}}, \mathbf{h}_{\text{target}}) \\ w &\sim \text{GS}(\text{score}_s(u)) \end{aligned}$$

$\mathbf{W}_{\text{hid}} \in \mathbb{R}^{d_h \times 2d}$ and $\mathbf{W}_{\text{lex}} \in \mathbb{R}^{|\mathcal{V}| \times d_h}$ are trainable weight matrices, and $\text{score}_s(u) \in \mathbb{R}^{|\mathcal{V}|}$ denotes the unnormalized scores for tokens. Unless stated otherwise, we assume vocabulary size $|\mathcal{V}| = 128$.

The NN-Listener receives the same graph along with the sampled message w and infers the target referent. It encodes the graph using another graph neural network GNN_l, identical in architecture to the Speaker's, and computes compatible scores between the message and family members:

$$\begin{aligned} &[\mathbf{v}_1, \ldots, \mathbf{v}_{33}] = \text{GNN}_l(G) \\ &\text{score}_l(w, i) = \mathbf{e}_w^\top \cdot \mathbf{W} \cdot \mathbf{v}_i \quad \forall i \in [1, \ldots, 33] \end{aligned}$$

where $\mathbf{v}_i \in \mathbb{R}^d$ is the embedding of node i, $\mathbf{e}_w \in \mathbb{R}^{d_h}$ is the embedding of token w, and $\mathbf{W} \in \mathbb{R}^{d_h \times d}$ is a trainable bilinear transformation. The Listener selects the node with the highest score as its prediction. In all experiments, we use three shared-parameter graph neural layers, set $d = 80$ and $d_h = 20$, and the NN-Speaker and NN-Listener do not share parameters.

Because kinship structures are inherently relational, we encode the input with Relational Graph Convolutional Network (RGCN) (Schlichtkrull et al., 2018), designed for multi-relational graphs. In Appendix D we report alternative GNN architectures and hyperparameter configurations.

4.2.3 Training

From the kinship graph described above, we generate a dataset of 10,000 data points, each corresponding to a single game turn. Each data point consists of: (i) the full pruned kinship graph with the ego node uniformly sampled from $\{\text{Bob}, \text{Alice}\}$; (ii) a target node u uniformly selected from the remaining 32 nodes; (iii) a distractor set D of five nodes, uniformly sampled from the remaining nodes (i.e., excluding both the ego and the target node).

We split the dataset into 80% for training and 20% for validation. The two agents are trained jointly to minimize the following loss:

$$L = -\sum_{(u,D)} \log \frac{\exp(\text{score}_l(w, u))}{\sum_{v \in \{u\} \cup D} \exp(\text{score}_l(w, v))}$$

where w is the message generated by the Speaker for target u.[2] We use the Adam optimizer with a learning rate of 1×10^{-3}, training for 500 epochs using mini-batches of size 50.

4.2.4 Evaluation

To assess the generalizability of the learned communication protocol, we evaluate agents under a stricter criterion designed for faithful comparison with human-produced (HP) kinship systems. As in Kemp and Regier (2012), the NN-Listener must identify the target node from among *all* 32 possible family members (rather than the 6-candidate context used during training). The evaluation set thus consists of $2 \times 32 = 64$ data points, covering each $\text{ego} \in \{\text{Bob}, \text{Alice}\}$ paired with all 32 non-ego members.

Unlike during training, the NN-Speaker operates deterministically at evaluation time, producing the most likely token: $w^* = \arg\max_w \text{score}_s(u)[w]$; hence $q_s(w|u) = [[w = w^*]]$. The NN-Listener computes a probability distribution over all candidates using a softmax over the score function: $q_l(u|w) \propto \exp(\text{score}_l(w, u))$.

5 Experiments

We conduct computational experiments to validate our theoretical framework on the optimality of kinship naming across four languages: English (en), Dutch (nl), Spanish (es), and Vietnamese (vi).

5.1 Systems

To empirically evaluate the prediction that the semantic system exhibits an optimal framework if and only if the decoder is Bayesian, we need to simulate a large number of listener decoders. For each human language, we evaluate kinship naming between an HP-Speaker and one of the following types of HP-Listeners:

[2] During training, w sampled from Gumbel-Softmax is a distribution over $\mathcal{V}$ (while a discrete token in evaluation).

- *Bayesian* HP-Listener: uses a decoder identical to the Bayesian decoder employed by the HP-Speaker. This type is employed in the work of Kemp and Regier (2012); Zaslavsky et al. (2018).

- *Non-Bayesian* HP-Listener: uses a decoder that deviates from the Bayesian decoder by randomly changing its decision with a flip rate of $r_e \in \{0.1\%, 0.5\%, 1\%, 2\%\}$. [3]

Each Non-Bayesian condition is simulated with a population of 10 million listeners per language, resulting in a total of 40 million simulated listeners across all languages, along with four Bayesian listeners. Leveraging this large-scale simulation—and grounded in the well-established information-theoretic concepts of information loss and complexity—we assess whether communication with Bayesian listeners is indeed optimal, as predicted in Section 3.3.

We also evaluate the communication performance of our neural network agents (NN) under each language-specific communicative need distribution.[4] The communication setup is ego-specific, with ego $\in \{\text{Bob}, \text{Alice}\}$. Following the configuration described in Section 4 and summarized in Appendix B, we repeat training and evaluation 50 times to account for variability across runs.

5.2 Metrics

Since our primary interest lies in the optimality of the complexity-information loss tradeoff, our primary performance metric is the Euclidean *distance* to the optimal curve, defined in Section 3.3.

In addition, under each language's communicative need distribution, we evaluate *accuracy*, defined as the expected probability that a Listener correctly identifies the intended referent family member: $\mathbb{E}_{u\sim p,\, w\sim q_s(\cdot|u)}\left[q_l(u|w)\right]$. This is a *relaxed* variant of traditional accuracy, since the precise predictions of (natural language) Listeners are not directly observable due to polysemy (e.g., the term "brother" may refer to either elder brother or younger brother).

[3] It is important to note that the definition of these models is independent of the properties of the language. In other words, even though the non-Bayesian model is expected to perform worse than the Bayesian model in terms of decoding (or communicative success), this formulation is independent of the definitions of complexity and information loss, hence also of the extent to which their trade-off approaches optimality.

[4] We use the EGG framework (Kharitonov et al., 2019) and the PyTorch Geometric library (Fey and Lenssen, 2019). Our source code is publicly available at abc.anonymized.xyz.

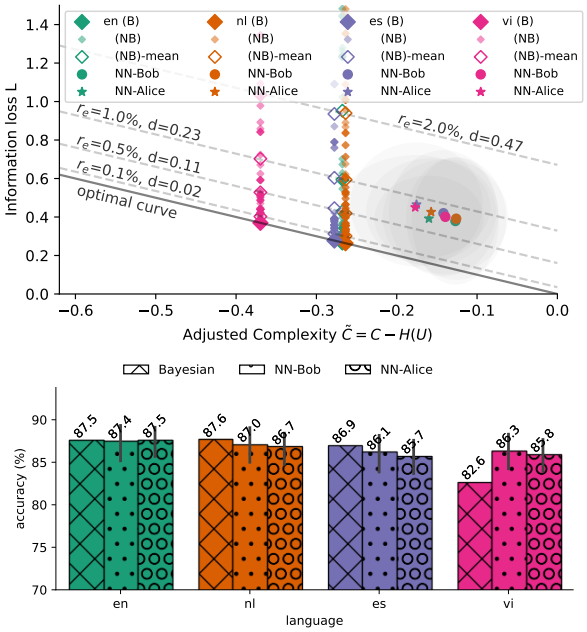

Figure 4: (Top) Trade-offs for HP systems with Bayesian (B) and Non-Bayesian (NB) listeners, and neural network (NN) systems (averaged over 50 runs; standard deviations shown as light gray ellipses). Trade-offs under Non-Bayesian HP-Listener conditions collectively (denoted as NB-mean) form lines that are approximately parallel to the optimal curve. Each line is annotated with the flip rate r_e and its distance d to the optimal curve. For visual clarity, only 100 randomly selected Non-Bayesian listeners are shown per language. (Bottom) Accuracy of NN Bayesian HP.

5.3 Results

The impact of non-Bayesian HP-Listeners. We evaluate four non-Bayesian HP-Listeners with error rates $r_e \in \{0.1\%, 0.5\%, 1\%, 2\%\}$. Figure 4-top shows the trade-off under each of these HP-Listener conditions, plus the condition without noise (Bayesian HP-Listener). The figure reveals that, across all languages, communication with the *Bayesian* HP-Listener lies exactly on the optimal curve (i.e., zero distance), while communication with non-Bayesian Listeners deviates progressively further from it. Moreover, HP-Listeners with lower error rates consistently yield more favorable trade-offs than those with higher error rates. These findings align with the theoretical prediction outlined in Section 3.3:trade-off optimality depends on how closely the HP-Listener's decoder approximates the Bayesian decoder of the HP-Speaker.

Interestingly, we observe that human communication systems in the three Indo-European languages—English, Dutch, and Spanish—exhibit similar levels of adjusted complexity and informa-

tion loss (and also accuracy as shown in Figure 4-bottom). In contrast, Vietnamese shows higher information loss but lower adjusted complexity, suggesting that it has more capacity to increase complexity in order to improve informativeness.

HP vs NN. Since the NN models are trained to minimize information loss, they are not inherently constrained by complexity. Fortunately, we find that early stopping serves as an effective mechanism for limiting complexity, therefore we select the checkpoint with accuracy closest to that of human communication (see Figure 4-bottom).

Figure 4-top shows that the NN communication system achieves trade-offs that are closer to the optimal curve than human communication with the non-Bayesian HP-Listener under $r_e = 1\%$ flipping rate. This suggests that communication systems emerging from object naming tasks can approach theoretical optimality without incurring exhaustive complexity—similar to human communication.[5]

NN naming system evolves towards optimality. Building on the previous results, we examine the evolution of NN communication over time. Figure 5 shows the trade-offs under the communicative need distribution for English (results for other languages are in Appendix C).

Initially, the emergent communication exhibits moderate complexity but high information loss, as agents rely on only a few vocabulary terms, resulting in a simple but inefficient language. As training progresses, information loss gradually decreases at the cost of increasing complexity, with the trade-off trajectory approaching the optimal curve. This trend suggests that agents progressively expand their vocabulary usage to enhance communicative success, thereby reducing information loss while incurring greater complexity.

6 Conclusion

We have used an information-theoretic framework for quantifying the trade-off between information loss and complexity in kinship naming systems. Our work accommodates non-Bayesian listeners—an important extension, as listener deviations are common in real-world settings. Indeed, even human listeners with clinically normal hearing often struggle with decoding in daily life (Ruggles et al., 2011). Our work enables the estimation of trade-offs under such realistic conditions, whether in human communication or in applied scenarios where messages are subject to perturbation. Conditions for optimality can arise when the communication channel is relatively noiseless, speakers and listeners share similar linguistic biases and representations, and decoding approximates Bayesian inference.

[5]Note that information loss is not an absolute indicator of accuracy—cf. the case of Vietnamese. This is analogous to what is commonly observed between log-likelihood and accuracy in classification tasks.

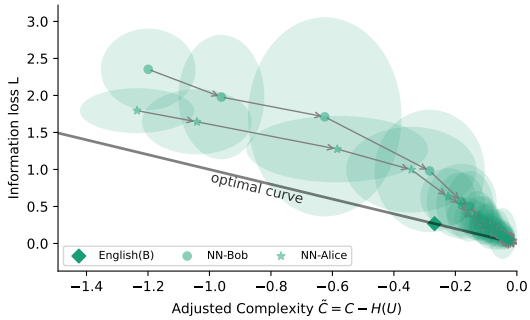

Figure 5: Trajectories of NN systems over time, recorded every 10 epochs, under the English communicative need distribution. Results are averaged over 50 runs, with standard deviation represented by ellipses.

Furthermore, we take into account the impact of cultural differences via communicative need distribution. By eliminating the dependence of the optimal trade-off curve on the need distribution, we enable clearer comparison of trade-offs across culturally shaped communication systems. For instance, Figure 4 shows that trade-offs vary across language families: Vietnamese, for example, achieves lower complexity relative to its own complexity capacity compared to Indo-European languages, albeit at the cost of informativeness. The figure also suggests that optimal human communication tends to cluster within an information loss range of approximately $[0.2, 0.4]$ and an adjusted complexity range of $[-0.4, -0.2]$. A deeper analysis of these patterns, along with an expanded comparison across more languages and linguistic families, is left for future work.

Finally, we have shown that kinship naming systems emerging from NN models exhibit a trade-off comparable to that of HP systems with less than 1% noise. This suggests that, at least within the constraints of our NN system, optimizing only for communication accuracy (or information loss) and applying early stopping (accounting for limited lifespan of individuals) are sufficient mechanisms to trigger the evolutionary dynamics that result in the observed trade-offs.

Limitations

As in Zaslavsky et al. (2018)'s Information Bottleneck framework, we measure complexity as the mutual information between the object and word random variables. This quantity is upper-bounded by the entropy of the communicative need distribution. However, this measure does not fully capture the richness of natural languages: in principle, a language can achieve unbounded complexity by continually expanding its vocabulary.

We restrict our NN communication systems to *one-token* messages, whereas natural languages often employ compositional expressions to refer to kinship relations—i.e., components of the expression systematically correspond to semantic properties of the kin category. For example, in English, the prefix *grand-* consistently denotes *parent-of-*, while in Spanish, the suffixes *-a* and *-o* typically indicate female and male family members, respectively. Such compositional strategies cannot emerge in our current setup. Additionally, because the NN-Speaker generates names deterministically, the emergent language lacks synonymy, which is common in natural languages (e.g., in English mother may be referred to as either "mother" or "mum").

References

Daniel Akkerman, Phong Le, and Raquel G. Alhama. 2024. The emergence of compositional languages in multi-entity referential games: from image to graph representations. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 18713–18723, Miami, Florida, USA. Association for Computational Linguistics.

G. Anand and T. Regier. 2023. Kinship terminologies reflect culture-specific communicative need: Evidence from hindi and english. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 45.

Shaked Brody, Uri Alon, and Eran Yahav. 2022. How attentive are graph attention networks? In *International Conference on Learning Representations*.

Jon W. Carr, Kenny Smith, Jennifer Culbertson, and Simon Kirby. 2020. Simplicity and informativeness in semantic category systems. *Cognition*, 202.

Rahma Chaabouni, Eugene Kharitonov, Emmanuel Dupoux, and Marco Baroni. 2021. Communicating artificial neural networks develop efficient color-naming systems. *Proceedings of the National Academy of Sciences*, 118(12).

Rahma Chaabouni, Florian Strub, Florent Altché, Eugene Tarassov, Corentin Tallec, Elnaz Davoodi, Kory Wallace Mathewson, Olivier Tieleman, Angeliki Lazaridou, and Bilal Piot. 2022. Emergent Communication at Scale. In *International Conference on Learning Representations*.

Mark Davies. 2002–2024. Corpus del español. Online linguistic resource providing historical and modern Spanish texts.

Mark Davies. 2010. The corpus of contemporary american english as the first reliable monitor corpus of english. *LLC*, 25:447–464.

Matthias Fey and Jan Eric Lenssen. 2019. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*.

Serhii Havrylov and Ivan Titov. 2017. Emergence of Language with Multi-agent Games: Learning to Communicate with Sequences of Symbols. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Charles Kemp and Terry Regier. 2012. Kinship categories across languages reflect general communicative principles. *Science*, 336:1049–1054.

Charles Kemp, Yang Xu, and Terry Regier. 2018. Semantic typology and efficient communication. *Annual Review of Linguistics*, 4:109–128.

Eugene Kharitonov, Rahma Chaabouni, Diane Bouchacourt, and Marco Baroni. 2019. EGG: a toolkit for research on Emergence of lanGuage in Games. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, pages 55–60, Hong Kong, China. Association for Computational Linguistics.

Angeliki Lazaridou and Marco Baroni. 2020. Emergent multi-agent communication in the deep learning era. *CoRR*, abs/2006.02419.

Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. 2017. Multi-Agent Cooperation and the Emergence of (Natural) Language. In *International Conference on Learning Representations*.

Nelleke Oostdijk, Martin Reynaert, Véronique Hoste, and Ineke Schuurman. 2013. The construction of a 500-million-word reference corpus of contemporary written Dutch. In *Essential Speech and Language Technology for Dutch: Results by the STEVIN-programme*, chapter 13. Springer Verlag.

Nam Pham. 2024. vietvault (revision bee0136).

Terry Regier, Charles Kemp, and Paul Kay. 2015. Word meanings across languages support efficient communication. *The handbook of language emergence*, pages 237–263.

Dorea Ruggles, Hari Bharadwaj, and Barbara G. Shinn-Cunningham. 2011. Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proceedings of the National Academy of Sciences*, 108(37):15516–15521.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *15th Extended Semantic Web Conference (ESWC)*, pages 593–607. Springer, Cham.

Claude E Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

Naftali Tishby, Fernando C Pereira, and William Bialek. 2000. The information bottleneck method. *arXiv preprint physics/0004057*.

Hy Van Luong. 1989. Vietnamese kinship: structural principles and the socialist transformation in northern vietnam. *The Journal of Asian Studies*, 48(4):741–756.

Noga Zaslavsky, Charles Kemp, Terry Regier, and Naftali Tishby. 2018. Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, 115:7937–7942.

Noga Zaslavsky, Mora Maldonado, and Jennifer Culbertson. 2021. Let's talk (efficiently) about us: Person systems achieve near-optimal compression. In *Proceedings of the annual meeting of the cognitive science society*, volume 43.

Noga Zaslavsky, Terry Regier, Naftali Tishby, and Charles Kemp. 2019. Semantic categories of artifacts and animals reflect efficient coding. In *Proceedings of the 41th Annual Meeting of the Cognitive Science Society, CogSci 2019: Creativity + Cognition + Computation, Montreal, Canada, July 24-27, 2019*, pages 1254–1260. cognitivesciencesociety.org.

A A Lower bound for Information Loss

In this appendix we derive a lower bound for information loss.

$$
\begin{aligned}
L =& -\mathbb{E}_{u\sim p}\mathbb{E}_{w\sim q_s(\cdot|u)} \log q_l(u|w) \\
=& -\sum_u p(u) \sum_w q_s(w|u) \log q_l(u|w) \\
=& -\sum_u p(u) \sum_w q_s(w|u) \\
& \log \left[p(u) \frac{q_s(w|u)}{p_s(w)} \cdot \frac{q_l(u|w) p_s(w)}{q_s(w|u) p(u)} \right] \\
=& -\sum_u p(u) \sum_w q_s(w|u) \log p(u) \\
& -\sum_u p(u) \sum_w q_s(w|u) \log \frac{q_s(w|u)}{p_s(w)} \\
& -\sum_u p(u) \sum_w q_s(w|u) \log \frac{q_l(u|w) p_s(w)}{q_s(w|u) p(u)} \\
=& -\sum_u p(u) \log p(u) \\
& -\sum_{u,w} p(u) q_s(w|u) \log \frac{q_s(w|u)}{p_s(w)} \\
& +\sum_w p_s(w) \sum_u \tilde{q}_s(u|w) \log \frac{\tilde{q}_s(u|w)}{q_l(u|w)} \\
=& H(U) - C + \mathbb{E}_{w\sim p_s} \left[D_{\mathrm{KL}}(\tilde{q}_s \| q_l) \right]
\end{aligned}
$$

where:

- $H(U) = -\sum_u p(u) \log p(u)$ is the entropy of the communicative need distribution;
- C is the complexity, as defined in Equation 1;
- $\tilde{q}_s(u|w) = \frac{q_s(w|u)\, p(u)}{p_s(w)}$ is the Bayesian decoder of the Speaker.

Since the KL divergence is always non-negative, the information loss is lower-bounded by:

$$L \geq H(U) - C.$$

This bound is achieved when $q_l = \tilde{q}_s$, i.e., when the Listener's decoder is identical to the Bayesian decoder of the Speaker.

B Hyper-parameters

We show relevant hyper-parameters for all experiments in Table 1. Gumbel-softmax temperature controls the Gumbel-softmax sampling distribution: lower values tend towards a one-hot encoding, whereas higher values tend towards a uniform encoding.

C The evolution of NN Kinship System

We show in Figure 6 how NN communication evolved during training in the environments of the four languages: English (en), Dutch (nl), Spanish (es), and Vietnamese (vi).

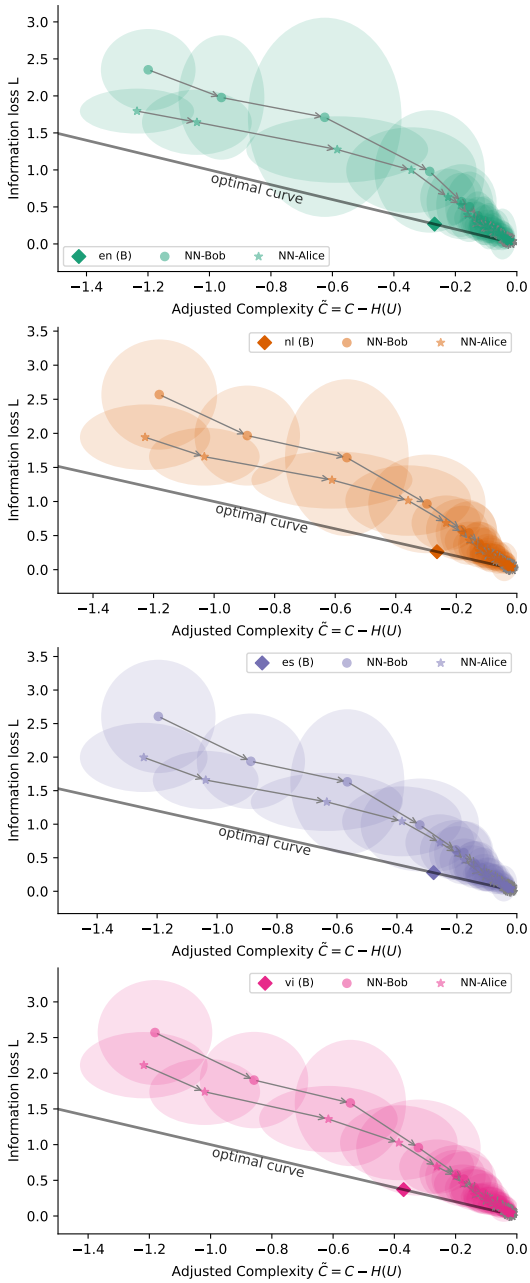

Figure 6: Evolution of complexity and information loss in ego-specific NN communication (average over 50 runs) under four communicative need distributions English, Dutch, Spanish, and Vietnamese.

D Impact of Model Architecture on Performance

We investigate the impact of architectural choices that led to the NN-agents used in the main study.

Table 1: Hyperparameter settings used in our experiments. The third column reports the values selected for the main study, while the last column lists the values considered during architecture search.

	Hyperparameter	**Value (main study)**	**Values (architecture search)**
Model Architecture	embedding dimensions d	80	–
	hidden dimension d_h	20	–
	Graph neural net	RGCN	RGCN, GATv2Conv
	# graph net layers	3	–
	Vocabulary size $\|\mathcal{V}\|$	128	16, 32, 64, 128, 256
	Graph pruning	True	True, False
Training	Optimizer	Adam	–
	Learning rate	1×10^{-3}	–
	Batch size	50	–
	# distractors	5	–
	Gumbel-softmax temperature	1.5	–

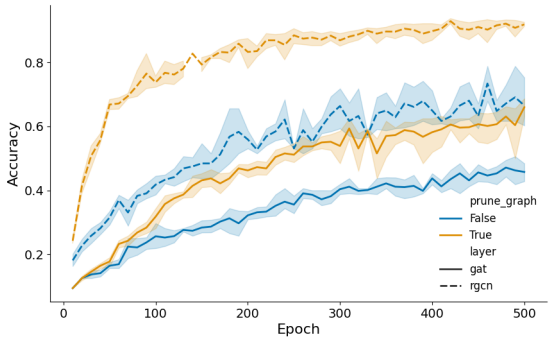

Figure 7: Evaluation accuracy, for simulations with and without pruning (n=40 runs, varying initialization).

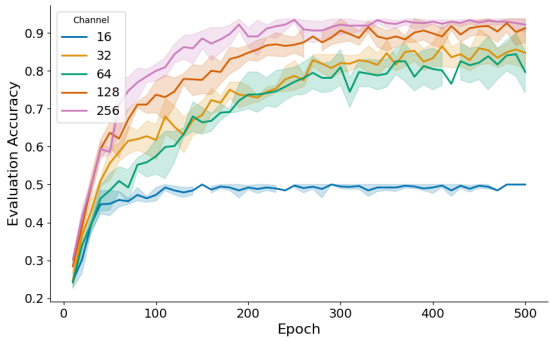

Figure 8: Evaluation accuracy, for simulations with pruning and RGCN layer (n=50 runs in total, with 10 different initializations for each vocabulary size).

Specifically, we examine three factors: *graph pruning*, *layer type*, and *channel capacity* (i.e., vocabulary size) in the environment of uniform communicative need distribution.

In the first study, based on the configuration described in Section 4 and summarized in Appendix B, we construct four variants by systematically varying the use of graph pruning and the choice of layer type (either RGCN (Schlichtkrull et al., 2018) or GATv2Conv (Brody et al., 2022)). As shown in Figure 7, both graph pruning and the use of RGCN layers are critical for achieving high communicative success, each contributing approximately 20 percentage points to the final communication accuracy of the NN-agents.

In the second study, we vary the vocabulary size (16, 32, 64, 128, 256) to examine the effect of channel capacity. Figure 8 demonstrates that a sufficiently large vocabulary relative to the size of the object set (32 kinship terms in our case) is crucial. For instance, a vocabulary size of 16 constrains communication to approximately 50% accuracy. In contrast, increasing the vocabulary size to 32 or greater substantially improves accuracy to 80% and above.

E Human Communication with Suboptimal HP-Listeners

We investigate the impact of suboptimal HP-Listeners on human communication across four language environments: English, Dutch, Spanish, and Vietnamese. Figure 9 reports both the distance to the optimal trade-off curve and the accuracy of the corresponding communication systems. The results reveal a linear relationship between the HP-Listener's error rate and both distance and accuracy. This finding supports our theoretical framework, confirming that communication with non-Bayesian HP-Listeners—i.e., those that deviate more from the Bayesian Listener—results in trade-offs that lie further from the optimal curve.

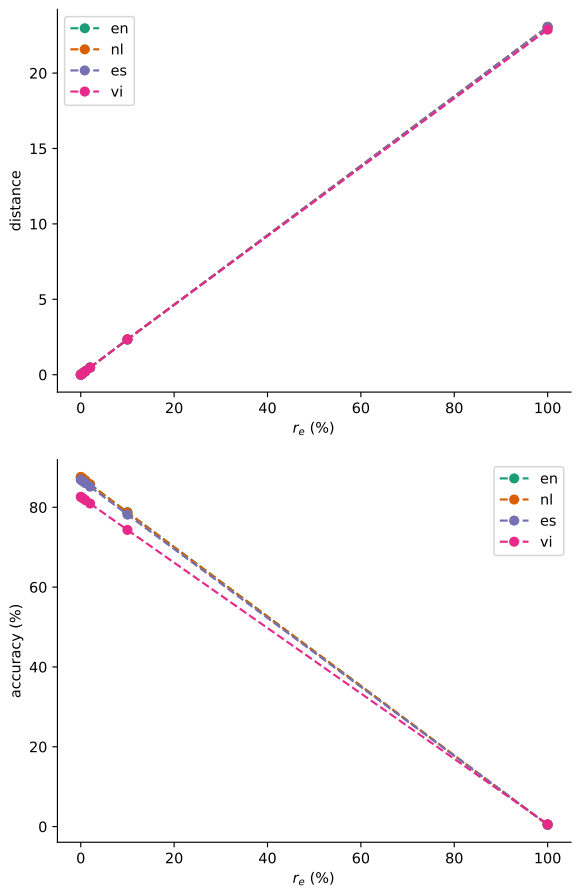

Figure 9: (Top) Distance to the optimal curve and (Bottom) accuracy of human communication with suboptimal Listeners at varying flip rates $r_e \in \{0, 0.001, 0.005, 0.01, 0.02, 0.1, 1.0\}$. Dashed lines indicate the linear relationship between flip rate and distance/accuracy.

F Kinship counts

We extract counts of family-member and kinship-term pairs (see Table 2) from text corpora in four languages:

- **English:** We use the Corpus of Contemporary American English (COCA) (Davies, 2010), a widely-used and balanced corpus of American English. It contains over one billion words from 1990–2019, covering eight genres such as spoken language, fiction, news, academic writing, and web content.

- **Dutch:** We use the SoNaR corpus (Oostdijk et al., 2013), a 500-million-word reference corpus of contemporary Dutch that includes both written and spoken data. SoNaR integrates material from various sources such as newspapers, newsletters, books, websites, and transcripts.

- **Spanish:** We use the NOW (News on the Web) corpus from the Corpus del Español (Davies, 2002–2024), which includes approximately 7.6 billion words from web-based newspapers and magazines across 21 Spanish-speaking countries.

- **Vietnamese:** We use the VietVault corpus (Pham, 2024), a dataset filtered and curated from Common Crawl dumps prior to 2023. We sample a 5GB subset of the full 80GB of raw Vietnamese text, spanning multiple domains.

Table 2: Counts for (family-member, term) pairs from text corpora.

			English		Dutch		Spanish		Northern Vietnamese	
Gen.	**Label**	**Full Name**	**Term**	**Count**	**Term**	**Count**	**Term**	**Count**	**Term**	**Count**
1st	MM	mother's mother	Grandmother	3949	Grootmoeder	529	Abuela	8577.5	bà ngoại	210
			Grandma	882	Oma	548.5	Yaya	10	bà	411.5
			Gran	25.5						
	MF	mother's father	Grandfather	3189	Grootvader	720	Abuelo	8075.5	ông ngoại	125
			Grandpa	399.5	Opa	361.5	Yayo	2	ông	1210
	FM	father's mother	Grandmother	3949	Grootmoeder	529	Abuela	8577.5	bà nội	266
			Grandma	882	Oma	548.5	Yaya	10	bà	411.5
			Gran	25.5						
	FF	father's father	Grandfather	3189	Grootvader	720	Abuelo	8075.5	ông nội	290
			Grandpa	399.5	Opa	361.5	Yayo	2	ông	1210
2nd	M	mother	Mother	65458	Moeder	18009	Madre	64464	mẹ	18004
			Mom	29849	Mama	1511	Mama	1295		
			Mommy	707	Ma	1186	Mami	1071		
			Momma	388			Mamá	52914		
	F	father	Father	63318	Vader	19939	Padre	77214	bố	5021
			Dad	31100	Papa	896	Papa	1250	cha	3390
			Daddy	3017	Pa	1346	Papi	512		
							Papá	47494		
	MZy	mother's younger sister	Aunt	1022	Tante	204.75	Tía	1404.75	dì	243
			Auntie	29			Tita	1.5		
	MBy	mother's younger brother	Uncle	1501.75	Oom	509	Tío	1404.75	cậu	237
							Tito	2.5		
	MZe	mother's elder sister	Aunt	1022	Tante	204.75	Tía	1404.75	bác gái	2.5
			Auntie	29			Tita	1.5	bác	96.5
	MBe	mother's elder brother	Uncle	1501.75	Oom	509	Tío	1404.75	bác trai	2
							Tito	2.5	bác	96.5
	FZy	father's younger sister	Aunt	1022	Tante	204.75	Tía	1404.75	cô	435
			Auntie	29			Tita	1.5		
	FBy	father's younger brother	Uncle	1501.75	Oom	509	Tío	1404.75	chú	388
							Tito	2.5		
	FZe	father's elder sister	Aunt	1022	Tante	204.75	Tía	1404.75	bác gái	2.5
			Auntie	29			Tita	1.5	bác	96.5
	FBe	father's elder brother	Uncle	1501.75	Oom	509	Tío	1404.75	bác trai	2
							Tito	2.5	bác	96.5
3rd	Zy	younger sister	Sister	9587.5	Zus	2029.5	Hermana	13610.5	em	2425.5
			Sis	43.5	Zusje	826	Tata	82	em gái	849
	By	younger brother	Brother	11687	Broer	2968	Hermano	22723	em	2425.5
			Bro	63.5	Broertje	565	Tete	4	em trai	459
	Ze	elder sister	Sister	9587.5	Zus	2029.5	Hermana	13610.5	chị	1752
			Sis	43.5			Tata	82		
	Be	elder brother	Brother	11687	Broer	2968	Hermano	22723	anh	3173
			Bro	63.5			Tete	4		
4th	D	daughter	Daugther	23571	Dochter	10378	Hija	68744	con	5296.5
					Dochtertje	1580			con gái	2543
	S	son	Son	28815	Zoon	8737	Hijo	94575	con	5296.5
					Zoontje	3753			con trai	2370
	ZyD	younger sister's daugter	Niece	326.25	Nicht	71.75	Sobrina	746	cháu	275.33
					Nichtje	121.75			cháu họ	4
	ZyS	yougher sister's son	Nephew	339.25	Neef	177.75	Sobrino	1067	cháu	275.33
					Neefje	76.5			cháu họ	4
	ByD	younger brother's daughter	Niece	326.25	Nicht	71.75	Sobrina	746	cháu	275.33
					Nichtje	121.75			cháu họ	4
	ByS	younger brother's son	Nephew	339.25	Neef	177.75	Sobrino	1067	cháu	275.33
					Neefje	76.5			cháu họ	4
	ZeD	elder sister's daughter	Niece	326.25	Nicht	71.75	Sobrina	746	cháu	275.33
					Nichtje	121.75			cháu họ	4
	ZeS	elder sister's son	Nephew	339.25	Neef	177.75	Sobrino	1067	cháu	275.33
					Neefje	76.5			cháu họ	4
	BeD	elder brother's daughter	Niece	326.25	Nicht	71.75	Sobrina	746	cháu	275.33
					Nichtje	121.75			cháu họ	4
	BeS	elder brother's son	Nephew	339.25	Neef	177.75	Sobrino	1067	cháu	275.33
					Neefje	76.5			cháu họ	4
5th	DD	daughter's daughter	Granddaughter	370	Kleindochter	87	Nieta	1304	cháu	275.33
									cháu ngoại	30
	DS	daughter's son	Grandson	478	Kleinzoon	117	Nieto	1685.5	cháu	275.33
									cháu ngoại	30
	SD	son's daughter	Granddaughter	370	Kleindochter	87	Nieta	1304	cháu	275.33
									cháu nội	25
	SS	son's son	Grandson	478	Kleinzoon	117	Nieto	1685.5	cháu	275.33
									cháu nội	25