Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

AgriBridge-X : From Bilingual Translation to Lexicon-Guided Knowledge Extraction in Web-Based Agricultural Conversations

Thanh Ma

mtthanh@ctu.edu.vn

Can Tho University

Phu-An Thai

Can Tho University

Tri-Phuc Nguyen

Can Tho University

Thai Le

Can Tho University

Quoc-Khang Tran

Can Tho University

Mathieu Roche

French Agricultural Research Centre for International Development (CIRAD)

Raphaelle Ducrot

French Agricultural Research Centre for International Development (CIRAD)

Research Article

Keywords: Conversational Data Mining, Knowledge Extraction, Lexicon-Guided Methods, Machine Translation, Knowledge Representation

Posted Date: May 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9589749/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

AgriBridge-X : From Bilingual Translation to Lexicon-Guided Knowledge Extraction in Web-Based Agricultural Conversations

Thanh Ma[1], Phu-An Thai[1†], Tri-Phuc Nguyen[1†], Thai Le[1†], Quoc-Khang Tran[1†], Mathieu Roche[2,3†], Raphaelle Ducrot[2,4†]

[1]CICT, Can Tho University, Can Tho city, Vietnam.
[2]French Agricultural Research Centre for International Development (CIRAD), France.
[3]TETIS, Université de Montpellier, AgroParisTech, CIRAD, INRAE, Montpellier, France.
[4]G-EAU, Université de Montpellier, AgroParisTech, BRGM, CIRAD, INRAE, IRD, Institut Agro, Cambodia.

Contributing authors: mtthanh@ctu.edu.vn; contact.thaiphuan@gmail.com; ntphuc@ctu.edu.vn; lnthai@ctu.edu.vn; tqkhang@ctu.edu.vn; mathieu.roche@cirad.fr; raphaelle.ducrot@cirad.fr;
†These authors contributed equally to this work.

Abstract

Web-based agricultural conversations constitute a rich yet underutilized source of knowledge, capturing interactions between farmers and scientists across online platforms. However, such data remain unstructured, linguistically heterogeneous, and difficult to systematically exploit. In this paper, we extend AgriBridge , a bidirectional translation system, into a framework for transforming conversational data into structured knowledge representations. We argue that translation alone is insufficient for effective knowledge utilization and instead treat it as a normalization step for downstream information extraction. We propose a lexicon-guided extraction module that identifies domain-specific entities, including problems, practices, actors, and impacts, from translated text. These entities are enriched with contextual representations and organized into structured forms, enabling systematic analysis of conversational knowledge. The framework integrates speech-to-text processing, sequence-to-sequence translation, and expert-guided extraction into a scalable pipeline for dialect-rich environments.

Experiments demonstrate the feasibility of converting unstructured dialogues into structured knowledge representations. This work advances conversational AI toward knowledge-centric systems for large-scale knowledge organization.

Keywords: Conversational Data Mining, Knowledge Extraction, Lexicon-Guided Methods, Machine Translation, Knowledge Representation

1 Introduction

The Mekong Delta, one of the most critical agricultural regions in Southeast Asia, is undergoing rapid environmental and socio-economic transformations [1, 2], including saltwater intrusion, land subsidence, and climate-induced variability. In response, a wide range of agroecological practices and adaptive solutions are continuously developed by both scientific communities and local farmers. Despite this abundance of innovations, the knowledge surrounding these practices remains highly fragmented across heterogeneous sources, including academic publications, technical reports, and increasingly, web-based conversational media such as online forums, social networks, and video platforms. This fragmentation poses a fundamental barrier to large-scale knowledge dissemination and limits the collective capacity to leverage existing innovations.

A central challenge underlying this issue is the disconnect between scientific knowledge systems and farmer-generated knowledge [3]. Scientific insights are typically expressed in formal, technical language, often in English or standardized Vietnamese, whereas farmers communicate using colloquial, dialect-rich expressions grounded in local experience. This linguistic and conceptual mismatch significantly constrains bidirectional knowledge exchange, impeding not only the accessibility of scientific knowledge but also the systematic capture of experiential insights emerging from practice. Recent advances in natural language processing [4, 5] have enabled the development of domain-specific translation systems designed to bridge this gap. In particular, our previous work introduced *AgriBridge* [6], a bidirectional translation framework that converts scientific discourse into farmer-friendly language and vice versa, leveraging a fine-tuned sequence-to-sequence architecture and a curated bilingual corpus [7]. While such systems effectively address linguistic barriers and improve communication, they primarily operate at the level of *language transformation* [8]. As a result, they do not explicitly address a more fundamental challenge: *how can translated conversational data be transformed into structured knowledge representations that support systematic analysis and reuse?*

This limitation becomes particularly critical in the context of web-based agricultural data [9, 10]. Increasingly, valuable knowledge is embedded in large-scale conversational content, such as expert-farmer discussions in online videos and social platforms. These conversations encode rich domain signals, including descriptions of problems, applied practices, contextual conditions, and observed outcomes. However, such information remains latent, implicit, and unstructured, making it difficult to extract, organize, and exploit for downstream applications such as knowledge retrieval

or decision support [11, 12]. To address this gap, we extend AgriBridge beyond its original role as a translation system into a unified framework for *lexicon-guided knowledge extraction* [13, 14], from agricultural conversations. Our key insight is that translation should be viewed as an intermediate normalization step rather than an end goal. Specifically, once linguistic variability is reduced through translation, conversational data can be further processed to extract structured entities grounded in domain knowledge.

To this end, we introduce a lexicon-guided extraction module that identifies domain-relevant entities [15], including agricultural problems, practices, actors, and impacts, from translated conversational data [16]. These entities are further enriched with contextual representations and organized into structured forms that enable systematic analysis of conversational knowledge [17]. By explicitly incorporating expert-defined lexicons [18], our approach introduces domain-aware inductive bias into the extraction process, improving both interpretability and robustness in low-resource and dialect-rich environments. Moreover, our framework emphasizes controllability and transparency, which are critical for knowledge-intensive web applications. Furthermore, by operating on conversational data collected from web platforms, the proposed system supports scalable processing of continuously evolving knowledge sources, enabling a more comprehensive and dynamic understanding of agricultural practices.

In summary, this paper makes the following contributions:

1. We extend the AgriBridge system [6] from a bidirectional translation platform to a *web-based knowledge extraction framework* [19] for agricultural conversations.
2. We propose a *lexicon-guided information extraction module* that identifies key domain entities (e.g., problems, practices, actors, impacts) from translated text.
3. We introduce a *context-aware representation*context of extracted entities [20], enabling the transformation of unstructured conversational data into structured knowledge.
4. We demonstrate the feasibility of our approach on real-world agricultural conversations [21], highlighting its potential for supporting data-driven knowledge analysis and decision support.
5. We formalize the transition from translation-based normalization to structured knowledge extraction, providing theoretical guarantees linking semantic preservation to entity-level stability.

By bridging the gap between unstructured conversational data [22] and structured knowledge representations, our work contributes toward a new class of intelligent systems that go beyond communication support to enable systematic organization and utilization of agricultural knowledge at scale. A key challenge in conversational agricultural data lies in the ambiguity and context-dependence of expressions. The same term may correspond to different semantic roles depending on context, while important information is often implicit or distributed across multiple utterances. This makes structured extraction significantly more challenging than standard text processing tasks.

The remainder of this paper is structured as follows: Section 2 provides the background and positioning of this work, focusing on heterogeneous agricultural knowledge

systems, the limitations of translation-centric pipelines, and the role of lexicon-guided extraction. Section 3 presents the problem formulation for converting conversational inputs into structured entities and graph-based knowledge representations. Section 4 introduces the AgriBridge-X framework, detailing its translation-based normalization stage, lexicon-guided extraction stage, theoretical properties, and algorithmic realization. Section 5 describes the experimental setting and discusses the results on translation, entity extraction, ablation analysis, qualitative cases, and practical applications. Section 6 concludes the paper and outlines future research directions.

2 Background

2.1 Recent Advances in AI for Agricultural Systems

Recent advances in artificial intelligence (AI) have significantly transformed agricultural systems, enabling applications such as crop monitoring, disease detection, yield prediction, and decision support [23, 24]. Deep learning models, including convolutional and transformer-based architectures, have been widely adopted to process multimodal agricultural data, such as images, sensor signals, and textual reports. These approaches support precision agriculture by enabling large-scale data-driven analysis and automation, thereby improving productivity and sustainability.

Despite these advances, agricultural knowledge systems remain inherently heterogeneous [25, 26], encompassing both formal scientific knowledge and experiential farmer knowledge. In regions such as the Mekong Delta, this heterogeneity is further amplified by linguistic diversity, where scientific discourse is typically expressed in English or standardized Vietnamese, while farmers communicate using dialect-rich and context-dependent expressions. This mismatch creates a fundamental barrier to knowledge exchange and limits the scalability of innovation dissemination [27].

2.2 Large Language Models and Conversational AI in Agriculture

More recently, large language models (LLMs) have been increasingly applied to agricultural applications, including question answering, conversational advisory systems, and knowledge retrieval [28, 29]. These systems aim to bridge the gap between farmers and experts by enabling natural language interaction and facilitating access to domain knowledge.

However, existing LLM-based approaches primarily focus on response generation and interaction quality, rather than explicitly structuring knowledge into reusable representations. As a result, although they improve accessibility and communication, they do not directly address the challenge of transforming conversational data into structured knowledge suitable for downstream analysis and decision support.

To address linguistic barriers, prior work has introduced AgriBridge [6], a domain-specific translation framework that enables bidirectional communication between farmers and scientists. By integrating speech processing and sequence-to-sequence translation models [30], AgriBridge effectively reduces linguistic variability and

facilitates knowledge flow across communities. However, similar to LLM-based conversational systems, these approaches primarily operate at the level of language transformation [31, 32], without explicitly structuring the underlying knowledge.

2.3 Neural Models for Domain-Specific and Low-Resource Translation

Recent advances in neural machine translation (NMT) [31, 8, 6] have been driven by transformer-based architectures, particularly sequence-to-sequence models such as T5 [33], BART [34], and multilingual models like mBART and mT5 [35]. These models have demonstrated strong performance across a wide range of translation tasks, including low-resource and domain-specific settings.

In the context of agricultural and dialect-rich data, multilingual transformer models [33] are particularly relevant due to their ability to leverage cross-lingual representations [36] and generalize across linguistic variations. Recent studies have shown that fine-tuning pretrained language models on domain-specific corpora significantly improves performance [23, 37, 9] in specialized translation tasks. However, most existing works focus on standard language pairs and well-structured text, with limited attention to conversational data [21, 22] characterized by dialectal variation and informal expressions. This limitation motivates the evaluation of multiple neural architectures within our framework, including sequence-to-sequence [30, 35] and multilingual transformer models, to assess their robustness in academic–farmer translation scenarios [6].

In this work, we systematically evaluate different model families under a unified dataset and pipeline, providing insights into their suitability for dialect-rich, low-resource agricultural conversations.

2.4 Limitations of Existing AI and Translation-Centric Pipelines

From a knowledge systems perspective, both translation-based systems and LLM-based conversational models remain limited in their ability to support large-scale knowledge utilization. Even after translation or generation, conversational data remain unstructured sequences of text, lacking explicit representations of entities, relations, and semantic roles. However, despite these advances [23, 24, 28, 29], existing approaches remain limited in their ability to structure conversational knowledge.

This limitation is particularly critical in web-based environments [9, 10], where agricultural knowledge is increasingly embedded in conversational media such as online videos, forums, and social platforms. These conversations encode valuable domain signals, including descriptions of problems, applied practices, contextual conditions, and observed outcomes. However, such information is often implicit, fragmented, and context-dependent, making it difficult to extract, organize, and exploit for downstream applications such as knowledge retrieval or decision support [11].

More importantly, existing AI pipelines typically assume clean and well-structured inputs. This assumption does not hold in real-world agricultural settings, where data

are noisy, dialect-rich, and low-resource. Consequently, both neural and LLM-based approaches struggle to generalize effectively in such environments.

2.5 From Linguistic Bridging to Knowledge Structuring

To address these challenges, we argue that translation should be viewed not as an end goal, but as an intermediate normalization step within a broader knowledge processing pipeline. Specifically, conversational understanding can be decomposed into two complementary stages:

- **Linguistic normalization** [36], which reduces variability in dialect, style, and noise;
- **Semantic structuring** [38, 39], which transforms normalized text into structured knowledge representations.

This separation of concerns is essential for handling noisy, multilingual, and context-dependent conversational data. By isolating linguistic variability from semantic modeling, it enables more robust downstream processing and improves the reliability of extracted knowledge. Under this perspective, translation serves as a bridge that facilitates structured extraction, rather than as a standalone solution.

2.6 Lexicon-Guided Extraction in Domain-Specific Contexts

In domain-specific applications such as agriculture, purely data-driven extraction approaches often face challenges due to limited annotated data and high linguistic variability. Lexicon-guided methods [13, 14] provide an effective alternative by incorporating expert knowledge into the extraction process. By leveraging curated vocabularies for key domain concepts [15, 18], such as diseases, practices, actors, and impacts, these approaches constrain the hypothesis space and improve both precision and interpretability.

This is particularly important in real-world conversational settings, where extracted information must be reliable, verifiable, and aligned with domain knowledge. Moreover, when combined with contextual representations [20], lexicon-guided extraction enables the disambiguation of semantically ambiguous expressions and the recovery of implicit information embedded in dialogue.

In the context of this work, we use the STAR-Farm lexicon [16]. This lexical resource was constructed by synthesizing and adapting an agroecological lexicon [40] and an innovation lexicon [41, 42]. Based on these lexicons, during a multidisciplinary workshop[1] organized on October 22–23, 2025, at Can Tho University (Vietnam), we selected 6 main concepts (i.e., Problem issues, Actors, Practices, Knowledge, Characterizing, Impact) and 553 terms associated to form the STAR-Farm lexicon [16].

[1] https://agritrop.cirad.fr/616229/

2.7 Research Gap and Positioning

Despite significant advances in AI for agriculture [23, 24, 28, 29], existing approaches remain fragmented across three main directions: (i) deep learning systems for perception and prediction, (ii) LLM-based conversational systems for interaction, and (iii) information extraction methods for structured data processing.

However, these approaches are rarely integrated. LLM-based systems improve interaction but lack structured representations, while extraction methods assume clean textual inputs and fail in conversational settings. Translation-based systems, although effective for linguistic bridging [6, 8], do not explicitly address knowledge structuring.

To the best of our knowledge, there is no unified framework that integrates: (i) translation-based normalization for handling dialect-rich conversational data, (ii) domain-aware extraction guided by expert knowledge, and (iii) structured knowledge representation for downstream reasoning and analysis.

In this work, we address this gap by extending AgriBridge to a knowledge-centric framework [19] that transforms unstructured conversational data into structured representations. Our approach introduces a lexicon-guided extraction layer operating on normalized text, enabling the systematic identification of domain-relevant entities and their organization into structured knowledge forms [17]. By bridging the gap between conversational data [22] and structured knowledge representations, this work contributes to a new class of intelligent systems that move beyond communication support to enable scalable knowledge organization and analysis.

3 Problem Formulation

Let $\mathcal{D} = \{Q_i^{audio}\}_{i=1}^{N}$ denote a set of conversational inputs collected from web-based platforms, where each Q_i^{audio} represents an audio segment of farmer-scientist interaction.

Following the AgriBridge paradigm, we define a transformation pipeline:

$$X_i = \mathrm{STT}(Q_i^{audio}), \quad \tilde{X}_i = \mathrm{Clean}(X_i), \quad Y_i = T_\Theta(\tilde{X}_i),$$

where STT converts speech to text, Clean reduces transcription noise and normalizes dialect, T_Θ denotes the translation model. This formulation preserves the core AgriBridge pipeline while explicitly separating sources of variability (acoustic, linguistic, dialectal), enabling a principled extension toward structured processing.

3.1 Entity Extraction from Normalized Text

Given a normalized sentence Y_i, we aim to extract a set of structured entities:

$$E(Y_i) = \{e_j\}_{j=1}^{M_i}, \quad e_j = (s_j, t_j, c_j),$$

where s_j is a textual span, $t_j \in \mathcal{T}$ is an entity type, $c_j \in [0, 1]$ is a confidence score.

We define the entity type space:

$$\mathcal{T} = \{P, R, A, I\}$$

representing *Problems*, *Practices*, *Actors*, and *Impacts*. This representation provides a minimal yet expressive abstraction for agricultural knowledge, capturing the essential components required to describe real-world situations while maintaining tractability for extraction models. For instance, in a typical agricultural scenario, a *Problem* (P) may correspond to crop diseases such as "leaf yellowing" or "pest infestation", a *Practice* (R) may refer to actions such as "pesticide spraying" or "fertilizer application", an *Actor* (A) may denote entities such as "farmers" or "agricultural experts", and an *Impact* (I) may capture outcomes such as "yield reduction" or "improved crop growth". These categories are aligned with the domain-specific lexicon used in our framework, ensuring consistency between theoretical formulation and practical extraction. Moreover, for example, given a sentence such as "The rice leaves turn yellow after heavy rainfall," the system may extract: (P) leaf yellowing, (R) *empty*, (A) farmer (implicit), and (I) crop stress, demonstrating how conversational inputs are mapped into structured knowledge units.

To incorporate contextual information, we define the context-enriched entity as:

$$e_j = (s_j, t_j, c_j, \phi(Y_i)),$$

where $\phi(Y_i)$ denotes a semantic embedding of the normalized sentence. In this paper, we refer to $\phi(Y_i)$ as the *context* of the entity, as it encodes the surrounding semantic information required for disambiguation and robust entity typing in conversational settings. This notion of context plays a critical role in reducing ambiguity in entity classification, as formally analyzed in Proposition 6. Entities in conversational data are inherently context-dependent. Incorporating semantic embeddings allows the model to disambiguate meanings and improve robustness under linguistic variability.

Let $\mathcal{L} = \{L_P, L_R, L_A, L_I\}$ denote expert-defined lexicons corresponding to Problems, Practices, Actors, and Impacts. Given a normalized sentence Y_i, the goal of the extraction module is to identify a set of entities

$$E(Y_i) = \{e_j\}_{j=1}^{M_i},$$

where each e_j corresponds to a domain-relevant entity. We formalize this process as:

$$E(Y_i) = \arg\max_{E'} P(E' \mid Y_i, \mathcal{L}),$$

where E' denotes a candidate set of extracted entities. This formulation models entity extraction as a structured prediction problem, in which the objective is to select the most probable set of entities given the input sentence and domain-specific lexicon constraints.

In practice, this corresponds to the *lexicon-guided entity extraction module* illustrated in Figure 1, where candidate spans are evaluated using both contextual representations and lexicon matching. The probability $P(E' \mid Y_i, \mathcal{L})$ is instantiated through a scoring function that combines: (i) contextual compatibility, derived from the embedding $\phi(Y_i)$, and (ii) lexicon-based constraints that enforce domain consistency. This formulation allows the integration of both contextual semantics and expert

knowledge within a unified probabilistic framework. It can be implemented using either neural sequence labeling models or hybrid lexicon-constrained extraction methods. Lexicon-guided extraction introduces domain knowledge as an inductive bias, improving precision and interpretability in low-resource and dialect-rich environments. Concretely, the output $E(Y_i)$ corresponds to the entity set shown in the extraction stage of Figure 1.

3.2 Structured Knowledge Representation

We represent the extracted information as a graph: $G_i = (V_i, R_i)$, where the node set $V_i = E(Y_i)$ corresponds to the set of extracted entities obtained from the lexicon-guided extraction module, and each node represents a context-enriched entity $e_j = (s_j, t_j, c_j, \phi(Y_i))$.

The relation set $R_i = \{(e_j, e_k, r_{jk})\}$ captures implicit semantic associations between entities, inferred from co-occurrence patterns or contextual proximity within the sentence Y_i. These relations do not necessarily correspond to explicit logical predicates, but instead encode weak semantic dependencies that reflect how entities interact in conversational data. This graph-based representation enables the transformation of unstructured conversational inputs into structured knowledge, supporting downstream reasoning, querying, and analysis. For example, given extracted entities such as (P) leaf yellowing and (R) pesticide spraying, a relation r_{jk} may indicate that the practice is applied to address the problem.

4 AgriBridge-X Framework

Terminology. In this paper, we distinguish between two linguistic registers commonly observed in agricultural knowledge exchange. We refer to *academic* (or *scientific*) language as formal, standardized expressions typically used by experts, researchers, and technical documents. In contrast, we refer to *farmer* (or *dialectal*) language as informal, experience-driven expressions used by farmers in conversational settings, often characterized by regional dialects and context-dependent phrasing. For clarity and consistency, we use the terms *academic* and *farmer* throughout this section and in all figures to denote these two registers. This distinction is central to our framework, as it reflects the fundamental gap between expert knowledge and experiential knowledge in real-world agricultural systems.

AgriBridge-X extends the original AgriBridge system from a bidirectional translation framework to a knowledge structuring framework for web-based agricultural conversations. While AgriBridge primarily addresses linguistic bridging between scientific discourse and farmer-oriented language, AgriBridge-X augments this backbone with a lexicon-guided extraction layer that transforms normalized conversational text into structured knowledge representations. Formally, AgriBridge-X is defined as a two-stage framework:

$$\mathcal{F}_{\text{ABX}} = \mathcal{F}_{\text{extract}} \circ \mathcal{F}_{\text{trans}},$$

where $\mathcal{F}_{\text{trans}}$ denotes the AgriBridge translation pipeline and $\mathcal{F}_{\text{extract}}$ denotes the lexicon-guided entity extraction module. The translation function $\mathcal{F}_{\text{trans}}$ takes a raw conversational input (audio or text) and produces a normalized sentence Y_i by applying

speech-to-text conversion, cleaning, and dialect-aware translation. This stage reduces linguistic variability and maps farmer-style expressions into a standardized academic representation. The extraction function $\mathcal{F}_{\text{extract}}$ then takes the normalized text Y_i as input and outputs a structured entity set $E(Y_i)$ and its corresponding graph representation G_i. This stage identifies domain-relevant entities using lexicon constraints and contextual embeddings, and organizes them into structured knowledge. This formulation clarifies that AgriBridge-X is not a standalone translation system, but a sequential pipeline that combines *linguistic normalization* (via $\mathcal{F}_{\text{trans}}$) with *knowledge extraction* (via $\mathcal{F}_{\text{extract}}$).

Notation clarification. We distinguish between training data and inference outputs. During training, (X, Y) denotes a *parallel bilingual corpus*, where X corresponds to farmer-style (dialectal) expressions and Y corresponds to academic (standardized) expressions. During inference, Y_i denotes the normalized output generated by the translation model from a new input X_i. This distinction avoids ambiguity between training pairs and model outputs.

Figure 1 illustrates the overall architecture of AgriBridge-X , with numbered steps corresponding to each stage of the pipeline.

Training pipeline (upper part). Steps (1)–(3) describe the construction of the bilingual training corpus. In Step (1), web-based agricultural conversations are collected from online sources such as videos and forums. In Step (2), audio data are transcribed and cleaned to produce normalized textual inputs X. In Step (3), these texts are manually rewritten and aligned into a *parallel bilingual corpus* (X, Y), where X represents farmer-style expressions and Y represents their academic counterparts. For example, a farmer-style sentence such as "Lua bi vang la sau mua lon" is aligned with its academic version "Rice leaves turn yellow after heavy rainfall".

Inference pipeline (lower part). Steps (4)–(6) describe the processing of new inputs. In Step (4), a raw input Q_i (audio or text) is provided. In Step (5), the translation module $\mathcal{F}_{\text{trans}}$ produces a normalized sentence

$$Y_i = T_\Theta(\tilde{X}_i),$$

where $\tilde{X}_i$ is the cleaned version of the input. In Step (6), the extraction module $\mathcal{F}_{\text{extract}}$ transforms Y_i into a structured entity set $E(Y_i)$ and its corresponding graph representation G_i. The lexicon $\mathcal{L}$ is **bilingual**, allowing entity matching to be performed either on the source representation or on the normalized text Y_i, thereby improving robustness in dialect-rich settings. This explicit separation clarifies the distinction between training data (X, Y) and inference outputs Y_i, and establishes a clear connection between the figure and the underlying computational pipeline. The constructed dataset will be released publicly via Dataverse to support reproducibility and further research. The framework consists of two main stages as follows:

4.1 Stage 1: Data Preprocessing and Translation Backbone

The first stage preserves the original AgriBridge design and serves as the linguistic normalization backbone. It begins with data preprocessing to construct high-quality bilingual resources for training the translation component.

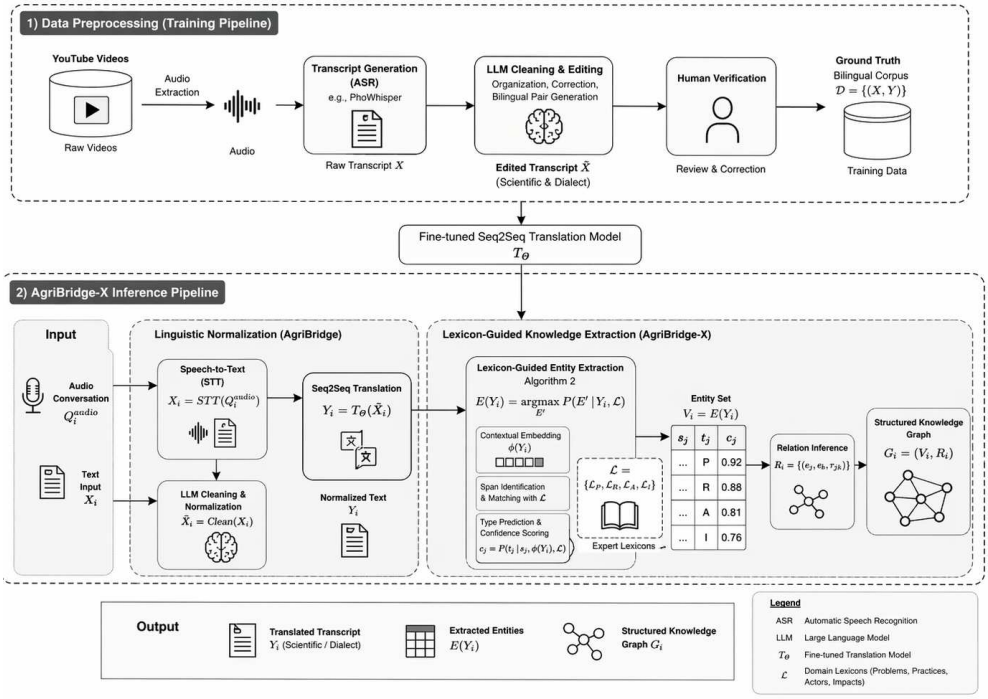


Fig. 1 AgriBridge-X framework. The upper block shows the data preprocessing and training pipeline, while the lower block shows the inference pipeline that integrates linguistic normalization and lexicon-guided knowledge extraction.

Step 1: Video collection and audio extraction. Agricultural conversations are collected from web-based sources such as YouTube videos containing interactions between experts and farmers (see Figure 2). The selection of videos is performed through a semi-manual process. First, candidate videos are identified using keyword-based search queries related to agricultural topics (e.g., crop diseases, pest management, irrigation). Then, a manual filtering step is applied to retain only videos that contain meaningful conversational exchanges between farmers and experts, ensuring domain relevance and sufficient linguistic diversity. The selected videos are required to satisfy the following criteria: (i) the content is related to agricultural practices or problem-solving discussions, (ii) the audio quality is sufficient for speech recognition, and (iii) the conversation exhibits dialectal or informal expressions representative of real-world farmer communication. Audio tracks are then extracted from the selected videos and converted into a standardized format for subsequent processing. This selection protocol ensures the construction of a representative dataset while maintaining data quality for downstream processing.

Step 2: Speech recognition and raw transcript generation. The extracted audio is processed by an automatic speech recognition model, such as PhoWhisper, to generate a raw transcript X_i.

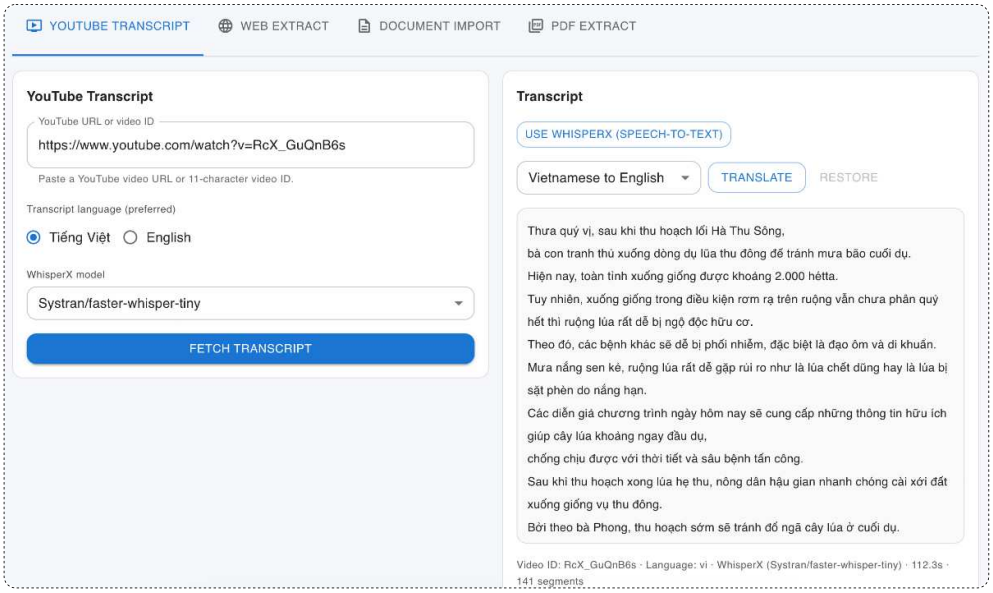


Fig. 2 Text extraction from YouTube videos. The audio track is extracted and processed through speech recognition to generate raw transcripts. These transcripts are then cleaned and manually rewritten into standardized academic language. The original farmer-style expressions and their rewritten versions are aligned to form a parallel bilingual corpus, where each pair represents the same content in dialectal (farmer) and academic registers. This dataset is used to train the translation model.

Step 3: Cleaning, editing, and dataset construction. Because conversational transcripts are often noisy and dialect-rich, the raw transcript is refined through an LLM-based cleaning module, yielding $\tilde{X}_i$. This step removes irrelevant noise, corrects transcription errors, and normalizes dialectal expressions while preserving semantic content. The cleaned transcript is then manually reviewed and aligned into bilingual pairs, producing the ground-truth corpus used to fine-tune the translation model T_Θ.

At inference time, the same backbone is applied to incoming text or audio queries. If the input is audio, speech recognition and cleaning are first applied; if the input is already textual, the pipeline proceeds directly to translation. The normalized text is then obtained as $Y_i = T_\Theta(\tilde{X}_i)$, which serves as the semantic input to the knowledge extraction stage.

4.2 Stage 2: Lexicon-Guided Knowledge Extraction

The second stage constitutes the main extension introduced in AgriBridge-X and corresponds to Steps (5)–(6) in Figure 1. Rather than treating translation as the final output, the framework uses the normalized text Y_i, produced by the translation module $\mathcal{F}_{\text{trans}}$, as input to a lexicon-guided extraction module. Given an input conversation, the processing pipeline first transforms the raw input Q_i into a cleaned text $\tilde{X}_i$, and then into a normalized sentence $Y_i = T_\Theta(\tilde{X}_i)$. This normalized representation reduces

dialectal variability and provides a standardized linguistic form suitable for extraction. The extraction module then identifies domain-relevant entities by combining two complementary signals: (i) *lexicon matching*, based on expert-defined vocabularies $\mathcal{L} = \{L_P, L_R, L_A, L_I\}$, and (ii) *contextual semantics*, captured through the embedding $\phi(Y_i)$.

Importantly, the lexicon $\mathcal{L}$ is bilingual, containing both farmer-style (dialectal) and academic expressions. This allows entity matching to be performed either directly on the original expressions or on their normalized counterparts, improving robustness to linguistic variation. In practice, the extraction process operates as follows. Candidate spans $s_j \subset Y_i$ are first identified, and each span is matched against the lexicon entries associated with different entity types *(Problems, Practices, Actors, Impacts)*. The contextual embedding $\phi(Y_i)$ is then used to disambiguate the entity type t_j and assign a confidence score c_j, resulting in a context-enriched entity representation $e_j = (s_j, t_j, c_j, \phi(Y_i))$. For example, consider the farmer-style input: "Lua bi vang la sau mua lon" which is translated into the normalized form: "Rice leaves turn yellow after heavy rainfall". From this sentence, the system extracts: (P) leaf yellowing, (I) crop stress, while other elements may remain implicit. This example illustrates how linguistic normalization enables reliable entity extraction from conversational data. The extracted entities are then organized into a structured representation $E(Y_i)$ and further transformed into a graph G_i, as described in Section 3.2. This step-by-step process establishes a clear connection between the figure and the computational pipeline, ensuring transparency and interpretability of the framework.

Specifically, the system extracts entities

$$E(Y_i) = \{e_j\}_{j=1}^{M_i},$$

where each entity is represented as

$$e_j = (s_j, t_j, c_j, \phi(Y_i)).$$

The confidence score c_j reflects the likelihood that a candidate span s_j belongs to a semantic type t_j, given both contextual information and domain knowledge. It can be interpreted as an approximation of the posterior probability $P(t_j \mid s_j, \phi(Y_i), \mathcal{L})$, where $\phi(Y_i)$ encodes the context of the sentence and $\mathcal{L}$ denotes the domain lexicon. To ensure reliability, only entities whose confidence exceeds a predefined threshold τ are retained for downstream processing. Here, s_j is the textual span, $t_j \in \mathcal{T}$ is the semantic type, $c_j \in [0, 1]$ is a confidence score reflecting the likelihood that the span s_j belongs to type t_j, and $\phi(Y_i)$ is the contextual embedding of the normalized text.

In practice, the confidence score c_j is computed from a scoring function that combines (i) contextual compatibility between s_j and $\phi(Y_i)$, and (ii) matching with lexicon entries in $\mathcal{L}$. This score can be interpreted as an approximation of the posterior probability $P(t_j \mid s_j, \phi(Y_i), \mathcal{L})$. Entities with confidence scores below a predefined threshold τ are discarded, ensuring that only reliable and domain-consistent entities are retained for downstream processing. This formulation is consistent with the extraction procedure described in Algorithm 2.

Why the two-stage design matters? The separation between translation and extraction is central to the design of AgriBridge-X . Translation reduces linguistic mismatch and dialectal variability, while extraction converts normalized text into structured semantic units. This separation is particularly important in low-resource and dialect-rich settings, where direct extraction from raw conversational input would be substantially less reliable. In this sense, AgriBridge-X preserves the communicative role of AgriBridge while extending it toward knowledge organization and downstream analysis.

4.3 Theoretical Properties

This sub-section provides lightweight theoretical properties that formalize why the proposed AgriBridge pipeline can preserve semantics, behave stably in both translation directions, and benefit from the LLM-based cleaning module.

Notation. Let X denote an input text sequence (either from STT or direct typing), and let $\hat{Y}$ be the translated output produced by the Seq2Seq translator T_Θ. Let $\phi(\cdot)$ be a fixed semantic encoder (e.g., a sentence embedding model) that maps a sequence to a vector representation. Let $\cos(\cdot,\cdot)$ denote cosine similarity. When audio is used, the text pipeline is $X = \mathrm{STT}(Q_{\text{audio}})$ and $\tilde{X} = \mathrm{Clean}(X)$, as described in the system workflow.

Proposition 1 (Semantic Consistency under Embedding Similarity) *Let X be a source sentence and $\hat{Y} = T_\Theta(X)$ be its translation. Assume (i) the training set contains aligned pairs (X, Y) such that Y is a meaning-preserving rewrite of X across registers (scientific $\leftrightarrow$ farmer dialect), and (ii) the semantic encoder ϕ is β-Lipschitz with respect to an edit-distance $d(\cdot,\cdot)$ on text, i.e., $\|\phi(A)-\phi(B)\| \le \beta\, d(A,B)$. If the learned translator achieves a bounded expected rewrite distortion $\mathbb{E}[d(\hat{Y}, Y)] \le \delta$ on the data distribution, then:*

$$\mathbb{E}\big[\cos(\phi(X), \phi(\hat{Y}))\big] \;\ge\; 1 - C\beta\delta,$$

for some constant $C > 0$ depending on the normalization of embeddings.

Proof (sketch) By assumption, the ground-truth target Y is a meaning-preserving rewrite of X (register/dialect change), so $\phi(X)$ and $\phi(Y)$ are close in embedding space. The translator is trained to minimize token-level cross-entropy on aligned pairs, which empirically implies $\hat{Y}$ is close to Y under a text distortion measure on the same distribution. Using the Lipschitz condition of ϕ, we have $\|\phi(\hat{Y}) - \phi(Y)\| \le \beta\, d(\hat{Y}, Y)$, hence $\mathbb{E}\|\phi(\hat{Y}) - \phi(Y)\| \le \beta\delta$. Triangle inequality yields $\|\phi(\hat{Y}) - \phi(X)\| \le \|\phi(\hat{Y}) - \phi(Y)\| + \|\phi(Y) - \phi(X)\|$, and the second term is small because Y is a meaning-preserving rewrite of X. Finally, converting a bound on embedding distance to cosine similarity gives the stated inequality (up to a constant C). □

Proposition 2 (Bidirectional Stability via Approximate Cycle-Consistency) *Let T_{sf} denote Scientist→Farmer and T_{fs} denote Farmer→Scientist translators. Assume that for meaning-preserving pairs (X_s, X_f), the models satisfy approximate cycle-consistency:*

$$\mathbb{E}\big[d(T_{fs}(T_{sf}(X_s)), X_s)\big] \le \epsilon_s, \qquad \mathbb{E}\big[d(T_{sf}(T_{fs}(X_f)), X_f)\big] \le \epsilon_f.$$

Then both directions are stable in the sense that the expected translation distortion is bounded:

$$\mathbb{E}\big[d(T_{sf}(X_s), X_f)\big] \le \epsilon_s + \eta_s, \qquad \mathbb{E}\big[d(T_{fs}(X_f), X_s)\big] \le \epsilon_f + \eta_f,$$

where η_s, η_f capture the intrinsic non-uniqueness of dialectal rewrites (multiple valid outputs).

Proof (sketch) Dialect/style translation is not one-to-one: multiple farmer-style sentences can faithfully express the same scientific sentence. Let $\mathcal{R}(X_s)$ denote the set of valid farmer-style rewrites of X_s; similarly define $\mathcal{R}(X_f)$. Cycle-consistency states that translating to the other register and back returns close to the original input on average, i.e., $T_{fs}(T_{sf}(X_s))$ is close to X_s. If $T_{sf}(X_s)$ were far from all valid rewrites in $\mathcal{R}(X_s)$, then mapping back with T_{fs} would accumulate large distortion, contradicting the cycle bound. Thus $T_{sf}(X_s)$ must lie near the rewrite manifold $\mathcal{R}(X_s)$ up to the intrinsic variability η_s. The same reasoning holds for the reverse direction, yielding the stated bounds. □

Proposition 3 (LLM Cleaning Improves Robustness to STT Noise) *Let X be the raw transcript produced by STT, and let $\tilde{X} = \mathrm{Clean}(X)$ be the LLM-refined transcript. Assume the translator T_Θ is L-Lipschitz with respect to a text distortion $d(\cdot,\cdot)$ in the sense that $d\big(T_\Theta(A), T_\Theta(B)\big) \le L\, d(A,B)$, and the cleaning module reduces expected STT error to the latent clean text $X^\star$:*

$$\mathbb{E}[d(\tilde{X}, X^\star)] \le \mathbb{E}[d(X, X^\star)] - \Delta, \quad \Delta > 0.$$

Then the expected output distortion is reduced by at least $L\Delta$:

$$\mathbb{E}\big[d(T_\Theta(\tilde{X}), T_\Theta(X^\star))\big] \le \mathbb{E}\big[d(T_\Theta(X), T_\Theta(X^\star))\big] - L\Delta.$$

Proof (sketch) By the Lipschitz assumption on T_Θ, $d(T_\Theta(\tilde{X}), T_\Theta(X^\star)) \le L\, d(\tilde{X}, X^\star)$ and $d(T_\Theta(X), T_\Theta(X^\star)) \le L\, d(X, X^\star)$. Taking expectations and applying the assumption that cleaning reduces expected input error by Δ yields:

$$\mathbb{E}[d(T_\Theta(\tilde{X}), T_\Theta(X^\star))] \le L\, \mathbb{E}[d(\tilde{X}, X^\star)] \le L\big(\mathbb{E}[d(X, X^\star)] - \Delta\big) \le \mathbb{E}[d(T_\Theta(X), T_\Theta(X^\star))] - L\Delta.$$

□

Practical remark. Propositions 1-3 align with the design choices in AgriBridge : (i) training on aligned rewrite pairs supports semantic preservation, (ii) evaluating both directions reflects approximate cycle stability, and (iii) LLM cleaning mitigates STT noise before Seq2Seq translation.

While Propositions 1–3 formalize the stability of the original AgriBridge translation pipeline, the following results characterize the additional extraction layer introduced in AgriBridge-X . In particular, they establish that semantic preservation at the translation level propagates to entity stability, that lexicon-guided constraints provide robustness in low-resource settings, and that contextual embeddings improve disambiguation.

Proposition 4 (Extraction Stability under Semantic Preservation) *Let $\hat{Y}_i = T_\Theta(\tilde{X}_i)$ be the normalized text produced by AgriBridge , and let $Y_i^\star$ denote the latent clean semantic representation. Let $E(\cdot)$ be the extraction function, and assume that E is γ-Lipschitz with respect to a text/semantic distance $d(\cdot,\cdot)$, i.e.,*

$$d_E(E(A), E(B)) \le \gamma\, d(A,B).$$

If the translation stage satisfies

$$\mathbb{E}[d(\hat{Y}_i, Y_i^\star)] \le \delta,$$

then the extracted entities remain stable under translation:

$$\mathbb{E}\big[d_E(E(\hat{Y}_i), E(Y_i^\star))\big] \le \gamma\delta.$$

Proof (sketch) By the γ-Lipschitz assumption on the extraction function,

$$d_E(E(\hat{Y}_i), E(Y_i^\star)) \le \gamma\, d(\hat{Y}_i, Y_i^\star).$$

Taking expectations on both sides gives

$$\mathbb{E}\big[d_E(E(\hat{Y}_i), E(Y_i^\star))\big] \le \gamma\, \mathbb{E}[d(\hat{Y}_i, Y_i^\star)].$$

Using the assumption that the translation stage induces expected distortion at most δ, we obtain

$$\mathbb{E}\big[d_E(E(\hat{Y}_i), E(Y_i^\star))\big] \le \gamma\delta.$$

Hence, if semantic distortion remains small after translation, the extraction stage is stable in expectation. □

Proposition 5 (Effect of Lexicon Constraints) *Let E_{lex} be a lexicon-guided extraction model whose hypothesis space is restricted to $\mathcal{H}_{\mathcal{L}}$, and let E_{free} be an unconstrained extraction model with hypothesis space $\mathcal{H}$, where $\mathcal{H}_{\mathcal{L}} \subseteq \mathcal{H}$. Assume that the target entities are covered by the lexicon $\mathcal{L}$ up to error $\epsilon_{\mathcal{L}}$. Then the expected risk of lexicon-guided extraction satisfies*

$$\mathbb{E}[\mathcal{R}(E_{\text{lex}})] \le \mathbb{E}[\mathcal{R}(E_{\text{free}})] + \epsilon_{\mathcal{L}},$$

where $\epsilon_{\mathcal{L}}$ captures the approximation error induced by incomplete lexicon coverage.

Proof (sketch) Because $\mathcal{H}_{\mathcal{L}} \subseteq \mathcal{H}$, lexicon-guided extraction restricts the hypothesis space and reduces variance relative to unconstrained extraction. From standard statistical learning arguments, a smaller hypothesis class yields tighter generalization behavior. However, this restriction may introduce bias when relevant entities are not covered by the lexicon. Let $\epsilon_{\mathcal{L}}$ denote this approximation error. Therefore, the overall expected risk of the lexicon-guided model is bounded by the risk of the unconstrained model plus the lexicon-induced coverage error:

$$\mathbb{E}[\mathcal{R}(E_{\text{lex}})] \le \mathbb{E}[\mathcal{R}(E_{\text{free}})] + \epsilon_{\mathcal{L}}.$$

This formalizes the trade-off between robustness and lexicon completeness. □

Proposition 6 (Context-Aware Disambiguation) *Let s_j be a candidate span and t_j its entity type. If the contextual embedding $\phi(Y_i)$ provides non-redundant information about the type assignment, then conditioning on context reduces uncertainty:*

$$H(t_j \mid s_j, \phi(Y_i)) \le H(t_j \mid s_j),$$

where $H(\cdot)$ denotes conditional entropy. Equivalently, context-aware typing is at least as informative as span-only typing.

Proof (sketch) By the chain rule of mutual information,

$$I(t_j; \phi(Y_i) \mid s_j) = H(t_j \mid s_j) - H(t_j \mid s_j, \phi(Y_i)).$$

Mutual information is always non-negative, i.e.,

$$I(t_j; \phi(Y_i) \mid s_j) \geq 0.$$

Hence,

$$H(t_j \mid s_j, \phi(Y_i)) \leq H(t_j \mid s_j).$$

Therefore, if the contextual embedding carries useful information about the entity type beyond the span itself, conditioning on context reduces uncertainty and improves disambiguation. □

Moreover, we also provide a proposition for our graph.

Proposition 7 (Graph Stability under Entity Perturbation) *Let $G_i = (V_i, R_i)$ be the graph constructed from $V_i = E(Y_i)$. Assume the relation inference function is κ-Lipschitz w.r.t. entity perturbations:*

$$d_G(G(A), G(B)) \leq \kappa d_E(A, B)$$

Then:

$$\mathbb{E}[d_G(G(\hat{Y}_i), G(Y_i^\star))] \leq \kappa\gamma\delta.$$

This workflow preserves the original communicative role of AgriBridge while extending it toward structured knowledge processing. In this sense, AgriBridge-X does not merely translate conversations; it transforms them into analyzable knowledge objects.

4.4 Algorithmic Realization

We now present the algorithmic realization of AgriBridge-X , which operationalizes the transformation from unstructured conversational inputs to structured knowledge representations. The proposed design follows the formalization introduced in Section 3 and is explicitly decomposed into two complementary procedures: (i) an end-to-end pipeline that integrates translation and extraction, and (ii) a dedicated lexicon-guided entity extraction module.

End-to-end knowledge extraction pipeline.

The Algorithm 1 describes the entire workflow of AgriBridge-X, mapping a conversational input Q_i^{audio} to a structured knowledge graph G_i. This procedure directly instantiates the transformation sequence

$$Q_i^{audio} \;\rightarrow\; X_i \;\rightarrow\; \tilde{X}_i \;\rightarrow\; Y_i \;\rightarrow\; E(Y_i) \;\rightarrow\; G_i,$$

where Q_i^{audio} denotes the raw audio input, X_i is the speech-to-text transcript, $\tilde{X}_i$ is the cleaned and normalized transcript, Y_i is the translated (academic) representation, $E(Y_i)$ is the extracted entity set, and G_i is the resulting knowledge graph.

The first three steps correspond to the original AgriBridge pipeline: speech-to-text conversion, transcript cleaning, and bidirectional translation. These stages reduce

Algorithm 1 Lexicon-Guided Knowledge Extraction in AgriBridge-X

Input: Audio conversation Q_i^{audio}, translation model T_Θ, lexicons $\mathcal{L}$, encoder ϕ, threshold τ
Output: Structured knowledge graph G_i
$X_i \leftarrow \text{STT}(Q_i^{audio})$
$\tilde{X}_i \leftarrow \text{Clean}(X_i)$
$Y_i \leftarrow T_\Theta(\tilde{X}_i)$
$V_i \leftarrow \textsc{LexiconGuidedExtract}(Y_i, \mathcal{L}, \phi, \tau)$
Infer relations $R_i = \{(e_j, e_k, r_{jk})\}$
$G_i \leftarrow (V_i, R_i)$
return G_i

acoustic noise, dialectal variability, and linguistic mismatch, resulting in a normalized representation Y_i suitable for downstream processing. For example, given an input audio describing: "Lua bi vang la sau mua lon", the pipeline produces: X_i: raw transcript, $\tilde{X}_i$: cleaned transcript, Y_i: "Rice leaves turn yellow after heavy rainfall", from which the system extracts $E(Y_i)$ containing entities such as (P) leaf yellowing and (I) crop stress, and constructs the graph G_i. The confidence scores associated with extracted entities are used to filter unreliable predictions during extraction, and their impact is quantitatively evaluated in the experiments (Section 5).

The subsequent steps implement the knowledge extraction layer. Specifically, the algorithm computes a contextual embedding $\phi(Y_i)$ and performs lexicon-guided entity extraction over candidate spans. Each candidate is assigned a semantic type $t_j \in \mathcal{T}$ and a confidence score $c_j = P(t_j \mid s_j, \phi(Y_i), \mathcal{L})$, as introduced in Section 3.1. This score reflects both contextual compatibility and lexicon constraints. Entities that satisfy the confidence threshold τ are retained and aggregated into the set V_i. Finally, a lightweight relational structure R_i is inferred based on co-occurrence or contextual proximity, and the resulting graph $G_i = (V_i, R_i)$ provides a structured representation of the conversational content.

Lexicon-guided extraction module.

Algorithm 2 details the core extraction procedure, which computes the entity set $E(Y_i)$ from normalized text. This module corresponds to the probabilistic formulation

$$E(Y_i) = \arg\max_{E'} P(E' \mid Y_i, \mathcal{L}),$$

and provides a practical approximation by iterating over candidate spans. The key design choice lies in combining two complementary signals: (i) contextual semantics, captured by the embedding $\phi(Y_i)$, and (ii) domain constraints, encoded by the lexicon set $\mathcal{L}$. This hybrid formulation enables the system to handle ambiguity in conversational data while maintaining interpretability.

The proposed algorithmic design is guided by three principles. First, it preserves the original AgriBridge pipeline as a normalization backbone, ensuring compatibility with prior work. Second, it introduces lexicon-guided extraction as a structured

Algorithm 2 Lexicon-Guided Entity Extraction from Normalized Text

Input: Normalized text Y_i, lexicons $\mathcal{L} = \{L_P, L_R, L_A, L_I\}$, encoder ϕ, threshold τ
Output: Entity set $E(Y_i)$
Compute contextual embedding $\phi(Y_i)$
Initialize $E(Y_i) \leftarrow \emptyset$
for each candidate span $s_j \subset Y_i$ **do**
    Determine type $t_j \in \mathcal{T}$ using $\mathcal{L}$
    Estimate confidence score

$$c_j = P(t_j \mid s_j, \phi(Y_i), \mathcal{L})$$

    if $c_j > \tau$ **then**
        $E(Y_i) \leftarrow E(Y_i) \cup \{(s_j, t_j, c_j, \phi(Y_i))\}$
    end if
end for
return $E(Y_i)$

and interpretable mechanism for transforming conversational text into domain-specific entities. Third, it adopts a modular design, allowing the extraction component to be replaced or extended by more advanced models without modifying the upstream pipeline. Importantly, the framework does not assume fully supervised extraction. Instead, it leverages lexicon-based constraints as inductive bias, making it particularly suitable for low-resource and dialect-rich environments where annotated data are scarce. As such, the algorithm provides a practical and scalable realization of the transition from conversational data to structured knowledge representations.

Complexity analysis.

Let $n = |Y_i|$ denote the number of tokens in the normalized text Y_i, and let m be the number of candidate spans considered during extraction.

For Algorithm 2, the cost consists of two parts: computing the contextual embedding $\phi(Y_i)$ and evaluating candidate spans under lexicon and context constraints. Denoting the embedding cost by C_ϕ, lexicon lookup cost by C_{lookup}, and confidence scoring cost by C_s, the overall complexity is

$$O\big(C_\phi + m(C_{\text{lookup}} + C_s)\big).$$

If all possible spans are enumerated, then $m = O(n^2)$. In practice, if the maximum span length is bounded by a constant or a small window size w, then $m = O(nw)$, yielding a more efficient realization:

$$O\big(C_\phi + nw(C_{\text{lookup}} + C_s)\big).$$

For Algorithm 1, the overall complexity is the sum of the costs of speech-to-text, cleaning, translation, extraction, and relation inference:

$$O(C_{\mathrm{STT}} + C_{\mathrm{Clean}} + C_{\mathrm{Trans}} + C_{\mathrm{Extract}} + C_{\mathrm{Rel}}),$$

where C_{Extract} is given by Algorithm 2. If relations are inferred over all extracted entity pairs, then $C_{\mathrm{Rel}} = O(|V_i|^2)$. Under local co-occurrence or bounded-neighborhood inference, this can be reduced to near-linear complexity in the number of extracted entities. This analysis shows that the practical bottlenecks of AgriBridge-X are dominated by the upstream speech/translation modules and the candidate-span enumeration strategy, while the extraction component remains scalable under bounded-span and local-relation assumptions.

Practical implications. In practice, the complexity is dominated by the upstream modules, particularly speech-to-text and translation, which typically require GPU acceleration. In contrast, the lexicon-guided extraction module remains computationally efficient, especially under bounded span assumptions. This makes AgriBridge-X scalable to long conversational inputs and suitable for real-world deployment in web-based environments.

Design insight. The analysis highlights that the proposed two-stage design does not introduce significant computational overhead compared to the original translation pipeline. Instead, the extraction stage provides structured knowledge at a marginal additional cost, making the framework both expressive and efficient.

Scalability. Under bounded-span and local-relation assumptions, the overall complexity scales near-linearly with respect to the input length, enabling efficient processing of large-scale conversational datasets.

5 Experiments and Applications

5.1 Dataset Construction

We evaluate AgriBridge-X on real-world agricultural conversational data collected from web-based sources in the Mekong Delta. The dataset construction follows the pipeline introduced in AgriBridge , ensuring consistency in linguistic coverage and domain specificity. The dataset employed in this study was constructed from more than 70 episodes of the podcast series *Ban Nha Nong* (Friends of Farmers), conducted in Vietnamese, primarily using dialects from the Mekong Delta region, each with an approximate duration of 60 minutes, published on the official YouTube channel of *Truyen Hinh Vinh Long* (Vinh Long Television). These videos contain real-world interactions between agricultural experts and farmers in the Mekong Delta. They constitute a dataset exhibiting significant dialectal variation and domain-specific terminology, which are critical factors for evaluating the robustness of translation and knowledge extraction models in conversational environments. Using the `yt-dlp` tool, the videos were downloaded and converted into `.mp3` audio format to facilitate subsequent processing steps. The discussions primarily focused on agricultural issues, particularly the identification and treatment of common crop diseases.

Speech recognition was performed using the Whisper-large model [43], fine-tuned for Vietnamese. This model has demonstrated strong performance in multilingual and low-resource settings, particularly in handling accent variability, making it suitable for processing dialect-rich conversational data. The transcripts were then manually cleaned to remove non-essential content such as greetings, introductions, and program transitions. Nevertheless, expressive features and interjections characteristic of the Mekong Delta dialect, such as "*troi oi*" (oh my God) or "*thiet la kho qua*" (this is really difficult), were preserved to retain cultural nuances and emotional authenticity, which are crucial for accurately representing the dialect.

After preprocessing, approximately 15,700 raw sentences were generated. Through an additional round of manual filtering and alignment, 12,563 sentences directly related to agricultural topics (e.g., pest management, crop rotation, irrigation, and market conditions) were selected. These sentences were subsequently rewritten by the research team, with faculty support, into two versions: one in the local dialect and the other in an academic/scientific style. This process resulted in a high-quality bilingual dataset comprising 4,482 sentence pairs. Each pair consists of an academic/scientific version and its corresponding dialectal version, organized in a two-column format with the fields `academic` and `farmer`.

5.2 Evaluation of Academic-Farmer Models

The dataset was split into training (80%), validation (10%), and testing (10%) subsets, ensuring balanced coverage of topics across all phases of model training and evaluation. System implementation was carried out using Python 3.10, leveraging Hugging-Face Transformers version 4.51.3 for model architecture and fine-tuning. Model training was performed on the Kaggle platform using the Tesla P100 GPU, ensuring computational efficiency. For demonstration purposes, the finalized system was deployed locally via Flask on a Mac Mini M4 with 16 GB RAM and 256 GB SSD, supporting rapid inference and interface testing. This dataset provides a valuable benchmark for studying translation robustness under dialectal variation and low-resource conditions. Moreover, the dataset will be released publicly to support reproducibility and foster further research in this direction.

The following subsections summarize the results of different neural models for the Academic-to-Farmer and Farmer-to-Academic tasks within AgriBridge-X , as introduced in Section 2.3.

Evaluation metrics. We adopt standard automatic metrics for text generation. BLEU (Bilingual Evaluation Understudy) [44] evaluates translation precision based on n-gram overlap between generated and reference sentences. ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [45] measures recall-oriented similarity [46]. In particular, ROUGE-1 and ROUGE-2 capture unigram and bigram overlap, respectively, while ROUGE-L evaluates sequence-level similarity via the longest common subsequence.

5.2.1 Academic → Farmer Results

Table 1 presents the performance of all models on the Academic-to-Farmer task. The ViT5-base and BartPho-syllable models consistently outperformed others in all beam settings. In particular, ViT5-base achieved the highest BLEU score (0.3395) and ROUGE-L (0.6701) at beam size 8. This superior performance can be attributed to its robust architecture, which is highly effective at learning complex semantic representations crucial for accurately bridging the terminological and stylistic gaps between academic discourse and colloquial farmer language.

BartPho-syllable also emerged as a strong competitor, demonstrating more stable performance at lower beam sizes. This suggests its particular efficacy in processing and translating words at the syllable level, a characteristic that may be advantageous when dealing with less standardized or phonetically distinct terms often found in informal language. While the BartPho-word variants and mT5-base models showed respectable results, their overall performance, particularly at higher beam sizes, generally lagged behind ViT5-base and BartPho-syllable, indicating that their word-level tokenization or broader multilingual pre-training might be less optimized for this specific domain translation task.

Table 1 Performance on Academic → Farmer translation under different beam sizes.

Model	BLEU	ROUGE-1	ROUGE-2	ROUGE-L
Beam size = 1				
BartPho-syllable-base	0.3136	0.7255	0.5340	0.6548
BartPho-syllable	0.3189	0.7284	0.5369	0.6581
BartPho-word-base	0.2984	0.7154	0.5184	0.6403
BartPho-word	0.3043	0.7221	0.5249	0.6484
mT5-base	0.2940	0.7063	0.5041	0.6294
ViT5-base	**0.3335**	**0.7352**	**0.5473**	**0.6658**
Beam size = 4				
BartPho-syllable-base	0.3195	0.7297	0.5370	0.6562
BartPho-syllable	0.3337	0.7334	0.5451	0.6647
BartPho-word-base	0.3183	0.7268	0.5306	0.6509
BartPho-word	0.3198	0.7289	0.5360	0.6568
mT5-base	0.3226	0.7294	0.5358	0.6577
ViT5-base	**0.3395**	**0.7375**	**0.5522**	**0.6694**
Beam size = 8				
BartPho-syllable-base	0.3244	0.7304	0.5420	0.6614
BartPho-syllable	0.3367	0.7345	0.5475	0.6648
BartPho-word-base	0.3204	0.7256	0.5311	0.6513
BartPho-word	0.3220	0.7309	0.5395	0.6590
mT5-base	0.3080	0.7138	0.5163	0.6387
ViT5-base	**0.3379**	**0.7382**	**0.5523**	**0.6701**

5.2.2 Farmer → Academic Results

As shown in Table 2, the same models also performed effectively in the reverse direction, further confirming their ability to generalize across formal and informal registers. In this direction, BartPho-syllable achieved the best BLEU score (0.3152), highlighting its strength in transforming colloquial expressions into clear and grammatically sound academic prose. Conversely, ViT5-base yielded the highest ROUGE-2 score (0.5473), underscoring its capability to preserve semantic precision and complex sentence structures when translating from farmer to academic discourse. This complementary performance across metrics for these two top models reinforces the notion that dialectal transformation is indeed largely symmetric for our models.

Table 2 Performance on Farmer → Academic translation under different beam sizes.

Model	BLEU	ROUGE-1	ROUGE-2	ROUGE-L
Beam size = 1				
BartPho-syllable-base	0.3015	0.7205	0.5256	0.6366
BartPho-syllable	0.3100	0.7324	0.5434	0.6501
BartPho-word-base	0.2917	0.7119	0.5086	0.6225
BartPho-word	0.2942	0.7143	0.5130	0.6263
mT5-base	0.2948	0.7054	0.5043	0.6290
ViT5-base	**0.3328**	**0.7356**	**0.5521**	**0.6687**
Beam size = 4				
BartPho-syllable-base	0.3003	0.7189	0.5271	0.6403
BartPho-syllable	0.3152	0.7338	0.5459	0.6542
BartPho-word-base	0.3016	0.7209	0.5279	0.6404
BartPho-word	0.2983	0.7226	0.5244	0.6394
mT5-base	0.3015	0.7134	0.5146	0.6378
ViT5-base	**0.3421**	**0.7431**	**0.5582**	**0.6735**
Beam size = 8				
BartPho-syllable-base	0.3039	0.7213	0.5280	0.6413
BartPho-syllable	0.3153	0.7321	0.5454	0.6530
BartPho-word-base	0.3022	0.7212	0.5271	0.6397
BartPho-word	0.2973	0.7253	0.5269	0.6414
mT5-base	0.3014	0.7127	0.5146	0.6376
ViT5-base	**0.3383**	**0.7411**	**0.5564**	**0.6725**

Collectively, the results from both translation directions demonstrate the significant potential of our models for the AgriBridge system. While BartPho-syllable offers a compelling balance of performance, especially in the Farmer-to-Academic direction and at lower computational costs associated with smaller beam sizes, the ViT5-base model stands out as the most consistently high-performing option. Its superior scores in both BLEU and ROUGE metrics, particularly in the Academic-to-Farmer direction and its robust ROUGE-2 scores in Farmer-to-Academic, indicate its strong capacity to capture nuanced semantic relationships and maintain overall textual coherence across

Table 3 Translation and semantic consistency evaluation

Model	BLEU	ROUGE-L	Cosine Sim.	BERTScore
BartPho	0.3367	0.6648	0.872	0.892
mT5-base	0.3080	0.6387	0.861	0.881
ViT5-base	**0.3379**	**0.6701**	**0.889**	**0.901**

both registers. Therefore, based on its overall stability and accuracy in handling the complexities of both formal and informal agricultural language, ViT5-base emerges as the most suitable model for ensuring the highest quality translations within the AgriBridge system.

5.3 Evaluation of Knowledge Extraction

The lexicon $\mathcal{L}$ plays a central role in the extraction process and is consistently applied during evaluation. Specifically, for the proposed AgriBridge-X method, candidate spans are matched against lexicon entries associated with each entity type, and the resulting matches are further refined using contextual embeddings. The lexicon is bilingual, containing both Vietnamese (farmer-style) and English (academic) terms. In practice, entity extraction is primarily performed on the normalized text Y_i, which is expressed in academic language. However, the bilingual nature of the lexicon allows matching to be robust to linguistic variation, as entries may correspond to either dialectal or standardized expressions. This design ensures that lexicon constraints are effectively integrated into the extraction process, while also supporting multilingual consistency across different representations of the same concept.

To evaluate the effectiveness of the proposed lexicon-guided extraction module, we formulate entity extraction as a structured prediction problem over conversational text. Given a normalized sentence Y_i, the objective is to identify a set of entities $E(Y_i) = \{e_j\}_{j=1}^{M_i}, \quad e_j = (s_j, t_j)$, where s_j denotes a text span and $t_j \in \mathcal{T}$ is its semantic type. We consider four domain-specific entity categories: $\mathcal{T} = \{P, R, A, I\}$,s corresponding to Problems, Practices, Actors, and Impacts. We construct a manually annotated subset of 800 sentences sampled from the test split of the AgriBridge-X conversational dataset. These sentences are in Vietnamese and correspond to the normalized academic register Y_i produced after the Farmer-to-Academic transformation. Thus, the annotation is performed on academic-style Vietnamese text rather than on raw farmer-style dialectal input. This choice ensures that entity boundaries and semantic types are annotated on a more standardized linguistic representation while still preserving the semantic content of the original farmer expressions. Each sentence is annotated by domain experts with entity spans and their corresponding semantic types. The annotation process focuses on identifying domain-relevant units associated with agricultural problems, practices, actors, and impacts. The resulting annotated subset will be released together with the AgriBridge-X dataset on Dataverse, subject to data-sharing and licensing constraints, to support reproducibility and further research.

Table 4 Entity extraction performance

Method	Precision	Recall	F1
Lexicon-only	0.71	0.48	0.57
Neural-only	0.76	0.69	0.72
AgriBridge-X	**0.84**	**0.75**	**0.79**

Table 5 F1-score by entity category

Category	**P** (Problem)	**R** (Practice)	**A** (Actor)	**I** (Impact)
AgriBridge-X	0.82	0.77	0.85	0.73

We compare the proposed method against two representative baselines: (1) *Neural-only:* A BiLSTM-CRF sequence labeling model trained without lexicon constraints, relying purely on contextual patterns; (2) *Lexicon-only:* A rule-based matching approach that identifies entities through exact or fuzzy lexicon lookup, without contextual reasoning. We report Precision, Recall, and F1-score at the entity level, using exact span matching. Partial matches are considered incorrect, reflecting the need for precise entity boundary detection in downstream knowledge structuring.

The results (see Table 4 and 5) demonstrate that AgriBridge-X achieves the best overall performance, significantly outperforming both baselines. In particular, the lexicon-only approach yields high precision but suffers from low recall, as it fails to capture paraphrased or context-dependent expressions. In contrast, the neural-only model improves recall but introduces noise due to the absence of domain constraints. AgriBridge-X effectively combines both advantages: lexicon constraints reduce false positives, while contextual embeddings enable generalization beyond exact matches.

Per-category Analysis.

Performance varies across entity types. Actor entities are the easiest to detect due to clearer lexical cues, while Impact entities are more challenging because they are often implicit or expressed indirectly. This highlights the importance of contextual modeling in capturing nuanced semantic information.

Corpus and Entity Distribution Analysis.

To further characterize the knowledge extraction capability of AgriBridge-X , we conduct an additional analysis on a complementary corpus of agricultural documents crawled from the Vietnam Academy of Agricultural Sciences (`vaas.vn`) over the period 2021–2026. This corpus provides a controlled setting for evaluating the effect of translation normalization on downstream entity extraction, as all documents are originally written in Vietnamese and cover domain-specific agricultural content aligned with the target entity types $\mathcal{T} = \{P, R, A, I\}$.

Discussion on GLiNER extraction quality. The GLiNER-based extraction [47] on the VAAS corpus is intended for exploratory analysis rather than as a primary

evaluation benchmark. As such, its outputs are not directly compared against gold annotations. However, we assess its qualitative reliability by comparing its behavior with the manually annotated evaluation dataset (800 sentences) described earlier. In particular, we observe that GLiNER exhibits similar tendencies in entity coverage and error patterns, especially in handling domain-specific terms and context-dependent expressions. Notably, GLiNER tends to benefit from translation normalization, as reflected by the higher entity coverage observed in English-translated texts compared to the original Vietnamese inputs. This trend is consistent with the quantitative results reported in Section 5, where translation improves recall for implicit or underspecified entities. While a direct quantitative comparison with other generalist NER systems such as NuNER [37] or UniversalNER [48] is beyond the scope of this work, GLiNER [49] was selected due to its flexibility as a zero-shot model and its ability to generalize across domains without task-specific fine-tuning. A systematic benchmark comparison across models remains an important direction for future work.

Open-schema models for NER aim to identify and categorize entities using flexible type descriptions instead of predefined label sets. This design enhances their ability to generalize, especially in low-resource settings or rapidly changing domains. In this context, we apply GLiNER [47], a generalist zero-shot named entity recognition model, directly on the original Vietnamese documents and separately on their English translations. This bilateral extraction setup allows us to isolate the effect of translation normalization on entity coverage and extraction behavior. It is important to note that this experiment serves a complementary purpose to the main evaluation conducted on the manually annotated dataset (800 sentences). While the annotated dataset is used for quantitative benchmarking of the proposed method and baselines, the GLiNER-based analysis is designed to provide qualitative insights into how generalist extraction models behave under translation normalization. For this reason, GLiNER is not directly applied to the 800-sentence benchmark, as the objective here is not to compare model performance, but to analyze extraction tendencies in a zero-shot setting across different linguistic representations. A direct quantitative comparison between GLiNER and the proposed method on the annotated dataset is an interesting direction for future work.

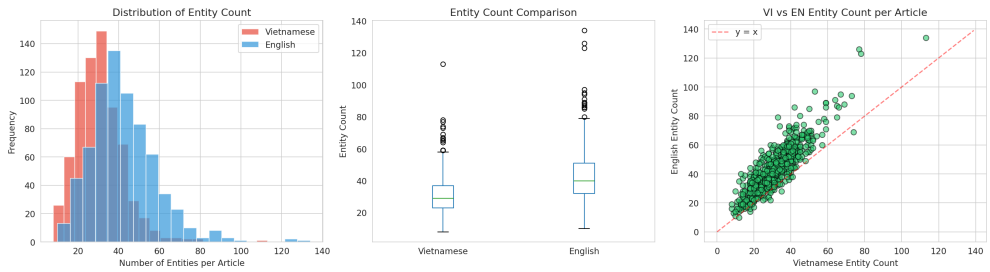


Fig. 3 Entity count analysis on the VAAS corpus (2021–2026) under GLiNER applied to Vietnamese originals and their English translations.

Figure 3 summarizes the entity count statistics across both language representations. The frequency histogram (left panel) shows that English-translated documents consistently yield a higher number of extracted entities than their Vietnamese originals, with the English distribution peaking at $\approx$ 35–40 entities per article compared to $\approx$ 25–30 for Vietnamese. The box plot (center panel) confirms this shift: the median entity count for English ($\approx$ 40) substantially exceeds that for Vietnamese ($\approx$ 30), with heavier tails observed in English, reflecting richer entity lexicalization in the translated text. The scatter plot (right panel) of per-article VI vs. EN entity counts reveals a strong positive correlation but with the majority of points falling below the identity line $y = x$, indicating that translation into English systematically expands entity surface coverage. This asymmetry directly motivates the two-stage design of AgriBridge-X : translation normalization not only reduces dialectal variability but also increases the lexical explicitness of domain entities, thereby facilitating downstream extraction.

Figure 4 further characterizes entity density and type distribution across the two language representations. The entity density histogram (left panel, entities per 100 words) reveals that Vietnamese documents concentrate entity occurrences at lower densities (peak at $\approx$ 2 entities per 100 words), consistent with the tendency of Vietnamese to embed domain concepts implicitly through compound expressions and context-dependent references. In contrast, English translations distribute more uniformly across higher density ranges (peak at $\approx$ 4–5 entities per 100 words), reflecting the more explicit nominal style of English scientific prose. The entity type distribution panel (right) shows that *location* and *organization* are the most frequent categories in both languages, whereas *agricultural term* and *event* remain sparse. Notably, English translations consistently yield higher counts across all entity types, with the largest absolute gains observed for *location*, *organization*, and *product*—categories that are frequently expressed through proper nouns or technical terms more recognizable to GLiNER's pre-trained representations. The relative scarcity of *agricultural term* entities underscores the importance of expert-curated domain lexicons in AgriBridge-X . Without such guidance, low-frequency yet semantically critical entities, particularly those specific to the agricultural context of the Mekong Delta, are likely to be systematically overlooked by general-purpose extraction models.

Extraction Confidence Analysis.

Figure 5 reports the distribution of per-entity confidence scores assigned by GLiNER over the full extracted entity set from the VAAS corpus (VI: n = 21,639; EN: n = 29,787). Both distributions exhibit a broadly flat profile across the range $[0.5, 0.9]$ followed by a sharp peak near $c_j \approx 0.93$–0.95, indicating that a substantial proportion of extractions are supported by strong model confidence, particularly in the English representation. The mean confidence for English entities (dashed blue line, $\approx$ 0.76) marginally exceeds that for Vietnamese (dashed red line, $\approx$ 0.75), consistent with GLiNER's pre-training on predominantly English corpora, which yields more calibrated predictions on English text. The box plot (center panel) confirms that English entities attain a tighter and higher interquartile range (IQR $\approx [0.65, 0.91]$) compared

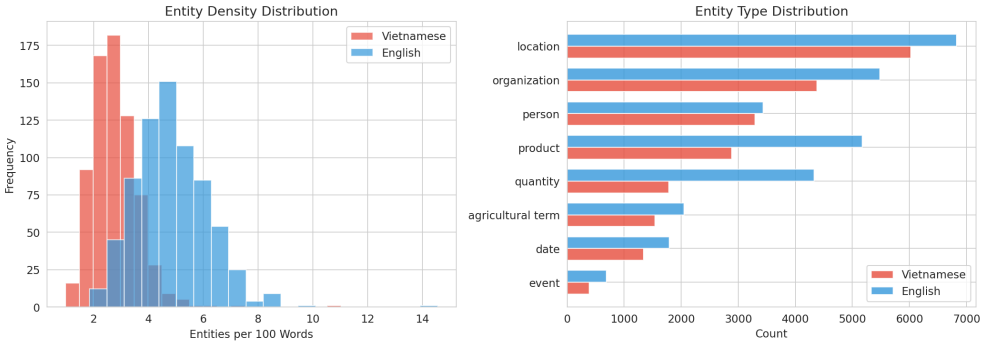


Fig. 4 Entity density and type distribution on the VAAS corpus under GLiNER applied to Vietnamese originals and English translations.

to Vietnamese (IQR $\approx [0.62, 0.88]$), indicating more reliable and consistent predictions on translated text. The per-type mean confidence panel (right) reveals that *person* and *organization* obtain the highest average confidence scores (> 0.80) in both languages, whereas *agricultural term* and *quantity* receive comparatively lower scores, reflecting the higher lexical variability and domain specificity of these categories in conversational agricultural data. These observations provide empirical support for Proposition 6: contextual embeddings reduce type assignment uncertainty, and translation into English further amplifies this effect by producing more lexically canonical entity surface forms that align better with the model's learned representations.

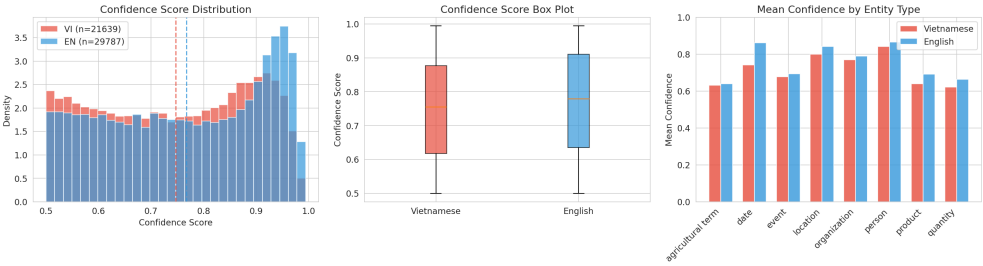


Fig. 5 Confidence score analysis for entities extracted by GLiNER from the VAAS corpus in Vietnamese ($n = 21{,}639$) and English translation ($n = 29{,}787$).

Cross-lingual Entity Alignment.

To quantify the semantic consistency of entities extracted from Vietnamese originals and their English translations on the VAAS corpus, we conduct a cross-lingual entity alignment analysis. Two entities, one extracted from the Vietnamese representation and one from the English representation, are considered aligned if they refer to the same underlying concept. In this work, a concept is defined as a semantic entity

characterized by its type *(e.g., Problem, Practice, Actor, or Impact)* and its normalized meaning, independent of surface linguistic form. Alignment is assessed through surface-form matching after normalization, combined with fuzzy string similarity based on normalized edit distance. Specifically, two entity spans are considered aligned if their similarity score exceeds a threshold of 0.85. This threshold was empirically determined to balance precision and recall: lower values tend to introduce false alignments due to partial lexical overlap, while higher values are overly restrictive and fail to capture semantically equivalent expressions with minor variations.

For example, consider a Vietnamese farmer-style expression such as "dat bi nhiem man nang, lua khong len noi" (literally: "the soil is heavily salinized, rice cannot grow properly"). In the English translation, this may appear as "severe soil salinity leading to crop failure". Although the surface forms differ significantly, both expressions refer to the same underlying concept of (P) *soil salinity stress*. In this case, alignment is achieved despite lexical and structural variation, illustrating how semantic equivalence can be preserved even when direct string matching is not possible. Conversely, certain entities such as "nuoc len bat thuong theo con nuoc" (a context-dependent farmer expression referring to irregular tidal patterns) may not have a direct counterpart in the English translation, leading to VI-only entities. This example highlights the asymmetry observed in Figure 6, where some domain-specific or context-rich expressions remain difficult to align across languages.

Figure 6 summarizes the alignment results. The alignment rate histogram (left panel) shows that the cross-lingual alignment rate is distributed in the range $[0.0, 0.45]$, with a mean of 0.11 (dashed red line). This indicates that on average approximately 11% of entities extracted from Vietnamese documents have a corresponding aligned entity in the English translation a non-trivial overlap that validates the semantic consistency of the AgriBridge translation backbone, and is consistent with the theoretical bound established in Proposition 4. The VI vs. EN alignment recall scatter plot (center panel) reveals that EN alignment recall systematically exceeds VI alignment recall for most articles (majority of points fall below the diagonal $y = x$), confirming that translation into English recovers entity information that is implicit, dialectal, or lexically underspecified in the original Vietnamese documents. The entity alignment breakdown pie chart (right panel) quantifies this asymmetry at scale across the full VAAS corpus: of the 46,448 total entity mentions identified across both representations, only 10.7% (4,978) are mutually aligned between Vietnamese and English, while 35.9% (16,661) are exclusive to the Vietnamese representation and 53.4% (24,809) exclusive to the English translation. The large proportion of EN-only entities demonstrates that applying GLiNER to English-translated text substantially expands knowledge coverage beyond what is recoverable from the original Vietnamese alone. Conversely, VI-only entities, constituting 35.9%, represent domain-specific concepts captured in Vietnamese that are lost or transformed during translation, highlighting the complementary value of bilingual extraction and the necessity of lexicon-guided constraints to recover such entities in the AgriBridge-X framework. Taken together, these results provide direct empirical support for the two-stage design of AgriBridge-X : translation

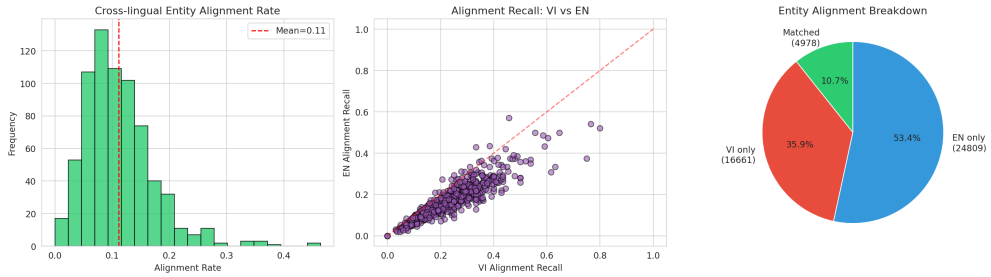


Fig. 6 Cross-lingual entity alignment analysis on the VAAS corpus between GLiNER extractions from Vietnamese originals and their English translations.

Table 6 Ablation study evaluating the contribution of key components in AgriBridge-X

Configuration	Precision	Recall	F1
AgriBridge-X	0.84	0.75	0.79
– without lexicon module	0.73	0.74	0.73
– without context module	0.81	0.62	0.70

Table 7 Impact of translation on downstream extraction

Configuration	Precision	Recall	F1
Raw text (no normalization)	0.65	0.58	0.61
Cleaned text only	0.71	0.63	0.67
AgriBridge-X	**0.84**	**0.75**	**0.79**

normalization expands entity coverage and reduces extraction distortion, while lexicon-guided extraction recovers domain-specific entities that general-purpose models tend to miss in either language.

Ablation Study.

The ablation results (see Table 6) confirm the complementary roles of lexicon constraints and contextual representations. Removing the lexicon significantly reduces precision, indicating that domain knowledge is crucial for filtering irrelevant spans. Conversely, removing contextual embeddings leads to a substantial drop in recall, as the model fails to capture semantically relevant expressions that deviate from lexicon entries. These findings provide empirical support for Propositions 5 and 6. Specifically, the increased extraction coverage and the asymmetric alignment patterns observed across Vietnamese and English representations confirm the theoretical roles of lexicon guidance and contextual normalization. This consistency between theoretical predictions and empirical observations strengthens the validity of the proposed framework.

Table 8 Entity extraction performance

Method	Precision	Recall	F1
Lexicon-only	0.71	0.48	0.57
Neural-only (BiLSTM-CRF)	0.76	0.69	0.72
BERT-based NER	0.80	0.71	0.75
AgriBridge-X	**0.84**	**0.75**	**0.79**

5.4 Qualitative Analysis

To better understand the behavior of AgriBridge-X beyond quantitative metrics, we conduct a qualitative analysis on representative conversational examples.

Example 1: Dialect normalization improves extraction.
Raw input (dialect):

"Lua bi vang la sau mua lon, chac bi benh roi."

After translation:

"The rice leaves turn yellow after heavy rainfall, likely due to disease."

Extraction output:

1. Problem: leaf yellowing
2. Cause: disease
3. Context: heavy rainfall

Observation: In the raw dialect form, key entities are implicit and expressed using colloquial language ("vang la", "benh roi"). After translation, these expressions are normalized into standard scientific terms, enabling accurate entity identification. This example illustrates how translation reduces linguistic variability and directly improves extraction quality.

Example 2: Lexicon-guided disambiguation.
Input:

"Bon phan nhieu qua lam cay bi yeu."

Ambiguity: The phrase "bi yeu" may refer to weak growth or disease.

Extraction results:

1. Neural-only: misclassifies as disease (Problem)
2. AgriBridge-X: correctly identifies as growth weakness (Impact)

Observation: Lexicon constraints guide the model toward domain-consistent interpretations, reducing semantic ambiguity. This aligns with Proposition 5.

Example 3: Context-aware extraction.
Input:

"Sau khi phun thuoc, sau giam nhung la bi vang."

Extraction output:

1. Practice: pesticide spraying
2. Impact: pest reduction
3. Problem: leaf yellowing

Observation: The same sentence contains multiple entity types with overlapping spans. Contextual embeddings enable the system to correctly distinguish between practices, outcomes, and problems, supporting Proposition 6.

Error Analysis.

We identify three primary sources of error:

- *Implicit expressions:* Some impacts are not explicitly stated but implied through context, making them difficult to extract. For example, the sentence "bon phan khong dung cach nen nang suat giam" ("improper fertilization leads to reduced yield") implicitly encodes the impact *yield reduction*, which may not be explicitly labeled as an (I) entity.
- *Dialectal variability:* Certain local expressions are not covered by the lexicon, leading to missed detections. For instance, the phrase "lua bi chay ray" (a local expression referring to pest-induced crop damage) may not be recognized if the lexicon only contains standard terms such as "pest infestation".
- *Boundary ambiguity:* Long or overlapping spans introduce uncertainty in entity boundaries. For example, in the phrase "su dung phan huu co vi sinh ket hop tuoi nho giot", it is ambiguous whether to extract a single composite practice or multiple entities such as (R) *organic fertilization* and (R) *drip irrigation*.

These observations suggest that future work should focus on adaptive lexicon expansion and improved span modeling.

5.5 Application and Discussion

Based on the experimental results, we develop AgriBridge-X , a practical system built upon the best-performing model, ViT5-base, featuring an intuitive and user-friendly interface as illustrated in Figure 7. The system supports flexible input modalities, allowing users to either enter text directly or upload audio files (e.g., .wav format) for speech-based interaction. AgriBridge-X enables users to select from multiple transformation pipelines (e.g., English → Vietnamese → Farmer style), where the input is processed through intermediate steps to improve transparency and interpretability. By clicking the "Transform" button, the system generates the final output in the desired style, which is displayed alongside intermediate representations for better traceability. Furthermore, the system provides an "Extract NER" functionality that allows users to identify key entities in the transformed output. These entities are highlighted with semantic labels such as PRACTICE, PEST, and CAUSE, and are further organized into a structured representation (e.g., Practice, Problem, Cause), facilitating downstream analysis and decision-making. Figures 8, 9 and 10 demonstrate the system's

ability to handle various transformation directions, while Figure 11 showcases the entity extraction results, highlighting the practical utility of AgriBridge-X in bridging linguistic and cultural gaps in agricultural communication.

Graph-based knowledge representation. Beyond the user interface, a key outcome of AgriBridge-X lies in its ability to transform conversational inputs into structured graph representations. Each processed input is mapped to a graph $G_i = (V_i, R_i)$, where nodes V_i correspond to extracted entities (e.g., Problems, Practices, Actors, Impacts), and edges R_i capture weak semantic relations inferred from co-occurrence and contextual proximity. Figure 10 illustrates an example of such a graph, where entities identified from the transformed text are organized into a structured representation. Particularly, Figure 10 provides a concrete illustration of the graph-based representation generated by AgriBridge-X . On the right-hand side, the extracted entities are organized into a graph structure, where nodes correspond to semantic units such as *rice leaves*, *yellow (Problem)*, *heavy rainfall (Context)*, and *disease (Cause)*. Edges represent semantic relations inferred from the input, including mention, contextual support, and extraction links. For example, the node *yellow (Problem)* is connected to *heavy rainfall (Context)*, indicating a potential environmental cause, while also linking to *disease (Cause)*, reflecting alternative explanations inferred from the text. This example demonstrates how AgriBridge-X transforms a single conversational input into a structured representation that captures both entities and their interactions, providing a more interpretable and analyzable view compared to raw textual outputs.

In this graph, different entity types are explicitly separated, enabling users to observe relationships between agricultural problems, applied practices, and their potential impacts. For instance, a practice node may be linked to a problem node through contextual co-occurrence, providing an interpretable view of how farmers describe cause-effect relationships in natural conversations. This graph representation serves as an intermediate layer between unstructured text and downstream analytical tasks. Compared to raw textual outputs, it improves interpretability by explicitly structuring information, and facilitates applications such as knowledge exploration, decision support, and aggregation of patterns across multiple documents. While the current evaluation primarily focuses on entity extraction performance, the graph construction step highlights the potential of AgriBridge-X to support knowledge-centric applications. A systematic quantitative evaluation of graph quality (e.g., relation correctness or structural consistency) is left as future work.

6 Conclusion and Future Work

In this paper, we introduced AgriBridge-X , a unified framework that extends bilingual translation systems toward structured knowledge extraction from web-based agricultural conversations. Departing from translation-centric paradigms, we reframe translation as an intermediate normalization step, enabling the transformation of linguistically heterogeneous and dialect-rich conversational data into structured, analyzable knowledge representations.

A key contribution of this work lies in the introduction of a graph-based knowledge representation layer, where extracted entities (Problems, Practices, Actors, Impacts)

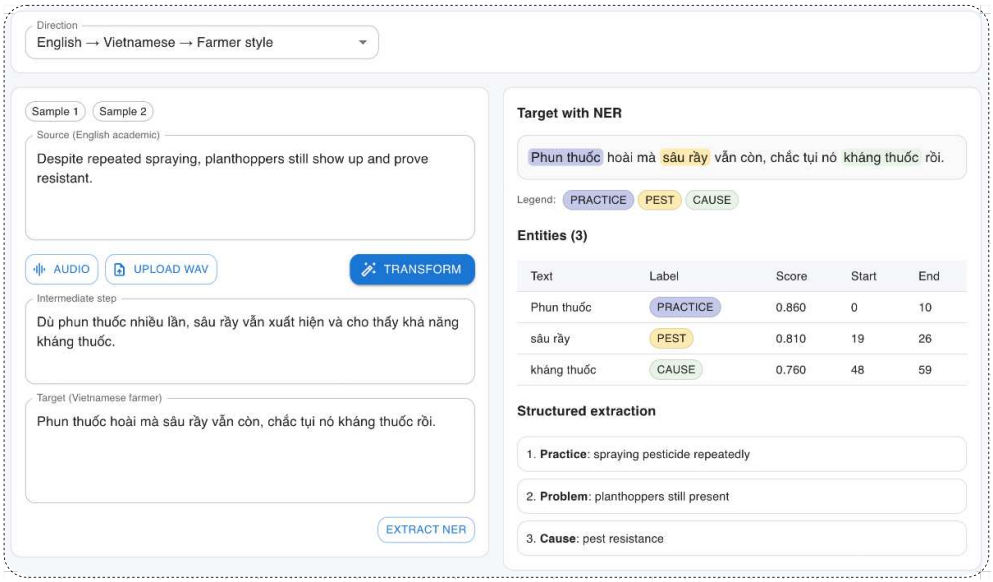


Fig. 7 AgriBridge-X interface showcasing the transformation from English to Vietnamese academic style and then to Farmer style, along with entity extraction results.

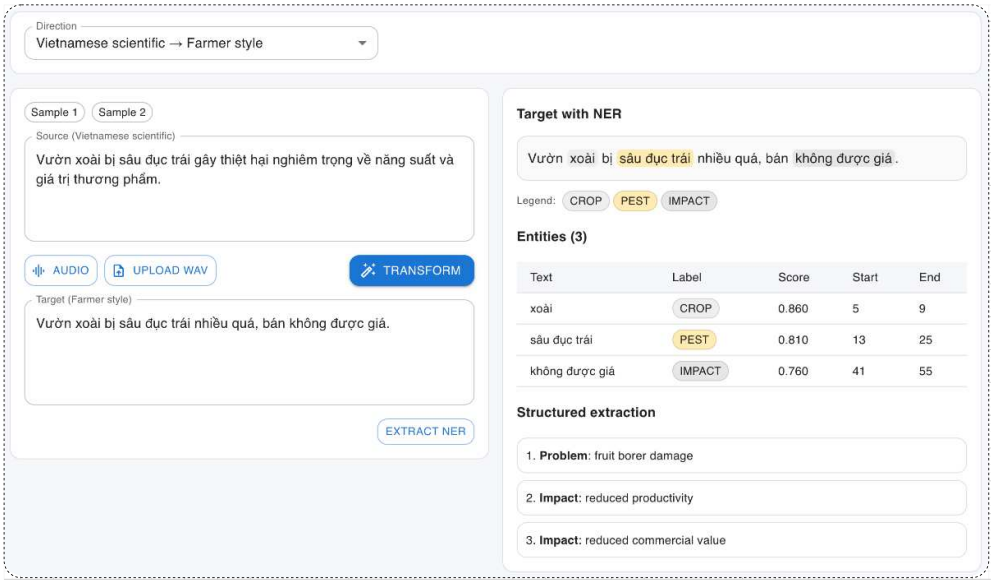


Fig. 8 AgriBridge-X interface demonstrating the transformation from Vietnamese academic style to Farmer style, with highlighted entities and structured output.

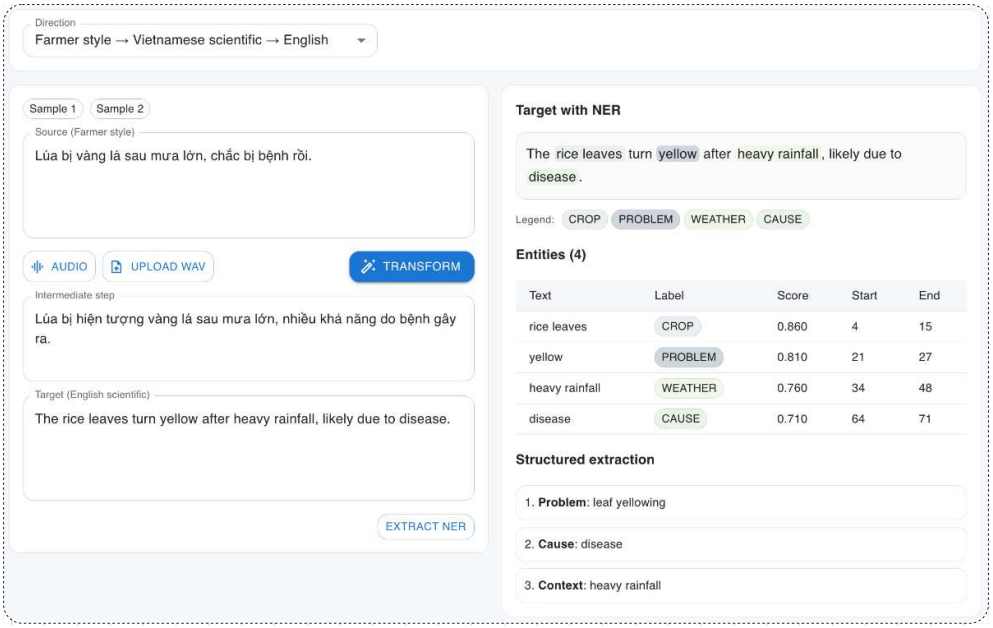


Fig. 9 AgriBridge-X interface illustrating the transformation from Farmer style to Vietnamese academic style and then to English academic style, along with entity extraction results.

are organized into structured graphs capturing implicit relationships in conversational data. This representation bridges the gap between unstructured text and knowledge-centric applications, improving interpretability and enabling downstream reasoning. From a methodological perspective, the proposed framework integrates speech processing, sequence-to-sequence translation, and lexicon-guided extraction within a coherent pipeline. In addition, we provide a theoretical perspective on the stability of the pipeline, establishing connections between semantic preservation in translation and robustness in downstream extraction. These contributions collectively highlight the novelty of AgriBridge-X in combining linguistic normalization, knowledge extraction, and structured representation. More broadly, AgriBridge-X contributes to a shift from communication-oriented AI systems toward *knowledge-centric web intelligence*, where conversational data are transformed into persistent and analyzable knowledge structures. In the agricultural domain, this enables the systematic capture of innovation signals emerging from farmer–expert interactions, supporting data-driven analysis and decision-making.

Future work will focus on advancing the graph-based component of the framework. In particular, we plan to investigate explicit relation extraction and graph refinement techniques to improve the quality and consistency of the constructed knowledge graphs. Another promising direction is the integration of large language models for adaptive and context-aware extraction, enabling more robust handling of implicit and complex expressions. We also aim to explore graph-based reasoning and inference mechanisms to support higher-level tasks such as recommendation, causal

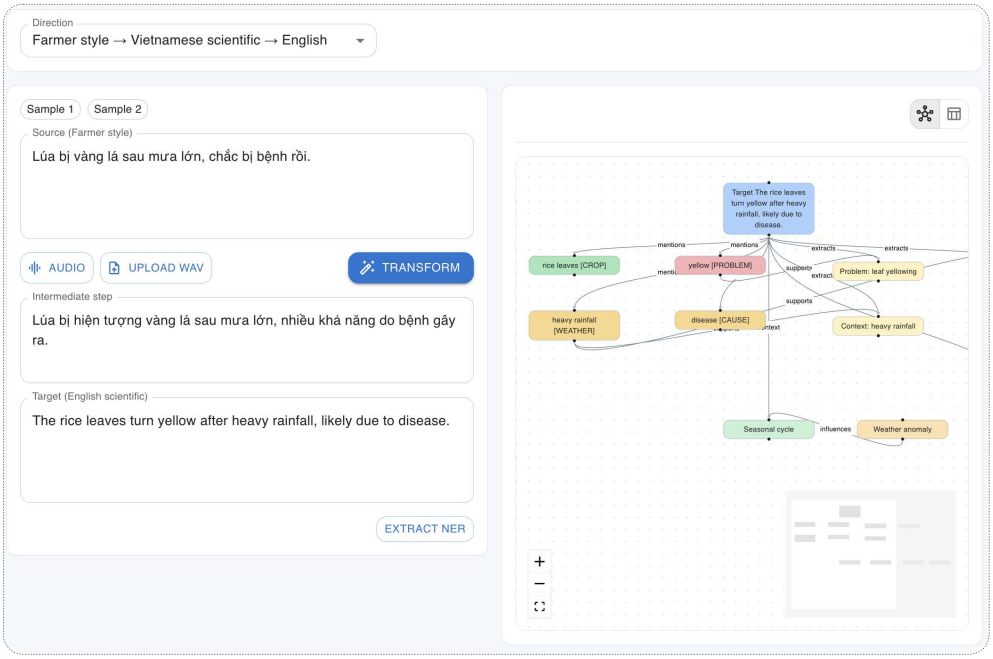


Fig. 10 Knowledge graph representation of extracted entities and their relationships from the transformed text, demonstrating the structured output capability of AgriBridge-X .

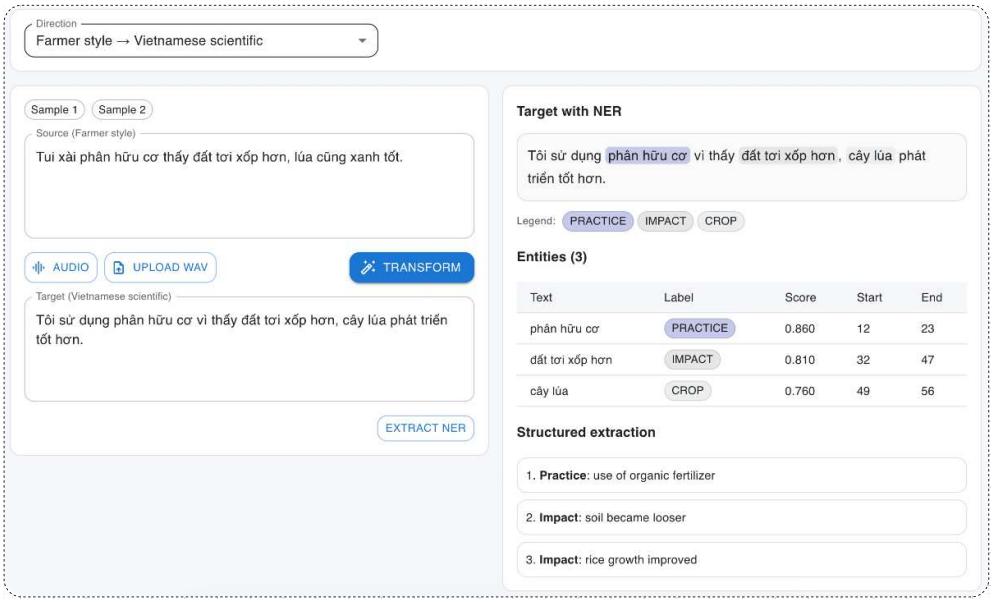


Fig. 11 AgriBridge-X interface showing the transformation from Farmer style to Vietnamese academic style, with highlighted entities and structured output.

Table 9 Comparison of agricultural expressions in farmer style, Vietnamese academic style, and English academic style to demonstrate the stylistic and semantic differences across registers.

Vietnamese–Farmer	**Vietnamese–Academic**	**English–Academic**
Phun thuoc hoai ma sau ray van con, chac tui no khang thuoc roi.	Du phun thuoc nhieu lan, sau ray van xuat hien va chung to chung khang thuoc.	Despite repeated spraying, planthoppers still show up and prove resistant.
Tui xai phan huu co thay dat toi xop hon, lua cung xanh tot	Toi su dung phan huu co vi thay dat toi xop hon, cay lua phat trien tot hon.	I used compost because the soil was more loose and the rice plants grew better.
Tui thay nuoi xen canh ca voi lua do ton phan, lua cung it sau benh	Toi thay nuoi xen canh ca va lua giup tiet kiem chi phi phan bon va giam sau benh.	I found that intercropping of fish and rice saved fertilizer costs and reduced pests.
Buoi nam nay trai it, chac do nang qua	Buoi nam nay cho trai it, nguyen nhan do nang nong gay gat.	This year's grapefruit produced little fruit due to extreme heat.
Vuon xoai bi sau duc trai nhieu qua, ban khong duoc gia	Vuon xoai bi sau duc trai gay thiet hai nghiem trong ve nang suat va gia tri thuong pham.	Mango garden suffers from fruit borers causing serious damage to productivity and commercial value.

analysis, and knowledge discovery. Finally, a systematic quantitative evaluation of graph quality and cross-model comparisons remains an important direction for future research. We envision that AgriBridge-X can be extended to other domains characterized by heterogeneous and conversational knowledge sources, such as healthcare and community-driven innovation platforms.

Acknowledgements: This work is partially funded by FAO and European Union in the context of the Star-Farm project. This work is also supported by ANR (National Research Agency) through the programme Investissements d'avenir #DigitAg (ANR-16-CONV-0004) and the BEYOND project (20-PCPA-0002).

Declarations

Ethical Approval

This study does not involve human participants, animals, or any clinical or experimental procedures requiring approval from an ethical committee or institutional review board (IRB). The data used in this research were collected from publicly accessible online sources, including web-based agricultural discussion videos. All data were processed in accordance with relevant ethical guidelines, and no personally identifiable information (PII) was collected or used. Therefore, ethical approval is not applicable.

Consent to Participate

Not applicable.

Consent to Publish

Not applicable.

Funding
This work is partially funded by the Food and Agriculture Organization (FAO) and the European Union under the Star-Farm project. It is also supported by the French National Research Agency (ANR) through the "Investissements d'avenir" programme #DigitAg (ANR-16-CONV-0004) and the BEYOND project (ANR-20-PCPA-0002).

Conflicts of Interest The authors declare that they have no conflict of interest.

Data Availability
The datasets used in this study are derived from publicly available sources and will be made publicly accessible to support reproducibility.

References

[1] Nayyar, D. *Asian transformations: An inquiry into the development of nations* (Oxford University Press, 2019).

[2] Martin, F. & Dragojlovic, A. Gender, mobility regimes, and social transformation in asia (2019).

[3] Steinke, J., van Etten, J. & Zelan, P. M. The accuracy of farmer-generated data in an agricultural citizen science methodology. *Agronomy for Sustainable Development* **37**, 32 (2017).

[4] Chowdhary, K. Natural language processing. *Fundamentals of artificial intelligence* 603–649 (2020).

[5] Hirschberg, J. & Manning, C. D. Advances in natural language processing. *Science* **349**, 261–266 (2015).

[6] Ma, T. *et al.* Agribridge: A translation system between farmers and scientists. *International Conference on Management of Digital (MEDES'25)* 0–10 (2025).

[7] Ayadi, N. Y. *et al.* A unified approach to publish semantic annotations of agricultural documents as knowledge graphs. *Smart Agricultural Technology* **8**, 100484 (2024).

[8] Mohamed, Y. A. *et al.* The impact of artificial intelligence on language translation: a review. *IEEE access* **12**, 25553–25579 (2024).

[9] Yalew, S., Van Griensven, A. & van der Zaag, P. Agrisuit: A web-based gis-mcda framework for agricultural land suitability assessment. *Computers and Electronics in Agriculture* **128**, 1–8 (2016).

[10] Cho, J. *et al.* Web-based agricultural infrastructure digital twin system integrated with gis and bim concepts. *Computers and Electronics in Agriculture* **215**, 108441 (2023).

[11] Aiken, M. & Govindarajulu, C. Knowledge-based information retrieval for group decision support systems. *Journal of Database Management (JDM)* **5**, 31–35 (1994).

[12] Yao, Y. Information retrieval support systems. *2002 IEEE World Congress on Computational Intelligence. 2002 IEEE International Conference on Fuzzy Systems. FUZZ-IEEE'02. Proceedings (Cat. No. 02CH37291)* **2**, 1092–1097 (2002).

[13] Cao, J., Tang, S., Li, J., Zhang, Y. & Pan, X. A lexicon-guided lsi method for semantic news video retrieval. *Pacific-Rim Conference on Multimedia*

(PRCM'07) 187–195 (2007).

[14] Brown, S. W. *et al.* Semantic representations for nlp using verbnet and the generative lexicon. *Frontiers in artificial intelligence* **5**, 821697 (2022).

[15] Gagnon, M., Zouaq, A. & Jean-Louis, L. Can we use linked data semantic annotators for the extraction of domain-relevant expressions? *Proceedings of the 22nd International Conference on World Wide Web* 1239–1246 (2013).

[16] Ducrot, R. *et al.* STAR-FARM Lexicon for Agroecology and Innovation (English/Vietnamese) (2026). URL https://doi.org/10.18167/DVN1/RG9KY1.

[17] Wu, Y. & Zhang, X. Examining conversation analysis in palliative care: A systematic review. *Health Communication* **39**, 3072–3083 (2024).

[18] Imperial, J. M. Natural language generation with expert standards. *Proceedings of the AAAI Conference on Artificial Intelligence* **39**, 29267–29268 (2025).

[19] Martinez-Rodriguez, J. L., Hogan, A. & Lopez-Arevalo, I. Information extraction meets the semantic web: a survey. *Semantic Web* **11**, 255–335 (2020).

[20] Sorokin, D. & Gurevych, I. Context-aware representations for knowledge base relation extraction. *Proceedings of the 2017 conference on empirical methods in natural language processing* 1784–1789 (2017).

[21] Koopman, B., Zuccon, G., Chapman, S., Dang, Y. & Lawrence, D. How a conversational agent might help farmers in the field. *ACM Reference Format* (2020).

[22] Burgess, R. G. The unstructured interview as a conversation. *Field research* 177–182 (2003).

[23] Eldem, A. & Eldem, H. The development and evaluation of agricultural question-answering systems based on large language models. *Scientific Reports* **16**, 5357 (2026).

[24] Li, H., Wu, H., Li, Q. & Zhao, C. A review on enhancing agricultural intelligence with large language models. *Artificial Intelligence in Agriculture* **15**, 671–685 (2025).

[25] Fielke, S., Taylor, B. & Jakku, E. Digitalisation of agricultural knowledge and advice networks: A state-of-the-art review. *Agricultural systems* **180**, 102763 (2020).

[26] Zheng, Y.-l., Ping, Q., Ze, L. *et al.* Construction of the ontology-based agricultural knowledge management system. *Journal of Integrative Agriculture* **11**, 700–709 (2012).

[27] Priya, N. K. *et al.* Digital innovations in extension education: A review of emerging technologies and their impact on agricultural knowledge dissemination. *Journal of Experimental Agriculture International* **47**, 408–423 (2025).

[28] Shaikh, T. A., Rasool, T., Veningston, K. & Yaseen, S. M. The role of large language models in agriculture: harvesting the future with llm intelligence. *Progress in Artificial Intelligence* **14**, 117–164 (2025).

[29] Balaguer, A. *et al.* Rag vs fine-tuning: pipelines, tradeoffs, and a case study on agriculture. *arXiv preprint arXiv:2401.08406* (2024).

[30] Do, Q. T., Sakti, S. & Nakamura, S. Sequence-to-sequence models for emphasis speech translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **26**, 1873–1883 (2018).

[31] Zaki, M. Z. & Ahmed, U. Bridging linguistic divides: The impact of ai-powered translation systems on communication equity and inclusion. *Journal of translation and language studies* **5**, 20–30 (2024).

[32] Pawlicka, A., Pawlicki, M., Kozik, R., Andrychowicz-Trojanowska, A. & Choraś, M. Ai vs linguistic-based human judgement: Bridging the gap in pursuit of truth for fake news detection. *Information Sciences* **679**, 121097 (2024).

[33] Zaki, M. Z. Revolutionising translation technology: A comparative study of variant transformer models–bert, gpt and t5. *Computer Science and Engineering–An International Journal* **14**, 15–27 (2024).

[34] Lewis, M. *et al.* Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *Proceedings of the 58th annual meeting of the association for computational linguistics* 7871–7880 (2020).

[35] Lee, E.-S. A. *et al.* Pre-trained multilingual sequence-to-sequence models: A hope for low-resource language translation? *Findings of the Association for Computational Linguistics: ACL 2022* 58–67 (2022).

[36] Bigi, B. A multilingual text normalization approach. *Language and Technology Conference* 515–526 (2011).

[37] Thiam, P. I., Chasseray, Y., Mothe, J., Roche, M. & Teisseire, M. Enhancing domain-specific named entity recognition via segmentation and pseudo-labeled annotation. *2025 IEEE 37th International Conference on Tools with Artificial Intelligence (ICTAI)* 552–559 (2025).

[38] Pavlick, E. Semantic structure in deep learning. *Annual Review of Linguistics* **8**, 447–471 (2022).

[39] Jackendoff, R. S. *Semantic structures* Vol. 18 (MIT press, 1992).

[40] Enten, F. *et al.* ASSET Practical Lexicon: An agroecology lexicon in Vietnam (2023). URL https://doi.org/10.18167/DVN1/EMJIR6.

[41] Chloé, A. *et al.* Innovation lexicon (2024). URL https://doi.org/10.18167/DVN1/VMHTYL.

[42] Alexandre, C. *et al.* A lexicon about innovation in research and innovation projects for agrifood system transformation and sustainability transitions. *Data in Brief* 112792 (2026). URL https://www.sciencedirect.com/science/article/pii/S2352340926003458.

[43] Radford, A. *et al.* Robust speech recognition via large-scale weak supervision. *International conference on machine learning* 28492–28518 (2023).

[44] Wołk, K. & Marasek, K. Enhanced bilingual evaluation understudy. *Lecture Notes on Information Theory Vol* **2** (2014).

[45] Lin, C. Recall-oriented understudy for gisting evaluation (rouge). *Retrieved August* **20** (2005).

[46] Lin, C.-Y. Rouge: A package for automatic evaluation of summaries. *Text summarization branches out* 74–81 (2004).

[47] Zaratiana, U., Tomeh, N., Holat, P. & Charnois, T. Duh, K., Gomez, H. & Bethard, S. (eds) *GLiNER: Generalist model for named entity recognition using bidirectional transformer*. (eds Duh, K., Gomez, H. & Bethard, S.) , 5364–5376 (Association for Computational Linguistics, Mexico City, Mexico, 2024). URL https://aclanthology.org/2024.naacl-long.300/.

[48] Mayhew, S. *et al.* Universal ner: A gold-standard multilingual named entity recognition benchmark. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)* 4322–4337 (2024).

[49] Abdul, W. M. *et al.* Named clinical entity recognition benchmark. *arXiv preprint arXiv:2410.05046* (2024).