



HAL
open science

Semantic Channel Equalizer: Modelling Language Mismatch in Multi-User Semantic Communications

Mohamed Sana, Emilio Calvanese Strinati

► **To cite this version:**

Mohamed Sana, Emilio Calvanese Strinati. Semantic Channel Equalizer: Modelling Language Mismatch in Multi-User Semantic Communications. GLOBECOM 2023 - 2023 IEEE Global Communications Conference, Dec 2023, Kuala Lumpur, Malaysia. pp.2221-2226, 10.1109/GLOBECOM54140.2023.10437371 . cea-04549339

HAL Id: cea-04549339

<https://cea.hal.science/cea-04549339>

Submitted on 30 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semantic Channel Equalizer: Modelling Language Mismatch in Multi-User Semantic Communications

Mohamed Sana, Emilio Calvanese Strinati
CEA-Leti, Université Grenoble Alpes, F-38000 Grenoble, France
Email : {mohamed.sana; emilio.calvanese-strinati}@cea.fr

Abstract—We consider a multi-user semantic communications system in which agents (transmitters and receivers) interact through the exchange of semantic messages to convey meanings. In this context, languages are instrumental in structuring the construction and consolidation of knowledge, influencing conceptual representation and semantic extraction and interpretation. Yet, the crucial role of languages in semantic communications is often overlooked. When this is not the case, agent languages are assumed compatible and unambiguously interoperable, ignoring practical limitations that may arise due to language mismatching. This is the focus of this work. When agents use distinct languages, message interpretation is prone to semantic noise resulting from critical distortion introduced by *semantic channels*. To address this problem, this paper proposes a new semantic channel equalizer to counteract and limit the critical ambiguity in message interpretation. Our proposed solution models the mismatch of languages with measurable transformations over semantic representation spaces. We achieve this using optimal transport theory, where we model such transformations as transportation maps. Then, to recover at the receiver the meaning intended by the teacher we operate *semantic equalization* to compensate for the transformation introduced by the semantic channel, either before transmission and/or after the reception of semantic messages. We implement the proposed approach as an operation over a codebook of transformations specifically designed for successful communication. Numerical results show that the proposed semantic channel equalizer outperforms traditional approaches in terms of operational complexity and transmission accuracy.

I. INTRODUCTION

Wireless networked intelligence and services are fuelling the evolution towards sustainable smart society, industry, and economies. Their extraordinary success and operational effectiveness have led to game-changer innovations in communication technologies, computer science, AI and hardware, making intelligent cyber physical systems an essential support for our daily lives. Intelligence consumes large amounts of data which need critically scarce resources to be generated, transported through the network, verified, processed and memorized. To accommodate such burgeoning new services, in parallel to the ongoing effort to investigate traditional approaches consisting in *e.g.* exploring higher frequencies, semantic communication (SemCom) appears as an unlocking solution and is now envisioned as a potential key enabler of future 6G networks [1]. This approach engineers communication systems to effectively recreate or infer the meaning of what has been communicated rather than to just optimize opaque data pipes aiming to reproduce exactly exchanged sequences of symbols. It can achieve significant source data compression gain, reducing the

amount of data to be sensed, communicated, and/or processed, thereby, reducing energy consumption, and latency of wireless communication systems. Communication of meanings to enable effective cooperation between intelligent agents requires reliable exchanges of semantics [2]. This is possible thanks to a common or shared language (either natural or artificial), which endows messages (or observations of the world) with semantic meanings [3]. Common language can be learned a priori *e.g.* via a (supervised) learning procedure. It can also emerge from an interactive consensus based on *e.g.* reinforcement learning (RL), by learning a communication protocol (an effective language to attain a goal driving cooperation) [4]. This requires communication attempts, thus incurring errors and communication costs [5], [6]. Alternatively, common language can be suggested by the environment, where agents ground their language in representations of the observed world [7]. Current work on semantic communications assumes that the interaction between agents is based on an agreed common language, possibly learned or given. This assumption considers that communication errors occur only because of *i)* syntactic wireless channel, *ii)* lossy semantic compression, or *iii)* limited expressivity of semantic extraction and interpretation function, ruling how knowledge is extracted, represented and interpreted. However, when agents use distinct languages, our previous work [2] shows that communication is prone to semantic noise that might stimulate interpretation errors causing defective cooperation strategies [8].

To the best of our knowledge, there is no existing framework that explicitly address the problem of language mismatch in semantic communications. It is the focus of this work, which proposes a new framework for semantic channel equalization (EQ). To do so, *1)* we first explore the implications of language mismatch (ruling knowledge representation and reasoning capabilities) and propose a new modelling of the semantic channel between cooperating agents. In our proposed modelling, a mismatch of languages translates to a mismatch of knowledge representation, captured by *measurable transformations* introduced on semantic representation space. Hence, *2)* we propose a novel framework based on optimal transport (OT) theory, which models such transformations as transport plans that we learn to construct a codebook. With the learned codebook of transformations, compensating language mismatch becomes a problem of identifying and implementing the right transformation during communication. *3)* We achieve this by defining a new measure to quantify the risk of misinterpreting the conveyed meaning, which we optimize using a Bayes-optimal approach. Eventually, this

This work was partially funded by the French government under the France 2030 ANR program “PEPR Networks of the Future” (ref. 22-PEFT-0010) and by the 6G-GOALS Project under the HORIZON program (no. 101139232).

work provides a solid foundation for robust and effective multi-user semantic communications. It can also be applied to multi-task learning and knowledge transferability [9].

II. SYSTEM MODEL

We consider the system model described in Fig. 1, where K users employ distinct languages (different logic), here artificially modeled by neural networks (NNs), to endow messages (observations of the world) with semantic meanings. In this figure, two transmitters (TX) k and l observe images of digits and encode underlying information for two receivers (RX) k and l , which try, respectively, to guess the digit on the image or its parity. Though the TXs act on the same observations of the world, they may structure differently information underlying observations for their respective RXs. Note that this reference example serves to easily explain our proposed method, which, obviously, is not limited to such use cases.

A. General framework: the central role of language

We refer to the logic used by each TX to encode observations as its *language generator*, the encoded information as *semantic message*, and the mechanism used at the RX to recover the meaning of the information as *language interpreter*.

Language generator. We model language generator as a mapping function $\lambda : \mathcal{M} \rightarrow \mathcal{X}$, which endows message or observation $m \in \mathcal{M}$ with semantic meaning $x \in \mathcal{X}$. Here, the message space \mathcal{M} and the semantics space \mathcal{X} are topological spaces, which can be continuous or discrete space. Also, λ can be deterministic or stochastic. In the latter case, we refer to $\lambda(m)$ as a realization of this stochastic process (e.g. a sampling from a generated probability distribution): $\lambda(m) = x \sim p(x|m)$. A language generator defines a way to partition a semantic space and associates to each partition a semantic meaning. Indeed, let $P = \{P_1, P_2, \dots\}$ be a partition of the probability space $(\mathcal{X}, \mathfrak{B}, \mu)$, where μ is a probability measure over \mathcal{X} and \mathfrak{B} is a σ -algebra on \mathcal{X} (i.e. a nonempty collection of subsets of \mathcal{X} closed under complement, countable unions and intersections). We will often think of a partition of \mathcal{X} as a list of measurable sets that cover \mathcal{X} . We refer to P_i as the i -th atom of the partition P . Thus $\mu(P_i)$ is the probability measure of atom P_i . An atom may capture a certain level of semantics. For instance, for an image classification task, an atom may be related to the features (semantic symbols) associated to a certain class of an image. Different partitioning of the semantic space captures

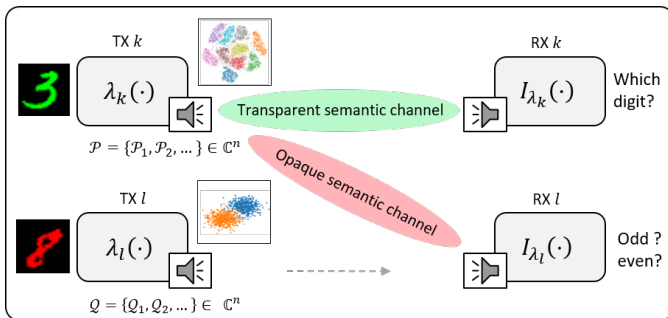


Fig. 1: General system model.

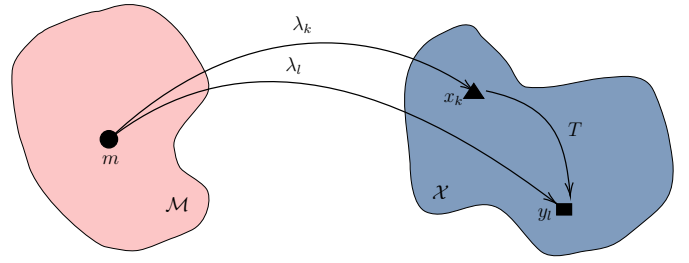


Fig. 2: Illustration language mismatch between user k and l .

different level of semantics. Also, different languages may partition differently the semantic space.

Language interpreter. Depending on the meaning the TX assigns to each transmitted data m (the atom to which m belongs to), generated semantic symbols get interpreted by the language interpreter using a mapping function, $I_\lambda : \mathcal{X} \rightarrow \mathcal{M}_\lambda$. Note that the target message space \mathcal{M}_λ may differ from \mathcal{M} . Thus, I_λ can but is not required to be the inverse function λ^{-1} . Besides, λ^{-1} is not necessarily defined unless λ is invertible: a bijective function, i.e. a one-to-one correspondence mapping. However, λ can be a one-to-many mapping: a message can be associated with different symbols, each conveying the same information. Conversely, λ can be a many-to-one mapping: multiple messages are represented with the same symbol. Hence, $(\mathcal{M}, \lambda, I_\lambda, \mathcal{M}_\lambda, \mathcal{X}, \mathfrak{B}, \mu)$ fully defines a language, whose expressivity is crucial for effective communication. Indeed, language may be limited: either *i*) not all messages of the world have a meaning in this language or *ii*) can be efficiently represented by the associated language generator or *iii*) the meaning accurately recovered by the language interpreter. This work does not focus on learning language generators or interpreters, which we assume are given. Instead, we focus specifically on communication between intelligent agents (with reasoning and processing capabilities that follows a language logic) to convey semantic messages to each other. These messages are interpretable by the RX depending on the meaning the TX assigns to them. Correct interpretation is not necessarily guaranteed, unless the TX and RX agree on and implement a common communication strategy or plan. For instance, considering the aforementioned example, RX k may correctly interpret semantic messages transmitted by TX k as they use the same language. In this case, we say that the semantic channel between TX k and RX k is transparent: they agree on the same logic for encoding and interpreting information. In contrast, RX l may fail to recover the meaning intended by TX k (in communication between TX k and RX l). In this case, the semantic channel between TX k and RX l is said to be opaque. It is the focus of this work, which proposes a new framework for semantic channel equalization.

B. Semantic channel modelling

Consider a communication between TX k and RX l . Let $P = \{P_1, P_2, \dots\}$ and $Q = \{Q_1, Q_2, \dots\}$ be the two finite partitions of the probability space $(\mathcal{X}, \mathfrak{B}, \mu)$ defined by their language generators λ_k and λ_l , respectively. Given a message $m \in \mathcal{M}$, TX k encodes and transmits the associated meaning $x_k = \lambda_k(m) \in P_i$ to RX l , which interprets x_k in its own language as $I_{\lambda_l}(x_k)$. However, in RX l language, data ‘ m ’

maps to $y_l = \lambda_l(m) \in Q_j$ (see Fig. 2). Thus, the atom P_i in TX k language is equivalent to atom Q_j in RX l language. As semantics are associated with atoms of partitions, RX l may misinterpret received symbol x_k , especially when $x_k \notin Q_j$. More specifically, when TX k communicates with RX l , only the observations that get mapped into $P_i \cap Q_j$ can be correctly interpreted in RX l language. Therefore, a mismatch of language translates to a mismatch of atoms of language partitions. Hereafter, we propose to model these mismatches with *measurable transformations*.

Definition 1. Let $(\mathcal{X}, \mathfrak{B})$ be a measurable space where \mathfrak{B} denotes a σ -algebra on \mathcal{X} . A function $T : \mathcal{X} \rightarrow \mathcal{X}$ is said to be a measurable transformation if for every set $B \in \mathfrak{B}$, **the preimage of B under T denoted $T^{-1}(B)$ is in \mathfrak{B}** ; i.e.:

$$T^{-1}(B) := \{x \in \mathcal{X} \mid T(x) \in B\} \in \mathfrak{B}. \quad (1)$$

Proposition 1. Let $P = \{P_1, P_2, \dots\}$ and $Q = \{Q_1, Q_2, \dots\}$ define two finite partitions of a probability space $(\mathcal{X}, \mathfrak{B}, \mu)$ such that $\mu(P_i) > 0$, $\forall P_i \in P$. Let $T \in \mathcal{T}$ be a measurable transformation. We define, $\forall (P_i, Q_j) \in P \times Q$:

$$\rho_{P_i \rightarrow Q_j}(T) = \frac{\mu(T^{-1}(Q_j) \cap P_i)}{\mu(P_i)}. \quad (2)$$

Intuitively, as show in Fig. 3, $\rho_{P_i \rightarrow Q_j}(T)$ quantifies the fraction of the volume of P_i that ends up in the set Q_j (information transfer from P_i to Q_j) under one iterate of T [10]. Since μ is a probability measure, we have $0 \leq \mu(T^{-1}(Q_j) \cap P_i) \leq \mu(P_i)$, thus, $0 \leq \rho_{P_i \rightarrow Q_j}(T) \leq 1$. To align atom P_i with atom Q_j , it is sufficient to find a transformation T that transports $x \in P_i$ into its equivalent semantic atom $T(x) \in Q_j$:

$$\mathbf{PO} : \arg \max_{T \in \mathcal{T}} \rho_{P_i \rightarrow Q_j}(T). \quad (3)$$

Transformation T can be viewed as a language “translator” bridging the semantic gap between the source and target atom.

Definition 2 (Language mismatch). Let J_P and J_Q define two finite index sets of two language partitions $P = \{P_1, P_2, \dots\}$ and $Q = \{Q_1, Q_2, \dots\}$, respectively. Let $\kappa : J_P \rightarrow J_Q$ define a function, which associates each atom $i \in J_P$ of partition P to the corresponding semantic atom $j \in \kappa(i) \subset J_Q$ in partition Q . Here, $\kappa(\cdot)$ can but is not required to be a one-to-one mapping. It can also be a many-to-one (e.g. the digits $\{1, 3, 5, 7, 9\}$ are all odd numbers) or one-to-many mapping (e.g. an even digit can be $\{0, 2, 4, 6, 8\}$). Let $T \in \mathcal{T}$ be a measurable transformation. We define

$$\rho_{P \rightarrow Q}(T) = \sum_{i \in J_P} \sum_{j \in \kappa(i)} \rho_{P_i \rightarrow Q_j}(T). \quad (4)$$

In particular, when T is the identity function, $\rho_{P \rightarrow Q}$ measures the mismatch of source and target language partitions.

From Proposition (1), language mismatch can be captured by measurable transformations acting on \mathcal{X} (see Fig. 3). These transformations model the semantic channel between users. To recover exactly at the RX the meaning intended by the TX sending a given symbol, either the RX manages to compensate for transformations introduced by the semantic channel (post-equalization), or the TX intentionally distorts intended meaning by selecting another symbol (pre-equalization).

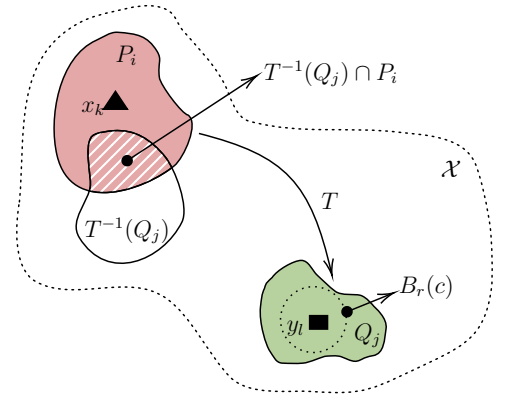


Fig. 3: Partitioning of the semantic representation’s space.

III. LEARNING MEASURABLE TRANSFORMATIONS

In this section, we are interested in finding and describing measurable transformations for modelling the semantic channel between two users. Our approach relies on optimal transport (OT) theory to devise measurable transformations. We first introduce the original Monge-Kantorovitch problem, then its discrete formulation. Eventually, we propose a new optimization framework, which relaxes the original problem constraints to learn semantic transport maps.

A. Background on the Monge-Kantorovitch Problem

Let us consider a communication between two users employing distinct languages, referred to as the source and target language, each partitioning the semantic space \mathcal{X} into finite partitions $P = \{P_1, P_2, \dots\}$ and $Q = \{Q_1, Q_2, \dots\}$ respectively. Hereafter, we assume $\mathcal{X} \subset \mathbb{C}^n$, where n is the dimension of the *complex* semantic space (Each transmitted data is represented with n complex semantic symbols). For any atom P_i of partition P and Q_j of partition Q , let $\mu_i \in \mathcal{P}(P_i)$ and $\nu_j \in \mathcal{P}(Q_j)$, where $\mathcal{P}(A)$ is the set of all probability measures over A . For any message m mapped to $x \in P_i$ by the source language, we seek a mapping $T : \mathcal{X} \rightarrow \mathcal{X}$, which transport x into the corresponding atom Q_j in the target language, i.e. such that $T(x) \in Q_j$. Specifically, we seek a transport map T that pushes μ_i onto ν_j , i.e. $T_{\#}\mu_i = \nu_j$, where $T_{\#}\mu_i$ is the *image measure* of μ_i by T , defined such that $T_{\#}\mu_i(A) = \mu_i(T^{-1}(A))$, $\forall A \subset \mathcal{X}$. Let $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}^+$ be a metric (e.g. Euclidean distance). The optimal transport map, when it exists, can be found by solving the following Monge transportation problem:

$$\arg \min_T \int_{P_i} d(x, T(x)) d\mu_i(x), \text{ s.t. } T_{\#}\mu_i = \nu_j. \quad (5)$$

The existence of such map is not always guaranteed as when for e.g. μ_i is supported on Dirac distributions and ν_j is not.

Kantorovitch relaxation. The Kantorovitch problem seeks a general coupling $\gamma \in \Gamma_{i,j}$ between P_i and Q_j by solving:

$$\arg \min_{\gamma \in \Gamma_{i,j}} \int_{P_i \times Q_j} d(x, y) d\gamma(x, y). \quad (6)$$

Here, $\Gamma_{i,j}$ is the set of all probabilistic measures in $\mathcal{P}(P_i \times Q_j)$ with marginals μ_i and ν_j . In contrast to T , the optimal coupling γ exists but may not be unique [11].

Case of discrete distributions. Usually the exact distribution μ_i or ν_j is not available. Only collections of samples from those distributions are accessible. Let $\mathbf{X}_i = \{x_{i,k} \in P_i, k = 1, \dots, N_{\mathbf{X}}\} \sim \mu_i$ and $\mathbf{Y}_j = \{y_{j,k} \in Q_j, k = 1, \dots, N_{\mathbf{Y}}\} \sim \nu_j$ define two finite samples from atom P_i and Q_j . The corresponding empirical distribution can be written as $\mu_i = \sum_{k=1}^{N_{\mathbf{X}}} p_k \delta_{x_{i,k}}$ and $\nu_j = \sum_{k=1}^{N_{\mathbf{Y}}} q_k \delta_{y_{j,k}}$, where δ_x denotes a unit point mass (Dirac) at the point $x \in \mathbb{C}^n$; p_k and q_k are probability masses associated to the k -th sample in \mathbf{X}_i and \mathbf{Y}_j , respectively, and belong to the probability simplex. The set of probabilistic coupling between those two distributions is the set of doubly stochastic matrices $\Gamma_{i,j}$ defined as:

$$\hat{\Gamma}_{i,j} = \{\gamma \in (\mathbb{R}^+)^{N_{\mathbf{X}} \times N_{\mathbf{Y}}} \mid \gamma \mathbf{1}_{N_{\mathbf{Y}}} = \mu_i, \gamma^T \mathbf{1}_{N_{\mathbf{X}}} = \nu_j\}, \quad (7)$$

where $\mathbf{1}_n$ is a n -dimensional vector of ones. The Kantorovich formulation of the optimal transport reads:

$$\arg \min_{\gamma \in \hat{\Gamma}_{i,j}} \langle \gamma, \mathbf{D} \rangle_{\mathcal{F}}, \quad (8)$$

where $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ denote the Frobenius dot product and $\mathbf{D} \in (\mathbb{R}^+)^{N_{\mathbf{X}} \times N_{\mathbf{Y}}}$ is the cost matrix related to the metric $d(\cdot, \cdot)$.

B. Learning semantic transport maps

Once γ is obtained, the source samples \mathbf{X}_i can then be transported into the target distribution using coupling-dependent barycentric mapping T_γ , which transport source sample \mathbf{X}_i into the convex hull of target sample \mathbf{Y}_j . When d is the squared ℓ_2 distance *i.e.*, $d(x, y) = \|x - y\|_2^2$ and the samples are drawn i.i.d from μ_i and ν_j , which we now assume, the mapping can be written in the following matrix form [12]:

$$T_\gamma(\mathbf{X}_i) = \text{diag}(\gamma \mathbf{1}_{N_{\mathbf{Y}}})^{-1} \gamma \mathbf{Y}_j = N_{\mathbf{X}} \gamma \mathbf{Y}_j. \quad (9)$$

Hence, the mapping T_γ depends on the coupling, which needs to be computed for every new data sample to be transported. To address this problem, similar to [12], we propose to jointly learn the probabilistic coupling $\gamma \in \hat{\Gamma}_{i,j}$ and a transformation $T \in \mathcal{T}$, coupling-free, which approximates the barycentric mapping T_γ . However, in our proposed language modelling, semantics are associated to the atoms of language partition. While the source language maps data $m \in \mathcal{M}$ into its semantic representation $x \in P_i$, the target language language maps it into $y \in Q_j$. Thus, what matters is to find a transformation T , which correctly transports x into the equivalent semantic atom, $T(x) \in Q_j$, where it can be correctly interpreted by the target language. Perfect alignment through the transformation T of (the points of) these two atoms is not needed: there is no need to guarantee $y = T(x)$ but rather that $\rho_{P_i \rightarrow Q_j}(T) = 1$. Therefore, we propose to solve the following problem:

$$\mathbf{P1}: \arg \min_{\gamma \in \hat{\Gamma}_{i,j}, T \in \mathcal{T}} f_{i,j}^{(r)}(\gamma, T) + \beta h(T), \quad (10)$$

where $\|\cdot\|_{\mathcal{F}}$ and denote the Frobenius norm; $\alpha, \beta > 0$ are some hyper-parameters that trade-off the optimization; $h(T)$ is a regularization term to ensure a better generalization of T . Here, in contrast to [12], we define $f_{i,j}^{(r)}(\gamma, T) = \|T(\mathbf{X}_i) - N_{\mathbf{X}} \gamma \mathbf{B}_r(\mathbf{Y}_j)\|_{\mathcal{F}}^2 + \alpha \langle \gamma, \mathbf{D} \rangle_{\mathcal{F}}$, where $\mathbf{B}_r(\mathbf{Y}_j)$ denotes a *barycentric mapping* of the samples in \mathbf{Y}_j into a ball of center c and radius r , contained in the convex hull of \mathbf{Y}_j , as

shown in Fig. 3. The center c of this ball can but is not required to be the centroid of \mathbf{Y}_j , which we adopt in the following. Thus, in contrast to [12], we seek a transport map T that transports samples \mathbf{X}_i from atom P_i of partition P into a ball $\mathbf{B}_r(\mathbf{Y}_j)$ of the target sample \mathbf{Y}_j ¹. The radius of this ball controls the confidence at the receiver in interpreting the meaning intended by the sender. It also robustifies transmitted semantic symbols against syntactic channel noise. In the following, we define $\mathbf{B}_r(\mathbf{Y}_j) = c + r(\mathbf{Y}_j - c)$, where $c = \mathbb{E}_{y \sim \nu_j}[\mathbf{Y}_j]$. In particular, $\mathbf{B}_1(\mathbf{Y}_j) = \mathbf{Y}_j$, in which case, Eq. (10) falls back to the original formulation in [12]. Also we consider only linear transformations: $\mathcal{T} = \{T, \exists \mathbf{A} \in \mathbb{C}^{n \times n}, \mathbf{b} \in \mathbb{C}^n, \forall x \in \mathcal{X}, T(x) = \mathbf{A}x + \mathbf{b}\}$. We show that non-linear language mismatch can be compensated by operating a low-complexity codebook of biased linear transformations.

IV. SEMANTIC CHANNEL EQ USING A CODEBOOK

A. Design of a codebook of transformations

Let $P = \{P_1, P_2, \dots\}$ and $Q = \{Q_1, Q_2, \dots\}$ define two language partitions. We now assume that $\kappa(\cdot)$ defined in (4) is either a one-to-one or a many-to-one mapping. In this case, $\forall i \in J_P, \kappa(i)$ is a singleton, allowing us to denote with $Q_{\kappa(i)}$ the equivalent atom in the target language (target atom) of the source atom P_i . Then, we construct a codebook of transformations

$$\mathcal{C}_{\mathcal{T}} = \{T_i = \text{solve}(\mathbf{P1}), \forall (i, \kappa(i)), i \in J_P\}, \quad (11)$$

obtained by solving Problem (10) for each couple $(i, \kappa(i))$ of source and target atom. Thus, the size of the codebook coincides with N_P , the number of semantic atoms of the source language, and does not depend on N_Q . To ease the notation, in the following, we write $\rho_i(T_k) = \rho_{P_i \rightarrow Q_{\kappa(i)}}(T_k)$, $\forall T_k \in \mathcal{C}_{\mathcal{T}}$, and, we define $\boldsymbol{\rho} = (\rho_i(T_k))_{i,k \in J_P}$, referred to as the *information transfer matrix*, which captures the codebook efficiency. Eventually, we define the entropy of the codebook as:

$$H_\mu(\mathcal{C}_{\mathcal{T}}) = -\frac{1}{N_P} \sum_{T \in \mathcal{C}_{\mathcal{T}}} \rho_i(T) \log(\rho_i(T)) + (1 - \rho_i(T)) \log(1 - \rho_i(T)). \quad (12)$$

B. Bayes-optimal selection of Codebook transformations

We focus on semantic channel pre-equalization. In this case, let $m \in \mathcal{M}$ be the observation that the sender k wants to convey to the receiver l . Instead of transmitting $x = \lambda_k(m)$, the sender intentionally distorts meaning intended at the receiver by transmitting another semantic symbol $y = T(x)$ obtained by appropriately selecting a transformation $T \in \mathcal{C}_{\mathcal{T}}$. Now, we are interested in a probabilistic selection strategy $\pi \in \Pi$, from a class of parametric probabilistic functions Π , such that $\pi(m, T)$ specifies the probability for selecting a transformation $T \in \mathcal{C}_{\mathcal{T}}$ given data $m \in \mathcal{M}$. To do so, we define

$$R(\pi, m) = 1 - \mathbb{E}_{T \sim \pi(m, T)} \left[\sum_{i \in J_P} \mu(P_i | m) \cdot \rho_i(T) \right], \quad (13)$$

where $\mu(P_i | m)$ is the probability that the semantic symbol $x = \lambda_k(m)$ associated to data m belongs to atom P_i of the

¹This is equivalent to defining a new metric $d(x, y) = \|x - \mathbf{B}_r(y)\|_2^2$

sender language partition P . Here, $R(\pi, m)$ quantifies the *a priori* (before syntactic channel) risk of misinterpretation at the RX by the target language when we employ the strategy π for transmitting data m . In particular, $R(\pi, m) \in [0, 1]$, where $R(\pi, m) = 0$ indicates zero risk of misinterpretation at the RX when strategy π is employed for transmitting m (as opposed to $R(\pi, m) = 1$). Then, we propose the following Bayes optimization to minimize the average risk of misinterpretation:

$$\mathbf{P2} : \arg \min_{\pi \in \Pi} \mathbb{E}_{m \sim p_{\mathcal{M}}(m)} [R(\pi, m)], \quad (14)$$

where $p_{\mathcal{M}}(m)$ is the probability that the source \mathcal{M}_k emits data m . Problem (14) can be efficiently and optimally solved by opportunistically maximizing $R(\pi, m)$ for each transmitted data m using Algorithm 1 to obtain a deterministic policy. Future work will explore stochastic transformation selection, by learning a unified parametric policy using *e.g.* RL.

Algorithm 1 Transformation selection

- S1.** Set the codebook $\mathcal{C}_{\mathcal{T}} = \{T_i, \forall i \in J_P\}$.
 - S2.** Observe new data m and let $\mathbf{u} = [\mu(P_i|m), \forall i \in J_P]^T$.
 - S3.** Compute $i^* = \arg \max_{i \in J_P} \rho^T \mathbf{u}$.
 - S4.** Select and apply T_{i^*} .
-

V. NUMERICAL RESULTS

We adopt the colour-MNIST dataset [5], consisting of images of digits of size $3 \times 28 \times 28$ randomly colored red, green, blue, and yellow. In this context, we consider our reference scenario of Fig. 1 where two TXs k and l observe the pixels value of random image, while the colour label and digit value are hidden, and encode underlying information for their respective RXs k and l , which try to infer the digit on the image or its parity. Accordingly, λ_k partitions the semantic space \mathbb{C}^n into $N_P = 10$ atoms corresponding to the 10 digits and λ_l into $N_Q = 2$ atoms. While images colour information is not used in interpretation, it introduces semantic noise, bringing additional complexity in semantic symbols representation. We use $N_Y = 5N_X = 5000$ image samples and set $r = 1$ (default), $n = 2$, thus, encoding each colour image with only two complex semantic symbols. We assume language generators and interpreters are given, modeled as auto-encoders, and obtained after an end-to-end learning process, minimizing a cross-entropy loss as in [2]. Language generators are composed of a convolutional layer of dimension $3n$, followed by a rectifier linear unit and a max-pool operator, another convolutional layer with dimension $8n$, followed by another max pooling operator, a dropout layer with probability 0.5, and a rectifier linear unit layer, whose output is flattened to feed a final linear layer of dimension $2n$. In contrast, each language interpreter only includes one linear layer of dimension $2n$. In addition, all convolutional layers use a kernel size equal to 5, and the max pool operators use a 2-sized kernel and stride. The power of the complex symbols x outputted by language generators is normalized so that $\mathbb{E}[|x|^2] = 1$. Once trained, the accuracy of the communication between TX k and RX k under error-free syntactic channel is $A_k = 78.4\%$, whereas for TX l and RX l it is $A_l = 94.4\%$. The 100% accuracy is not attained in particular due to the noise in data and the limited expressivity of the language generator

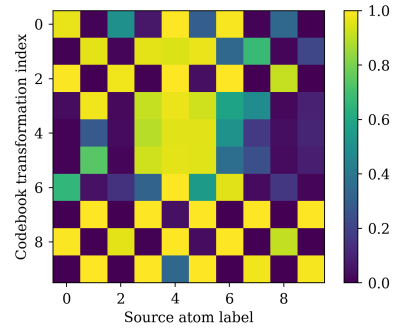


Fig. 4: Information transfer matrix ρ .

and interpreter. Now, we are interested in the communication between TX k and RX l , grounded to semantic errors, as they are not trained to match each other logic: their semantic channel is thus opaque, requiring equalization.

Operational zone of the codebook. We assess the operational zone of the codebook constructed by solving Eq. (11) using the Algorithm proposed in [12] by setting $\alpha = \frac{0.1nN_X}{\max(\mathbf{D})}$, $\beta = 10^{-8}n^{-1}N_X$, and $h(T) = \|\mathbf{A} - \mathbf{I}\|_{\mathcal{F}}^2$, where \mathbf{I} is the identity matrix of size n . Then, we compute the associated information transfer matrix ρ shown on Fig. 4. In practice, this is done *offline* using a Monte-Carlo sampling. From Fig. 4, the yellow diagonal in the matrix indicates that the optimized transformations in the codebook are successful in transporting samples from source atoms to target atoms. Fig. 4 also suggests that there exist equivalent transformations in the codebook. In other words, for a given source atom i and its associated target atom $\kappa(i)$, there exists more than one transformation with equal probability for transporting samples from P_i into $Q_{\kappa(i)}$. For instance, for the atom with label 0, the transformation at index 0, 2, and 8 exhibits similar performance. Similarly, there exist orthogonal transformations. For instance, the transformations at index 1, 3, 4, 5, 7, and 9 cannot accurately transport the sample from the source atom 0, to the associated target atom. These latter results suggest that the complexity can be further reduced by efficiently pruning transformations, which we will address in future work.

The need for compensating language mismatch. To show the effectiveness of our proposed equalization method, we compare it to different benchmarked approaches (see Fig. 5):

ClassCom: In the classical approach, TX k sends binary encoded images to RX l , which infers the parity information from the received noisy images. In method **A**, the RX first infers the digit by applying $I_{\lambda_k} \circ \lambda_k(\cdot)$ and then decodes its parity, whereas in method **B** it directly infers the parity by applying $I_{\lambda_l} \circ \lambda_l(\cdot)$. This approach employs quadrature amplitude modulation (QAM) of order 256, resulting in approximately $4.6n$ symbols per transmitted image. Obviously, method **B** exhibits better performance than **A**, as it does not require decoding the digit first. However, the two classical approaches are strongly impacted by noise in the syntactic channel. At 0 dB SNR, our proposed solution reaches 90% accuracy compared to 50% accuracy for *ClassCom* methods, which corresponds to random guessing of the digit parity. Even under error-free syntactic channel, the accuracy of method **B** caps to $A_l = 94.4\%$ whereas for method **A**, it is 88%, corresponding

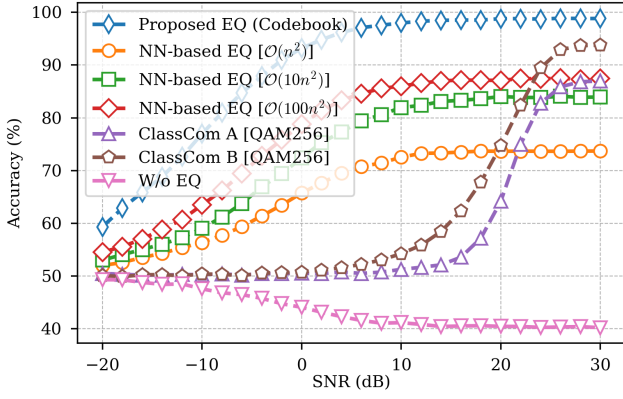


Fig. 5: Performance comparison of equalization techniques.

$\approx A_k + \frac{4}{9}(1 - A_k)$. Indeed, if the RX misinterpreted the digit, the probability that it misinterprets the parity is $\frac{5}{9}$. This result further confirms the value of semantic-based communication, which requires, however, semantic equalization mechanisms.

SemCom without EQ: The semantic channel is not equalized; that is, the codebook contains only the identity transformation. In this context, Fig. 5 shows that under an error-free syntactic channel, a communication between TX k and RX l results in 40.4% accuracy. This performance is even below the 50% accuracy expected for a random guess of the digit parity by RX l from a uniform distribution of symbols in the received constellation, which can be obtained by noising the syntactic channel. This poor accuracy is a result of the mismatch of semantic representation space and motivates the need for semantic channel equalization [2].

SemCom with NN-based EQ: The codebook contains only one transformation, modeled as an NN comprising one linear layer of dimension $2n$, thus having substantially the same complexity (expressed as the number of parameters) as our proposed linear transformations *i.e.* $\mathcal{O}(n^2)$. We train the NN by minimizing the cross-entropy loss, in an end-to-end fashion using the language generator λ_k together with the language interpreter I_{λ_l} . During the training, the weights of the language generator and interpreter are frozen so that they are not impacted by the learning procedure. Fig. 5 shows that, at 0 dB SNR, our proposed solution exhibits 30% accuracy improvement compared to the NN-based equalizer. Also, we evaluate the case where the NN-based equalizer is τ -times more complex (*i.e.* has τ -times more weights). In this case, it comprises two feed-forward layers of dimension τn and $2n$, respectively. Even when the complexity of the NN increases, the accuracy of the communication does not exceed 88%.

From semantic to communication effectiveness. In Fig. 6, we evaluate the impact of the ball radius r in Eq. (10) on the equalization performance. When r tends to zero, the accuracy of the communication increases as it becomes less sensitive to syntactic channel noise. Conversely, the entropy of the codebook decreases indicating a loss of diversity in the base semantic representation. Indeed, decreasing r concentrates the points of the target atoms, resulting in a seamless quantization. Consequently, we lose the information about the specificities (*e.g.* digit font or style) of each image projected into this atom.

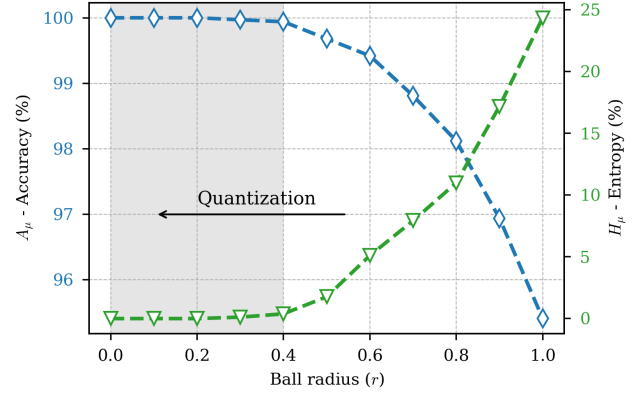


Fig. 6: Tradeoff semantic vs. effectiveness.

VI. CONCLUSION

We proposed a method for equalizing semantic channels in multi-user semantic communications, where senders and receivers employ distinct languages. In our proposed solution, language mismatch is first modeled by means of measurable transformations over semantic representation spaces. Then, we proposed a method for learning a codebook of transformations based on optimal transport theory, which models such transformations as transportation maps. Hence, the problem of semantic channel equalization translates to selecting the appropriate transformation in the codebook, which we achieve with a Bayes-optimal selection strategy. Overall, the proposed method is robust against both syntactic and semantic channel noise, improving transmission accuracy with reduced complexity compared to traditional approaches.

REFERENCES

- [1] E. Calvanese Strinati and S. Barbarossa, “6G networks: Beyond Shannon towards Semantic and Goal-Oriented Communications,” *Computer Networks*, vol. 190, p. 107930, 2021.
- [2] M. Sana and E. Calvanese Strinati, “Learning Semantics: An Opportunity for Effective 6G Communications,” in *proc. Consumer Communications & Networking Conference (CCNC)*, 2022, pp. 631–636.
- [3] G. Shi, Y. Xiao *et al.*, “From Semantic Communication to Semantic-Aware Networking: Model, Architecture, and Open Problems,” *IEEE Communications Magazine*, vol. 59, no. 8, pp. 44–50, 2021.
- [4] M. K. Farshbafan, W. Saad, and M. Debbah, “Curriculum learning for goal-oriented semantic communications with a common language,” *IEEE Trans. on Communications*, vol. 71, no. 3, pp. 1430–1446, 2023.
- [5] J. Foerster, I. A. Assael *et al.*, “Learning to communicate with deep multi-agent reinforcement learning,” *Advances in neural information processing systems*, vol. 29, 2016.
- [6] Z. Ding, T. Huang, and Z. Lu, “Learning Individually Inferred Communication for Multi-Agent Cooperation,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 22 069–22 079, 2020.
- [7] T. Lin, J. Huh, C. Stauffer, S. N. Lim, and P. Isola, “Learning to Ground Multi-Agent Communication with Autoencoders,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 15 230–15 242, 2021.
- [8] P. Wang, J. Wang, P. Paranamana, and P. Shafto, “A Mathematical Theory of Cooperative Communication,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 17 582–17 593, 2020.
- [9] M. Sana, N. di Pietro, and E. Calvanese Strinati, “Transferable and Distributed User Association Policies for 5G and Beyond Networks,” in *Proc. IEEE PIMRC*, 2021, pp. 966–971.
- [10] S. Sinha, U. Vaidya *et al.*, “On Information Transfer in Dynamical Systems with applications in Control of Non-equilibrium Dynamics,” in *proc. Indian Control Conference (ICC)*, 2019, pp. 326–331.
- [11] C. Villani, *Optimal transport: old and new*. Springer, 2009, vol. 338.
- [12] M. Perrot, N. Courty, R. Flamary, and A. Habrard, “Mapping Estimation for Discrete Optimal Transport,” *Advances in Neural Information Processing Systems*, vol. 29, 2016.