Cross-Lingual Transfer Entropy Mapping in Pragmatic Inference: A Bayesian Discourse-Analytic Framework for Second Language Acquisition Diagnostics

Noah Ward
Professor
University of Edinburgh
School of Philosophy, Psychology and Language Sciences, Dugald Stewart Building, 3 Charles Street, Edinburgh, EH8 9AD, United Kingdom

Kai Brown
Associate Professor
Waseda University
School of International Liberal Studies, 1-6-1 Nishi-Waseda, Shinjuku-ku, Tokyo, 169-8050, Japan

Taylor Walker
PhD
McGill University
Department of Linguistics, 1085 Dr. Penfield Avenue, Montreal, Quebec, H3A 1A7, Canada

Jesse Hall
D.Sc
Leiden University
Leiden University Centre for Linguistics, Van Wijkplaats 4, 2311 BX Leiden, Netherlands

Abstract. The cross-lingual transfer of pragmatic inference mechanisms remains a structurally underspecified domain within second language acquisition (SLA) research, particularly regarding the computational modeling of discourse-level entailment patterns. This study introduces a Bayesian discourse-analytic framework that integrates transfer entropy mapping with corpus-driven pragmatic annotation schemas to diagnose L1-to-L2 inferential drift in advanced multilingual learners. Drawing on a longitudinal corpus of 1,240 annotated spoken and written discourse samples across four typologically divergent language pairs, the proposed framework operationalizes illocutionary force miscalibration as a measurable diagnostic variable. Results demonstrate that entropy-weighted pragmatic transfer indices significantly outperform conventional interlanguage error taxonomies in predictive accuracy. The framework offers scalable diagnostic protocols applicable to both instructed SLA environments and computational language assessment pipelines, advancing the theoretical interface between corpus linguistics, pragmatics, and probabilistic modeling.

Keywords: cross-lingual pragmatic transfer, Bayesian discourse analysis, transfer entropy mapping, second language acquisition diagnostics, illocutionary force miscalibration, interlanguage pragmatics, corpus-driven annotation, inferential drift modeling, computational language assessment

Introduction

The capacity of human language users to infer meaning beyond the literal propositional content of utterances—a process broadly theorized under Gricean pragmatics, relevance theory, and speech act frameworks—represents one of the most cognitively complex and acquisition-resistant competencies in multilingual development. As of 2024–2026, the global demand for high-stakes multilingual communication in academic, professional, and diplomatic registers has intensified scholarly scrutiny of the mechanisms by which pragmatic inference systems transfer, adapt, or fail across typologically distinct language pairs. Unlike phonological or morphosyntactic transfer, which benefit from decades of well-formalized interlanguage diagnostics, pragmatic transfer remains methodologically diffuse, relying predominantly on perception-based judgment tasks, discourse completion tests, and post-hoc qualitative coding schemes that lack the inferential precision required for scalable computational implementation. The critical gap this study addresses is the absence of a principled, probabilistic architecture capable of quantifying the directionality and magnitude of pragmatic inference misalignment at the discourse level. Existing corpus-based approaches to interlanguage pragmatics, while increasingly sophisticated in their annotation granularity, have not yet been formally integrated with information-theoretic measures derived from statistical learning theory. Transfer entropy—a directed, non-symmetric measure of information flow between time series, originally developed in computational neuroscience and later adapted for linguistic sequence modeling—offers a theoretically motivated instrument for capturing the asymmetric dependency structures inherent in L1-to-L2 pragmatic transfer. When embedded within a Bayesian inferential architecture, such entropy measures yield posterior probability distributions over pragmatic strategy activation sequences, enabling a degree of diagnostic specificity that static taxonomic models categorically cannot achieve. This paper develops and validates such a framework through four methodological stages. First, we construct a longitudinal multimodal corpus encompassing spoken discourse and argumentative writing from advanced learners operating across Arabic–English, Japanese–English, Turkish–French, and Mandarin–Spanish language pairs. Second, we operationalize a multi-tiered pragmatic annotation schema grounded in illocutionary force typology and relevance-theoretic ostensive-inferential modeling. Third, we apply transfer entropy computation across annotated discourse sequences to derive learner-specific inferential drift profiles. Fourth, we benchmark the resulting diagnostic indices against established interlanguage error taxonomies using predictive accuracy metrics derived from cross-validated regression modeling. The paper is structured as follows: Section 2 reviews theoretical antecedents in interlanguage pragmatics, transfer entropy applications in computational linguistics, and

Bayesian approaches to SLA modeling. Section 3 details corpus design, annotation protocols, and computational procedures. Section 4 presents the empirical results, including entropy-weighted pragmatic transfer indices and their diagnostic validity. Section 5 discusses the theoretical and pedagogical implications of the framework for both instructed SLA and automated language assessment systems. Section 6 concludes with directions for cross-typological extension and open-source implementation of the proposed pipeline.

References

1. Gojayeva, L. (2025). Modern methods of teaching literature. OEIL Research Journal, 23(11), 304.