Design and methodological architecture of a multilingual corpus of interpreter-mediated public service telephone interactions

Daniel López Padilla, Raquel Lázaro Gutiérrez

Universidad de Alcalá, FITISPos-UAH, Nodo CLARIAH-CM
daniel.lopezp@uah.es , raquel.lazaro@uah.es

María José Vilella Sánchez, Fernando Espinoza Cuadros, Jorge Rico Ródenas

Sigma Cognition
mjvilella@sigmacognition.ai, fespinoza@sigmacognition.ai, jrico@sigmacognition.ai

Abstract

Multimodality in Social Sciences and Humanities (SSH) research is often associated with the integration of text and visual data. However, interpreter-mediated telephone interaction presents a different configuration of complexity, where acoustic, temporal, discursive, and pragmatic dimensions converge. This paper presents the design and methodological architecture of PRAGMACOR(Corpus Pragmatics and Telephone Interpreting: Analysis of Face-Threatening Acts, Ref. PID2021-127196NA-I00), a multilingual corpus of interpreter-mediated public service telephone interactions (Chinese–Spanish, English–Spanish, French–Spanish, German–Spanish), as a case study in multimodal and plurilingual SSH infrastructure.

The corpus integrates aligned audio recordings, orthographic transcriptions enriched with speech phenomena, temporal segmentation into speech acts, and multilayer pragmatic annotation of Face-Threatening Acts (FTAs), validated through a structured double-annotation and expert review process. Beyond textual data, the infrastructure captures prosodic overlap, turn-taking dynamics, and pragmatic mediation, enabling the study of cross-linguistic transfer and relational negotiation in asymmetrical institutional contexts. Datasets such as PRAGMACOR have proved essential to train LLMs for speech to speech translation (Sakai et al., 2024).

Attention is given to the ethical and technical design of the corpus, including local automatic transcription, systematic removal of personal identifiable information, and irreversible voice anonymization through spectral and temporal signal transformation. These procedures ensure both research usability and compliance with responsible data governance principles.

By conceptualising interpreter-mediated interaction as an acoustic-discursive multimodal object and plurilingual pragmatic process, this paper argues that PRAGMACOR provides a replicable model for the development of SSH-oriented infrastructures capable of supporting advanced research in multilingual communication, discourse analysis, and future evaluation of language technologies.

.

Keywords: Multimodal corpus design, Pragmatic annotation, Pragmatics, Multilingual Corpus

1. Introduction

Multimodality has become a central concern in Social Sciences and Humanities (SSH) research, particularly in fields engaging with digital archives, visual culture, and computational text analysis. In many cases, multimodality is conceptualised as the integration of text and image, or more broadly, the combination of heterogeneous data formats within digital infrastructures. However, less attention has been paid to forms of multimodality that emerge in spoken institutional interaction, especially in settings where communication is mediated across languages. Interpreter-mediated telephone encounters constitute a distinctive multimodal environment in which acoustic signals, temporal organisation, discursive structure, and pragmatic negotiation interact in complex ways.

Unlike face-to-face communication, telephone interaction lacks visual cues, yet it remains deeply multimodal. Meaning is co-constructed not only through lexical content but also through prosodic features, overlapping speech, turn-taking dynamics, and sequential organisation. When such interaction is mediated by an interpreter, an additional layer of complexity arises: linguistic transfer is intertwined with pragmatic mediation. Institutional asymmetries, face management, and relational positioning are negotiated across languages in real time, making these interactions particularly rich sites for examining multilingual and cross-cultural communication.

Despite this complexity, many multilingual speech corpora remain primarily text-oriented in their analytical design. Audio recordings are often treated as sources for transcription rather than as integral components of a layered multimodal object. Furthermore, the pragmatic dimension of interaction—such as the management of Face-Threatening Acts (FTAs), mitigation strategies, and relational effects—tends to be underrepresented in corpus infrastructures, even

Proceedings of Shaping Multilingual, Multimodal AI for the Social Sciences and Humanities (LLMs4SSH) @ LREC 2026, pages 153–157
11 May 2026.

though it plays a crucial role in institutional settings.

This paper presents PRAGMACOR, a multilingual corpus of interpreter-mediated public service telephone interactions (English–Spanish, French–Spanish, German–Spanish, and Chinese–Spanish), as a case study in the design of a multimodal and plurilingual research infrastructure. The corpus integrates aligned audio recordings, temporally segmented transcripts, multilayer pragmatic annotation, and rigorous validation procedures, all within a framework that prioritises ethical data governance and long-term reusability. Rather than conceptualising multimodality as the juxtaposition of separate media, PRAGMACOR operationalises it as the structured integration of acoustic, temporal, discursive, and pragmatic layers.

By foregrounding both multimodal architecture and plurilingual mediation, this paper argues that interpreter-mediated telephone interaction requires a reconceptualization of multimodal corpus design in SSH research. In doing so, it proposes an infrastructure model that is methodologically robust, ethically grounded, and adaptable to future developments in multilingual language technologies.

2. Conceptualising Multimodality in Interpreter-Mediated Telephone Interaction

Multimodality in SSH research has frequently been associated with the coexistence of distinct semiotic systems, most prominently text and image, within digital environments (Kress & van Leeuwen, 2006; Jewitt, 2014). These approaches have been particularly influential in media studies and visual communication. However, spoken institutional interaction—especially in telephone settings—calls for a broader understanding of modality. Rather than focusing on visual–textual combinations, interpreter-mediated telephone encounters reveal multimodality as the integration of acoustic, temporal, discursive, and pragmatic layers within a single interactional event.

From an acoustic perspective, prosodic variation, pauses, hesitations, speech rate, and intonation contours contribute to the construction of stance, affect, and relational positioning. Overlapping speech, which is frequent in institutional telephone communication, signals alignment, interruption, urgency, or resistance. These features are not peripheral embellishments to linguistic content; they are constitutive elements of interactional meaning.

At the temporal level, segmentation into speech acts and the organisation of turn-taking shape how institutional authority, accountability, and responsibility are negotiated. Telephone interaction unfolds in tightly structured sequences, where requests, clarifications, confirmations, and directives are embedded within institutional routines. When mediated by an interpreter, each segment is re-articulated across languages, introducing shifts in timing, emphasis, and pragmatic force.

The discursive layer further complicates this configuration. Interpreter-mediated interaction is not a linear transfer of propositional content but a process of multilingual recontextualisation. Speech acts produced in one language are reformulated in another within an asymmetrical institutional framework. This reformulation may preserve, attenuate, or intensify pragmatic force, affecting how Face-Threatening Acts (FTAs) are perceived and negotiated by participants.

In this sense, multimodality in interpreter-mediated telephone interaction is not reducible to multiple media formats; it emerges from the structured interplay of acoustic signal, temporal alignment, discursive segmentation, and pragmatic annotation. Capturing this interplay requires an infrastructure capable of integrating audio, transcription, alignment, and multilayer analytical categories within a single environment.

PRAGMACOR addresses this need by treating audio recordings not merely as sources for textual transcription, but as integral components of a multimodal research object. Through time-aligned segmentation and layered annotation, the corpus preserves the interactional architecture necessary for analysing how multilingual pragmatic mediation unfolds in institutional contexts. This design enables researchers to move beyond text-centric corpus analysis and engage with the dynamic, temporally situated nature of spoken communication.

3. Multilingual and Multilayered Corpus Architecture

PRAGMACOR has been designed as a multilingual and pragmatically annotated corpus of interpreter-mediated public service telephone interactions. The corpus currently includes interactions involving four language pairs—English–Spanish, French–Spanish, German–Spanish, and Chinese–Spanish—reflecting the linguistic diversity of institutional settings where telephone interpreting is deployed. This plurilingual configuration does not merely expand linguistic coverage; it enables the systematic study of cross-linguistic pragmatic mediation in comparable institutional scenarios.

3.1 Multilingual design and validation procedures

The corpus was developed through a structured annotation workflow involving professional transcribers and annotators who are Spanish speakers and native speakers of the respective

foreign language. Each language pair was supported by a dedicated team responsible for transcription, annotation, and review.

To ensure methodological robustness, the annotation process followed a double-annotation protocol: each interaction was independently annotated by two bilingual annotators. In cases of discrepancy, annotations were merged through a semi-automated procedure that displayed divergent labels for adjudication. A third bilingual annotator validated the final decision, and an expert linguist supervised the process—particularly during the initial phases—to reinforce shared criteria and resolve interpretative ambiguities.

This layered validation structure reflects the interpretative nature of pragmatic annotation, where categorisation depends not only on formal linguistic features but also on contextual judgement and theoretical grounding (Ide & Pustejovsky, 2017). The use of double annotation and adjudication procedures aligns with established practices in corpus linguistics aimed at ensuring analytical reliability while acknowledging interpretative variability (Artstein & Poesio, 2008).

3.2 Multilayer annotation and segmentation

All transcription and annotation tasks were carried out using EXMARaLDA, a tool that enables the definition of multiple annotation tiers aligned to the same temporal axis. Each audio file was automatically converted into a structured template with predefined speaker roles and language layers. This configuration allowed annotators to segment speech into temporally bounded units corresponding, as far as possible, to single speech acts.

Segmentation plays a crucial role in preserving the interactional architecture of telephone communication. Annotators were instructed to avoid including more than one speech act per segment and to limit segment duration, thereby facilitating subsequent pragmatic annotation. Because each speaker occupies a separate channel, overlapping speech can be represented explicitly, preserving the simultaneity characteristic of institutional telephone interaction.

Beyond orthographic transcription—enriched with features such as false starts and filled pauses—the corpus incorporates multilayer pragmatic annotation, including the identification and classification of FTAs. These annotations are anchored to specific time-aligned segments, ensuring that pragmatic interpretation remains tied to the acoustic and sequential context in which it occurs.

3.3 Integration of automatic processing and ethical safeguards

To optimise efficiency while maintaining data security, automatic transcription was performed locally using a Whisper model deployed on internal servers, preventing the transfer of sensitive audio data to third-party platforms. Automatically generated transcripts were subsequently revised by human annotators, reinforcing accuracy and contextual interpretation.

Ethical design was central to the corpus architecture. All personal identifiable information (PII) was systematically replaced with dedicated tags in the transcripts. To ensure full anonymisation, a two-stage process was implemented at the audio level. First, time-aligned segments containing personal data were replaced with silence. Second, an irreversible voice anonymisation pipeline was applied, incorporating spectral and temporal signal transformations—including fundamental frequency modification and controlled perturbation—to prevent biometric re-identification.

By integrating automatic processing, multilayer annotation, bilingual validation, and irreversible anonymisation within a unified workflow, PRAGMACOR operationalises a model of multilingual and multimodal corpus design that balances methodological precision, interpretative depth, and responsible data governance.

4. Plurilingual Pragmatic Mediation and Analytical Potential

Interpreter-mediated institutional interaction is not merely multilingual communication; it is a process of plurilingual pragmatic mediation unfolding within asymmetrical institutional frameworks. In such settings, linguistic transfer is inseparable from relational negotiation. Requests, directives, explanations, and justifications are reformulated across languages in real time, often under conditions of urgency or vulnerability. Capturing this complexity requires analytical categories that move beyond lexical equivalence and address the management of face, authority, and accountability.

Face-Threatening Acts (FTAs), understood as acts that potentially challenge a participant's public self-image (Brown & Levinson, 1987), are particularly salient in institutional encounters, where obligations, refusals, procedural constraints, and evaluative statements are frequent. The annotation of FTAs within PRAGMACOR provides a structured entry point into this dimension. FTAs, understood as acts that potentially challenge a participant's public self-image, are particularly salient in institutional encounters, where obligations, refusals, procedural constraints, and evaluative statements are frequent. In interpreter-mediated telephone interaction, such acts are not simply reproduced

across languages; they are re-articulated. The interpreter's rendition may preserve, mitigate, intensify, or subtly reframe the pragmatic force of the original utterance.

This plurilingual re-articulation has several analytical implications. First, it enables the study of cross-linguistic pragmatic transfer: how degrees of imposition, politeness strategies, or mitigation devices shift when speech acts move between linguistic systems. Second, it allows for the examination of relational positioning within institutional asymmetries. Public service encounters often involve unequal access to knowledge, authority, or resources; the interpreter's mediation can influence how these asymmetries are enacted discursively.

Because PRAGMACOR integrates time-aligned audio, segmented speech acts, and multilayer pragmatic annotation, researchers can examine these processes at different levels of granularity. For instance, it becomes possible to trace how a directive produced in one language is segmented, reformulated, and sequentially embedded in the target language. Overlaps, hesitations, or prosodic cues can be analysed alongside pragmatic labels, preserving the interactional context in which mediation occurs.

The plurilingual design of the corpus further expands its analytical scope. By including multiple language pairs involving Spanish as a common institutional language, PRAGMACOR facilitates comparative research across linguistic and cultural configurations. Patterns of mitigation, directness, or face management can be examined not only within individual interactions but also across language pairs, contributing to broader discussions on multilingual communication and cultural diversity in institutional contexts.

Importantly, this analytical potential extends beyond traditional discourse analysis. The structured, validated annotation framework positions the corpus as a foundation for future interdisciplinary research, including the evaluation of computational tools aimed at modelling multilingual pragmatic phenomena. While the present paper focuses on infrastructural design, the layered architecture of PRAGMACOR makes it adaptable to emerging methodologies in multilingual language technologies, without reducing its interpretative depth.

By conceptualising interpreter-mediated communication as plurilingual pragmatic mediation, PRAGMACOR demonstrates how multimodal corpus design can capture the relational and interactional dimensions of multilingual SSH research. Rather than treating language transfer as a neutral conduit of information, the corpus foregrounds the dynamic reconfiguration of meaning across languages and institutional roles.

5. Ethical, Infrastructural and Methodological Implications for SSH Research

The design of multilingual and multimodal corpora in institutional contexts entails specific methodological and ethical challenges. Telephone interpreting in public service settings frequently involves sensitive personal information, vulnerable populations, and asymmetric power relations. As a result, corpus development cannot be approached solely as a technical task; it must be embedded within a framework of responsible data governance and long-term research sustainability.

PRAGMACOR addresses these challenges through a layered anonymisation strategy that operates at both textual and acoustic levels. Personal identifiable information is systematically replaced with dedicated tags in transcripts, while audio segments containing such information are removed through time-aligned silencing. Crucially, voice anonymisation is implemented through irreversible spectral and temporal transformations of the signal, preventing biometric re-identification while preserving interactional properties relevant to research. This dual approach allows the corpus to remain analytically usable without compromising participant privacy.

From an infrastructural perspective, the integration of local automatic transcription tools with human validation procedures reflects a balanced model of technological support and expert interpretation. Automatic speech recognition accelerates processing while remaining embedded within a workflow that prioritises contextual judgement and theoretical grounding. Similarly, multilayer annotation in a time-aligned environment ensures that pragmatic categories remain anchored to the acoustic and sequential realities of interaction.

Methodologically, the corpus illustrates the importance of treating annotation not as a neutral labelling exercise but as an interpretative practice requiring training, calibration, and adjudication. The double-annotation and review system implemented in PRAGMACOR underscores the complexity of pragmatic classification, particularly when applied across languages and institutional contexts. This design foregrounds transparency and reproducibility by making the decision-making process traceable and structured.

Beyond its immediate analytical applications, the infrastructure proposed here has broader implications for SSH research. First, it offers a replicable model for integrating acoustic, temporal, discursive, and pragmatic layers within multilingual corpora. Second, it demonstrates how ethical safeguards can be embedded directly into the architecture of data processing rather than treated as external constraints. Third, it positions

interpreter-mediated interaction as a valuable site for studying multilingual communication in contexts where relational and institutional dynamics are central.

As SSH research increasingly engages with multilingual data and computational tools, the need for robust, ethically grounded, and multimodally structured infrastructures becomes more pressing. By conceptualising telephone interpreting as a multimodal and plurilingual object of study, PRAGMACOR contributes to this emerging landscape, providing a framework that supports both fine-grained discourse analysis and future interdisciplinary collaboration.

6. Bibliographical References

Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics, 34*(4), 555–596. https://doi.org/10.1162/coli.07-034-R2

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

Ide, N., & Pustejovsky, J. (2017). *Handbook of linguistic annotation*. Springer. https://doi.org/10.1007/978-94-024-0881-2

Jewitt, C. (Ed.). (2014). *The Routledge handbook of multimodal analysis* (2nd ed.). Routledge.

Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.

Sakai, Y., Makinae, M., Kamigaito, H., Watanabe, T. (2024). Simultaneous Interpretation Corpus Construction by Large Language Models in Distant Language Pair. arXiv:2404.12299. https://doi.org/10.48550/arXiv.2404.12299