Terminology-Augmented Generation for Intangible Cultural Heritage: A Controlled LLM-Based Translation Framework

Wanda Punzi Zarino[1], **Pilar Sánchez Gijón**[2]
Department of Economics and Law, University of Naples Parthenope, Italy[1]
Department of Translation, Interpreting and East Asian Studies, Autonomous University of Barcelona, Spain[2]
Via Acton, 38 - 80133 Naples, Italy[1]
Edifici K, Plaça del Coneixement - 08193 Bellaterra, Spain[2]
wanda.punzizarino001@studenti.uniparthenope.it, Pilar.Sanchez.Gijon@uab.cat

Abstract

This study examines the integration of a bilingual Italian–Spanish concept-oriented terminological resource into a controlled large language model (LLM) translation workflow within the domain of Campanian gastronomy. The termbase encodes structured conceptual, linguistic, and translational metadata, including grammatical information, translation strategies, and genre-sensitive usage recommendations. Through a local Model Context Protocol (MCP) architecture, the resource is dynamically connected to locally deployed LLMs, enabling the automatic identification and retrieval of relevant terminological units prior to generation. The system combines in-context terminological injection with deterministic post-processing enforcement: genre-specific policies are injected into the model prompt prior to generation and verified through a rule-based post-processing layer that enforces surface-level terminological consistency in the output. Two open-weight models — Mistral 7B Instruct and Gemma3 4B — are evaluated across three conditions and three discursive genres on a dataset of authentic texts. The findings suggest that the combination of terminological injection and deterministic enforcement can improve terminological compliance in controlled, domain-specific settings, while also highlighting differences in instruction-following behavior across models and genres.

Keywords: Terminology-augmented generation, Bilingual IT-ES termbase, Intangible cultural heritage

1. Introduction

The UNESCO Convention for the Safeguarding of the Intangible Cultural Heritage [UNESCO, 2003] defines intangible heritage as encompassing practices, knowledge, and skills transmitted across generations. Within this framework, gastronomy emerges as a culturally dense domain in which culinary practices intertwine material production, ritual traditions, territorial identity, and linguistic expression. Terminological units in this domain cannot be reduced to mere technical labels; rather, following Cabré [1993, 2003], they can be understood as linguistic realizations of structured conceptual knowledge, encoding production processes, historical trajectories, and socio-cultural values [Chessa et al., 2014, Grimaldi, 2017, Buccheri, 2023].

When such culturally embedded terminology enters translation contexts, additional complexity arises, as translation choices affect not only lexical form but also conceptual integrity and cultural mediation. The degree of communicative constraint further modulates terminological stability across text types [Sabatini, 1990, 1999], while established translation techniques provide an operational framework for addressing conceptual asymmetries in gastronomic contexts [Molina and Hurtado Albir, 2002].

Against this background, this paper explores the application of the Terminology-Augmented Generation (TAG) framework to the domain of Campanian traditional agri-food products. A concept-oriented Italian–Spanish bilingual termbase was developed, encoding structured conceptual and linguistic metadata, including grammatical information, translation strategy labels (e.g., borrowing, established equivalent, literal translation), and genre-sensitive usage recommendations. The resource is integrated into an LLM-based translation workflow through a local Model Context Protocol (MCP) server, which enables the deterministic retrieval of relevant terminological units prior to translation. The system combines in-context terminological injection with a deterministic post-processing enforcement layer, ensuring that both lexical and structural genre constraints are applied consistently. Two open-weight models — Mistral 7B Instruct and Gemma3 4B — are evaluated in parallel, allowing for a cross-model comparison that sheds light on the relationship between instruction-following capacity and the effectiveness of terminology-augmented generation.

Three research questions guide this study: RQ1: Can a concept-oriented bilingual termbase regulate terminological compliance across discursive genres in LLM-based translation? RQ2: Does the combination of in-context terminological injection and deterministic post-processing enforcement improve control over terminological realization compared to either component alone? RQ3: To what extent do differences in instruction-following capac-

Proceedings of the 2nd Workshop on Evaluating Text Difficulty in a Multilingual Context (DeTermIt! 2026), pages 74–82
11 May 2026.

ity across open-weight models affect the effectiveness of terminology-augmented generation?

The remainder of the paper is structured as follows. Section 2 reviews related work on RAG and TAG architectures. Section 3 introduces the corpus and terminological data. Section 4 describes the case study and the MCP-based integration architecture. Section 5 presents and discusses the results. Section 6 concludes with reflections on future research directions.

2. Background and Related Work

In recent years, the rapid development of Large Language Models (LLMs) has led to growing interest in integrating structured terminological resources into generative AI systems. One of the most influential architectural paradigms in this context is Retrieval-Augmented Generation (RAG), introduced by Lewis et al. [2020], which combines neural text generation with the dynamic retrieval of external documents. While RAG improves factual grounding by retrieving contextually relevant passages at inference time, it is primarily designed for unstructured text retrieval based on dense vector similarity and does not inherently support deterministic access to structured, concept-oriented terminological data. Subsequent studies have highlighted both the strengths and limitations of RAG, particularly with regard to noise in retrieved content and limited control over fine-grained domain-specific knowledge [Gupta et al., 2024, Lackner et al., 2025b].

Within specialized domains, researchers have begun exploring terminology-enhanced variants of RAG. Martín-Chozas et al. [2025], for instance, show that combining neural retrieval with curated lexical resources can improve retrieval effectiveness and answer quality in legal corpora. However, these approaches remain embedded within the broader RAG paradigm and continue to rely on similarity-based mechanisms.

In response to these structural limitations, Terminology-Augmented Generation (TAG) has emerged as a complementary paradigm to RAG [Fleischmann, 2025], enabling direct and deterministic access to structured termbases. The theoretical foundations of TAG have been further articulated by Di Nunzio [2025], who frames it as a generative architecture grounded in the dual conceptual and linguistic dimensions of terminology science, advocating a design that includes a terminology access layer, filtering and reasoning components, and human-in-the-loop validation mechanisms. Empirical studies have begun to substantiate its effectiveness. Lackner et al. [2025a] show that TAG significantly improves terminological adherence in LLM-based machine translation, while Lackner et al. [2025b] demonstrate that lightweight structured formats such as YAML and JSON enhance model performance in in-context learning settings. They further observe that Mistral 7B consistently produced unusable outputs in multilingual translation tasks — a finding that informs the comparative dimension of the present study. Taken together, these contributions position TAG as a robust framework for embedding curated terminological knowledge into generative pipelines, particularly in multilingual and domain-sensitive contexts where precision and expert validation are essential.

3. Corpus and Terminological Data

The corpus underpinning this study is derived from the official register of *Prodotti Agroalimentari Tradizionali* (PAT) of the Campania region [Regione Campania, 2025]. At the time of data collection, the regional register comprised 610 officially recognized products. For the purposes of this study, a subset of 28 entries was selected, corresponding to the subdomain of gastronomic preparations. This selection reflects a methodological choice to prioritize gastronomic preparations over raw materials, as culinary terms tend to exhibit greater terminological and translational complexity. Unlike names referring to primary agricultural products, they encode procedural knowledge, preparation methods, seasonal consumption patterns, and communal practices embedded in specific territorial, ritual, and socio-historical contexts.

The repertory displays marked lexical heterogeneity. Some terms include dialectal or regionally marked forms, such as *ciauliello*, whose semantic opacity may challenge non-local readers and translators and whose resistance to straightforward lexical equivalence reflects its function as a marker of territorial identity. Others reveal historically motivated lexical forms: *frittata di scammaro*, for instance, derives from socio-religious dietary practices associated with Lenten observance in the Kingdom of the Two Sicilies, where *scammaro* referred to “lean days” during which conventual cooking excluded eggs and meat. The corpus also includes terms anchored in festive contexts, such as *cicci di Santa Lucia*, which situates the preparation within a specific ritual calendar (13 December) and within the communal practices of Avellino and its surrounding area.

4. Case Study

The termbase, encoded in YAML format in line with recent findings on lightweight structured representations for in-context learning [Lackner et al., 2025a,b], constitutes the knowledge layer underpinning the experiment. Rather than functioning as a static glossary, it encodes formally structured

conceptual and translational data that can be programmatically accessed at run time. Each entry is organized into three layers. The conceptual layer encodes a concept identifier, domain classification, definitional content, production notes, cultural notes, area of production, and a source field referencing the official PAT register. The linguistic layer provides grammatical and lexical metadata for both Italian and Spanish, including part of speech, gender, and registered variants. The translational layer specifies translation strategy labels, preferred target equivalents, and genre-sensitive usage rules.

The experimental dataset consists of 30 authentic Italian texts distributed across three discursive genres: regulatory-administrative, promotional-touristic, and narrative-cultural (10 texts per genre). Rather than constructing artificial sentences designed to contain specific terminological units, the texts were drawn from real-world sources, including official PAT product register entries, gastronomic journalism, promotional food blogs, and literary or narrative culinary writing. Each text contains at least one terminological unit from the PAT-Campania termbase. The three genres were selected to represent distinct points on Sabatini's typological cline of communicative constraint [Sabatini, 1990, 1999], ranging from the highly constrained regulatory-administrative register, characterized by strict terminological stability and institutional formatting conventions, to the more loosely constrained narrative-cultural register, in which voice, perspective, and affective engagement take precedence over formal precision.

Integration into the translation workflow was achieved through a local MCP server, which exposes terminological content as callable tools accessible to the orchestration layer. In line with the conceptual framework proposed by Di Nunzio [2025], the system integrates a terminology access component together with filtering, reasoning, and generation modules, extended with a deterministic post-processing enforcement layer. Two open-weight models were evaluated in parallel — Mistral 7B Instruct and Gemma3 4B — both accessed through the Ollama framework. The entire system was implemented and executed locally within an isolated Python virtual environment, ensuring controlled dependency management, modular deployment, and reproducibility. No external retrieval services or third-party APIs were employed. The overall workflow is illustrated in Figure 1.

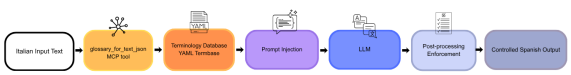

Figure 1: Terminology-augmented generation architecture.

The terminology retrieval phase follows a deterministic sequence prior to generation. Given an Italian source text, the system first invokes the MCP tool *glossary_for_text_json*, which performs automatic identification of relevant terminological units through boundary-aware Unicode-safe matching. The tool retrieves all matching entries from the termbase, including single-word terms, multi-word expressions, and registered variants, together with structured IT–ES correspondences, translation strategy labels, and genre-sensitive usage rules. When genre-specific modulation is required, the system subsequently invokes the *translate_usage* tool to retrieve the translation policy associated with a selected discursive genre. The retrieved policy specifies the output form, the preferred Spanish equivalent, and the field from which the expansion note should be drawn: production notes for promotional-touristic texts, cultural notes for narrative-cultural texts.

The generation module receives a shared system prompt (Appendix A) that defines the translator role and general translation instructions, together with a genre-specific glossary block (Appendix B) injected into the user message. The glossary block is constructed dynamically from the retrieved terminological policies: in regulatory-administrative contexts, the model is instructed to preserve the original Italian term formatted with Spanish angle quotation marks; in promotional-touristic contexts, to use the preferred Spanish equivalent immediately followed by a parenthetical expansion, reformulated based on the context and co-text of the passage; in narrative-cultural contexts, to introduce the preferred Spanish equivalent with an inline cultural gloss introduced by a comma. Crucially, the glossary block is not a static template but a dynamically constructed instruction set that reflects the specific terminological units detected in each source text and the genre rules encoded in the termbase.

Once the output is generated, the same genre-specific policies are applied deterministically through the post-processing enforcement layer. For regulatory-administrative texts, the enforcement layer verifies whether the Italian term is present in the output: if found without angle quotation marks it adds them, and if the preferred Spanish equivalent appears instead of the Italian term it replaces it with the angle-quoted Italian form. For promotional-touristic and narrative-cultural texts, the enforcement layer verifies whether the preferred Spanish equivalent is present: if absent but the Italian source term is found and differs from the Spanish equivalent, it substitutes the latter. The enforcement layer does not intervene on parenthetical expansions or inline glosses — a conservative design choice that avoids the risk of inserting contextually

inappropriate content.

5. Analysis of Results

5.1. Terminology Adherence

Terminology Adherence (TA) was computed automatically for all three conditions across both models and genres. For regulatory-administrative texts, TA measures whether the Italian source term is present in the model output; for promotional-touristic and narrative-cultural texts, it measures whether the preferred Spanish equivalent as specified in the termbase is present. Term presence was verified through boundary-aware, Unicode-safe string matching with apostrophe normalization, and TA was computed as a proportional score per text, averaged across texts within each genre and condition. Results are reported in Tables 1, 2, and 3.

The baseline condition reveals a fundamental limitation of unconstrained LLM-based translation: both models spontaneously preserve the expected terminology in fewer than half of the cases globally — and as rarely as 15% for narrative-cultural texts. Culturally opaque and dialectally marked terms exert no statistical pressure toward preservation in models trained on general multilingual corpora; in the absence of explicit guidance, probabilistic generation tends to resolve terminological opacity through free translation, paraphrase, or domestication. This is illustrated by cases such as *cicatielli con pulieio*, rendered as *cicatelli con pulígaro* by Mistral — a non-existent form resulting from the misinterpretation of the dialectal term *pulieio* — or *frittata di scammaro*, rendered as *frittata de escarola* by Gemma, where *scammaro* is erroneously mapped to *escarola* (chicory), obliterating the socio-religious meaning encoded in the term. The baseline thus operationalizes what the TAG framework is designed to address: the structural inability of unconstrained LLMs to govern their own terminological choices in specialized, culturally dense domains.

A notable asymmetry emerges at the baseline level between the two models on regulatory-administrative texts: Gemma3 4B preserves Italian denominations in 70.0% of cases, compared to 40.0% for Mistral 7B Instruct. This divergence points to a broader pattern observed throughout the evaluation: the two models respond to terminological guidance in qualitatively different ways, and their behavior cannot be reduced to a simple ranking in terms of general translation capability.

The introduction of in-context terminological injection produces consistent improvements across all genres and both models. Nevertheless, the injection condition also reveals a systematic weakness in Mistral's compliance behavior on promotional-touristic texts, where TA reaches only 50.0% against Gemma's 80.0%. This asymmetry is consistent with the findings of Lackner et al. [2025b], who observed that Mistral 7B produced unusable outputs in multilingual translation tasks.

The addition of the post-processing enforcement layer partially compensates for this limitation, raising Mistral's global TA from 70.0% to 85.0% and narrowing the gap with Gemma, which reaches 90.0% under INJ+ENF. The convergence of both models to 100.0% TA on regulatory-administrative texts under the full system condition shows that deterministic enforcement can reliably recover cases in which injection alone fails. The remaining gap on promotional-touristic texts — where Mistral reaches only 65.0% even under INJ+ENF — points to a different failure mode: cases in which the model translates the source term so freely that no recoverable surface form remains, making enforcement ineffective.

Genre	BASE	INJ	INJ+ENF
Regulatory	40.0%	80.0%	100.0%
Promotional	35.0%	50.0%	65.0%
Narrative	15.0%	80.0%	90.0%
Global	**30.0%**	**70.0%**	**85.0%**

Table 1: TA — Mistral 7B Instruct.

Genre	BASE	INJ	INJ+ENF
Regulatory	70.0%	100.0%	100.0%
Promotional	35.0%	80.0%	85.0%
Narrative	15.0%	85.0%	85.0%
Global	**40.0%**	**88.3%**	**90.0%**

Table 2: TA — Gemma3 4B.

5.2. Genre Constraint Compliance: Manual Evaluation

Manual evaluation was conducted on 120 translations produced under the INJ and INJ+ENF conditions across both models and all genres using the Genre Constraint Compliance (GCC; Table 4). GCC evaluates the quality of terminological modulation on a three-point scale: 0 indicates that no modulation is provided; 1 indicates that a modulation is present but consists of a partial or verbatim reproduction of the termbase note, resulting in an expansion that is not fully integrated into the surrounding co-text; and 2 that the modulation is contextually adapted to the passage and its co-text. For regulatory-administrative texts, the score captures typographical compliance rather than reformulation, since no modulation is required in this genre: 0 indicates missing angle quotation marks, 1 partial

Genre	Mistral INJ	Gemma INJ	Mistral INJ+ENF	Gemma INJ+ENF
Regulatory-administrative	80.0%	100.0%	100.0%	100.0%
Promotional-touristic	50.0%	80.0%	65.0%	85.0%
Narrative-cultural	80.0%	85.0%	90.0%	85.0%
Global	**70.0%**	**88.3%**	**85.0%**	**90.0%**

Table 3: TA cross-model comparison: Mistral 7B Instruct vs Gemma3 4B.

compliance, and 2 correct formatting.

The results reveal three main patterns across genres and models: consistent typographical compliance in regulatory-administrative texts, particularly for Gemma; systematic prompt leakage in Mistral across promotional and narrative genres; and a tendency in both models to extend the instructed explicitation behavior beyond the boundaries of the termbase.

Genre	Model	INJ	INJ+ENF
Regulatory	Mistral	0.90	1.40
	Gemma	2.00	2.00
Promotional	Mistral	0.10	0.90
	Gemma	0.60	0.60
Narrative	Mistral	0.00	0.00
	Gemma	0.70	0.50

Table 4: GCC by genre, model, and condition (0–2 scale).

The regulatory-administrative genre yields the clearest results. Gemma3 4B achieves a perfect GCC of 2.00 under both INJ and INJ+ENF, indicating that in every evaluated case the Italian term is correctly preserved with the required angle quotation marks. Mistral 7B reaches 0.90 under INJ and 1.40 under INJ+ENF — a pattern that reflects inconsistent typographical compliance rather than terminological failure: the term is present, but the formatting convention is not reliably applied. These results converge on the same conclusion as Section 5.1: for the most formally constrained genre in Sabatini's typology, Gemma3 4B is a more reliable partner for TAG-based translation.

The promotional-touristic genre reveals a more nuanced picture. Gemma achieves a GCC of 0.60 under both INJ and INJ+ENF, indicating that when parenthetical expansions are produced they tend toward partial rather than full contextual reformulation. Mistral's GCC of 0.10 under INJ reflects a qualitatively different failure mode: in the majority of annotated cases, expansions were either absent or constituted prompt leakage — content appended after the translation as external notes, often reproducing verbatim portions of the glossary block or the background note. This behavior is not captured by TA alone and underscores the importance of manual evaluation for a complete assessment of TAG effectiveness. A related pattern, observable in both models, concerns terminological overgeneralization: when instructed to add expansions for specific terminological units, both models occasionally extend this behavior to gastronomic terms present in the surrounding text but not covered by the termbase. This suggests that the models internalize the structural pattern of the glossary block instruction and replicate it beyond its intended scope. The phenomenon is particularly systematic in Mistral under INJ+ENF, where parenthetical expansions are added for *panzanella* (*ensalada de pan rallado, tomates, cebolla y aceituna*), *pappa al pomodoro* (*sopa de pan rallado y tomate*), and *fusilli alla 'nduja* (*fusilli con salsa de salchichón calabrés*), none of which appear in the termbase. In one particularly noteworthy case, Gemma3 4B applies a contextually integrated gloss to *'O rraù* — a dialectal Neapolitan form of *ragú napolitano* that does not appear as such in the termbase, but which the model appears to recognize as a variant of a known entry, producing *una preparación típica del almuerzo dominical napolitano, considerada un emblema del patrimonio enogastronómico de Campania*. The expansion is semantically grounded in the termbase entry for *ragú napolitano* and is positioned in a contextually motivated way, facilitating the target reader's comprehension of an otherwise opaque dialectal form. While none of these expansions are explicitly instructed by the termbase, they point to a broader tendency of instruction-following models to replicate prompted explicitation patterns beyond their intended scope — a behavior that, under favorable conditions, may produce genuinely useful output.

The narrative-cultural genre produces the sharpest contrast between the two models. Mistral achieves a GCC of 0.00 under both INJ and INJ+ENF — not because no expansions are produced, but because every expansion annotated was classified as prompt leakage or hallucination: content generated outside the translated text, disconnected from the surrounding narrative co-text, and frequently reproducing background note material verbatim or introducing extraneous information. Manual verification confirms that the preferred term is present in 8 out of 10 cases under both conditions, yet the model consistently

fails to integrate the required inline gloss, instead appending expansions as external notes that violate both the structural and the stylistic requirements of narrative prose. Gemma, by contrast, achieves the highest GCC across all genre-model-condition combinations: 0.70 under INJ. A representative case is *frittata di scammaro* (narrative, INJ), where Gemma produces an inline gloss — *un plato típico de la cocina pobre napolitana* — genuinely adapted to the surrounding narrative context, earning a GCC of 2. In one promotional-touristic case, Gemma positions the parenthetical expansion after the anaphoric *menestra ebolitana* rather than immediately after *Ciauliello* — a co-textually motivated placement that, while technically non-compliant with the glossary block instruction, produces a more natural and readable result. The decrease to 0.50 under INJ+ENF may reflect cases where the enforcement layer's term substitution disrupted the natural integration of the gloss — a tension between deterministic correction and contextually sensitive generation that points to a fundamental design challenge for TAG systems operating on loosely constrained text types.

Taken together, these results demonstrate that manual evaluation captures dimensions of TAG effectiveness that automatic TA cannot access. High TA scores are a necessary but not sufficient condition for genre-compliant terminological treatment: what ultimately determines the quality of the output is not merely whether the correct term is present, but whether it is integrated in a way that respects the communicative conventions of the target genre.

6. Conclusions and Future Work

This study has examined whether the integration of a concept-oriented bilingual termbase into an LLM-based translation workflow can effectively regulate terminological compliance and genre-specific realization in the domain of Campanian intangible cultural heritage.

The results suggest that terminology-augmented generation can improve terminological adherence across genres and models, with global TA scores reaching 85.0% for Mistral and 90.0% for Gemma under the full system condition. The combination of in-context injection and deterministic enforcement tends to outperform either component in isolation: injection alone increases adherence but fails to guarantee genre-specific realization, while enforcement cannot operate on terms that have not been properly introduced.

The findings allow us to directly address the research questions. Terminology injection proves effective in regulating terminological compliance across discursive genres (RQ1), with global TA scores increasing consistently from baseline across all genre–model combinations. Injection combined with deterministic enforcement produces uneven effects: while the full system condition consistently outperforms injection alone on TA, the added value of enforcement is strongly genre- and model-dependent — improving terminological realization on regulatory-administrative and promotional-touristic texts for Mistral, but showing limited or no benefit in others, and in some cases interfering with contextually appropriate realization (RQ2).

However, terminological adherence is only one dimension of TAG effectiveness. Manual evaluation reveals that a model can achieve high TA while systematically failing to integrate terms in a genre-appropriate way, as demonstrated by Mistral's narrative-cultural outputs, where the preferred term is present in 8 out of 10 cases yet consistently realized as an external appendix rather than an inline gloss.

The cross-model comparison highlights the central role of instruction-following capacity in TAG effectiveness. Gemma3 4B consistently outperforms Mistral 7B Instruct not only in quantitative terms but also in qualitative behavior, demonstrating a greater ability to integrate terminological modulation into the surrounding discourse. This gap becomes more pronounced in loosely constrained genres, where successful output depends on context-sensitive generation rather than formal compliance alone (RQ3).

Future work should extend the evaluation to larger datasets, additional language pairs, and more capable open-weight models, in order to enable a more systematic investigation of the relationship between model capacity and TAG effectiveness. From a terminological perspective, enabling more flexible access to different types of conceptual metadata — allowing context-driven selection rather than genre-based assignment — may improve the contextual adequacy of terminological modulation in loosely constrained text types. Finally, the development of automatic metrics for contextual reformulation remains an open challenge for scalable evaluation of terminology-aware generation in culturally sensitive translation settings. In this regard, further investigation is needed into how standard MT evaluation metrics, such as BLEU [Papineni et al., 2002] and COMET [Rei et al., 2020], behave under controlled, terminology-augmented generation conditions, where improvements in terminological compliance and genre-sensitive realization may not be fully captured by reference-based similarity.

7. Bibliographical References

L. Buccheri. *Parole del cibo in Campania. Cento voci del lessico gastronomico regionale*. Franco Cesati Editore, Firenze, 2023.

M.-T. Cabré. *La terminología. Teoría, metodología y aplicaciones*. Antártida/Empúries, Barcelona, 1993.

M.-T. Cabré. Theories of terminology: Their description, prescription and explanation. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 9(2):163–199, 2003.

F. Chessa, C. De Giovanni, and M. T. Zanola. *La terminologia dell'agroalimentare*. FrancoAngeli, Milano, 2014.

G. M. Di Nunzio. Terminology-augmented generation (TAG): Foundations, use cases, and evaluation paths. *Journal of Digital Terminology and Lexicography*, 1(1):97–104, 2025.

K. Fleischmann. Terminologiemanagement: Die Schlüsselkomponente für effiziente Kommunikation in Unternehmen. *Information – Wissenschaft & Praxis*, 76(4):169–176, 2025.

C. Grimaldi. *Il prodotto agroalimentare campano. Tra lingua, cultura e tradizione*. Aracne, Roma, 2017.

S. Gupta, R. Ranjan, and S. N. Singh. A comprehensive survey of retrieval-augmented generation (RAG): Evolution, current landscape and future directions. *arXiv preprint arXiv:2410.12837*, 2024.

A. Lackner, A. Vega-Wilson, and C. Lang. Terminology-augmented generation: A systematic review of terminology formats for in-context learning in LLMs. In *Proceedings of the 4th International Conference on Multilingual Digital Terminology Today (MDTT 2025)*, volume 3990 of *CEUR Workshop Proceedings*, 2025a.

A. Lackner, A. Vega-Wilson, and C. Lang. An evaluation of terminology-augmented generation (TAG) and various terminology formats for the translation use case. *Journal of Digital Terminology and Lexicography*, 1(2):31–47, 2025b.

P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Roktäschel, S. Riedel, and D. Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems*, volume 33, pages 9459–9474, 2020.

P. Martín-Chozas, P. Calleja, and C. R. Limón. Terminology Enhanced Retrieval Augmented Generation for Spanish Legal Corpora. In *Proceedings of the 5th Conference on Language, Data and Knowledge*, pages 147–152, 2025.

L. Molina and A. Hurtado Albir. Translation techniques revisited: A dynamic and functionalist approach. *Meta: Translators' Journal*, 47(4):498–512, 2002.

K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, 2002.

Regione Campania. Prodotti tradizionali – elenco dei prodotti agroalimentari tradizionali. https://agricoltura.regione.campania.it/tipici/prodotti_tradizionali.htm, 2025.

R. Rei, C. Stewart, A. C. Farinha, and A. Lavie. COMET: A neural framework for MT evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 2685–2702, 2020.

F. Sabatini. Analisi del linguaggio giuridico: Il testo normativo in una tipologia generale dei testi. In M. D'Antonio, editor, *Corso di studi superiori legislativi (1988–1989)*. Cedam, Padova, 1990.

F. Sabatini. "Rigidità-esplicitezza" vs "elasticità-implicitezza": Possibili parametri massimi per una tipologia dei testi. In G. Skytte and F. Sabatini, editors, *Linguistica testuale comparativa*. Museum Tusculanum Press, Copenaghen, 1999.

UNESCO. *Convention for the Safeguarding of the Intangible Cultural Heritage*. UNESCO, Paris, 2003.

A. System Prompt

The following system prompt was shared across all models and conditions (BASELINE, INJ, INJ+ENF):

```
You are translating from Italian into Spanish texts belonging to
the domain of traditional Campanian gastronomy.

##ROLE AND OBJECTIVE##
You act as a professional translator specialized in traditional
gastronomy and cultural heritage (IT->ES). Each term encodes
production processes, historical memory, ritual practices, and
territorial identity, and cannot be reduced to a simple technical
label. Your objective is to produce a complete and faithful
translation that preserves the conceptual integrity of culturally
embedded terms and reads naturally in Spanish.

##TRANSLATION INSTRUCTIONS##
- Translate the ENTIRE text from Italian into Spanish.
- Preserve the tone, register, and voice of the original.
- When terminological instructions are provided, follow them strictly.
- When an expansion or gloss is required, integrate it inline
  within the sentence, adapting its formulation to the co-text.

##REQUIRED OUTPUT##
Return ONLY the final translation in Spanish, complete and ready
for use, with no intercalated comments.

##QUALITY CHECK BEFORE DELIVERING##
- You have translated the entire text with no omissions.
- You have strictly followed the terminological instructions.
- The translation reads naturally in Spanish.
```

B. Glossary Block

The following examples illustrate the genre-specific glossary blocks injected into the user message prior to generation, constructed dynamically from the terminological policies retrieved via the MCP server. No glossary block was injected under the BASELINE condition; the system prompt in Appendix A was the only instruction provided to the model.

Regulatory-administrative

```
##TERMINOLOGICAL INSTRUCTIONS##
The text belongs to the regulatory-administrative discursive genre.
Preserve the following Italian denominations unchanged and format
them with Spanish angle quotation marks. No expansions needed.

- '[IT term]': keep in Italian, write exactly:
  «[IT term]»
```

Promotional-touristic

```
##TERMINOLOGICAL INSTRUCTIONS##
The text belongs to the promotional-touristic discursive genre.
For each term, use the Spanish equivalent immediately followed
by a parenthetical expansion inline: term (expansion).
Reformulate the background note based on the context and co-text.

- '[IT term]': translate as '[ES equivalent] (your expansion here)'
  Background note to reformulate: [term-specific production note
retrieved from the termbase at runtime]
```

Narrative-cultural

```
##TERMINOLOGICAL INSTRUCTIONS##
The text belongs to the narrative-cultural discursive genre.
For each term, use the Spanish equivalent immediately followed
```

```
by an inline cultural gloss introduced by a comma: term, gloss,
Reformulate the background note based on the context and co-text.

- '[IT term]': translate as '[ES equivalent], your gloss here,'
  Background note to reformulate: [term-specific cultural note
retrieved from the termbase at runtime]
```
