

Legal NLP for Comparing Judicial Considerations in Child Custody Cases Involving Domestic and Foreign Spouses

Sieh-Chuen Huang¹ Hsuan-Lei Shao^{2*}

¹College of Law, National Taiwan University

²Graduate Institute of Health and Biotechnology Law, Taipei Medical University

schhuang@ntu.edu.tw hlshao@tmu.edu.tw

Abstract

Legal decision-making in child custody cases is primarily articulated through judicial language. This study applies legal natural language processing (Legal NLP) to analyze judicial considerations in child custody cases involving domestic and foreign spouses in Taiwan. We design a comparative analytical task based on a curated corpus of divorce-related custody judgments to examine differences in emphasized legal considerations. Interpretable NLP methods are employed to identify linguistic patterns associated with judicial practice. The results indicate systematic differences in how courts frame and prioritize legal considerations across case types.

1 Introduction and Related Works

1.1 Research Background

The number of foreign spouses in Taiwan has increased steadily in recent decades, accompanied by a rise in cross-national marriages and subsequent divorce cases. As a result, legal disputes concerning child custody have become more frequent, particularly in cases involving domestic and foreign spouses. These disputes require courts to articulate complex legal considerations through written judicial reasoning.

Under Taiwan's legal framework, divorce may occur through consensual agreement or judicial adjudication, with mandatory mediation required when parties fail to reach agreement on custody arrangements [1]. Although mediation is culturally preferred, unresolved disputes ultimately require judicial determination, making court judgments a primary source for examining how legal

considerations are framed in practice. The growing presence of foreign spouses has also generated broader demographic and policy debates related to population structure, fertility, and social integration [2]. While policies addressing the rights of foreign spouses are often framed as part of Taiwan's commitment to equality, disagreements remain regarding whether legal practice achieves substantively equal treatment. These debates motivate the need for empirical analysis grounded in judicial texts rather than normative claims alone.

1.2 Legal NLP and Judicial Analysis

Recent advances in legal natural language processing (Legal NLP) have enabled systematic analysis of judicial texts, supporting tasks such as information extraction, document summarization, and legal reasoning analysis. Techniques including named entity recognition and context-aware extraction have been applied to identify key legal actors, concepts, and relationships within court decisions [3]. In multilingual and cross-domain legal settings, NLP methods have also been used to address linguistic variation and interpretability challenges [4].

Beyond predictive modeling, prior work has demonstrated that judicial decisions can be treated as structured linguistic artifacts suitable for empirical analysis using NLP techniques [5]. Building on this perspective, recent Legal NLP research increasingly emphasizes interpretable task design and domain-sensitive analysis. Following this line of work, the present study applies Legal NLP methods to analyze judicial considerations in child custody cases involving domestic and foreign spouses in Taiwan. By designing a comparative analytical task and focusing on interpretable

* corresponding author

linguistic patterns, this study contributes to methodological discussions in legal domain language processing and provides empirical insights into judicial reasoning in family law contexts.

2 Data and Task Design

2.1 Dataset Construction

The dataset used in this study consists of 205 divorce-related child custody judgments issued by Taiwanese courts. Judicial decisions were collected from publicly available court databases and screened to include cases in which child custody arrangements were explicitly addressed in the context of divorce. To support comparative analysis, cases were categorized based on the nationality status of the spouses, distinguishing between disputes involving two domestic spouses and those involving at least one foreign spouse.

All judgments were anonymized to remove personal identifiers, and non-textual metadata were excluded from analysis. The corpus was further curated to ensure consistency in document structure and legal context, focusing on the judicial reasoning sections relevant to custody determination. This curated dataset provides a consistent textual basis for examining differences in how judicial considerations are articulated across case categories.

2.2 Task Design

The research task is formulated as a comparative analytical task aimed at examining differences in judicial considerations articulated in child custody decisions involving domestic and foreign spouses. Rather than predicting judicial outcomes, the task focuses on how legal considerations are expressed in judicial reasoning, reflecting the interpretative nature of family law adjudication.

Specifically, the task involves identifying and comparing linguistic patterns associated with commonly referenced legal considerations in custody cases, including child welfare, parental caregiving capacity, family stability, and socio-economic circumstances. These considerations are operationalized through textual features extracted from the reasoning sections of judicial decisions.

Interpretable Legal NLP techniques are applied to extract salient linguistic features and to support systematic comparison across the two case categories. By framing the analysis as a comparative task grounded in judicial language, the study enables examination of whether and how courts emphasize different legal considerations depending on the nationality composition of the spouses. This task design prioritizes interpretability and domain sensitivity, aligning with recent methodological directions in legal domain language processing.

3 Methods

3.1 Preprocessing and Segmentation

Judicial texts analyzed in this study are drawn from Taiwanese court judgments concerning child custody determinations following divorce. Given the highly structured nature of judicial decisions, which typically consist of factual summaries, reasoning sections, and conclusions, the study begins with text preprocessing and structural segmentation to ensure consistency in legal context. During preprocessing, non-semantic elements such as formatting markers, page numbers, and section headers are removed. Chinese word segmentation is then applied to facilitate subsequent linguistic analysis. To focus specifically on judicial reasoning, only the reasoning sections of the judgments are retained for analysis, excluding factual descriptions and dispositive outcomes.

After preprocessing, each judgment is represented by its reasoning section as the basic unit of analysis. This design allows the study to concentrate on the language through which courts articulate legal considerations and ensures comparability across cases.

3.2 Linguistic Feature Extraction with TF-IDF

To analyze how judicial considerations are expressed in custody adjudication, this study adopts lexical-level linguistic features as the primary analytical representation. Word usage in judicial reasoning is treated as an observable indicator of how courts emphasize different legal considerations across cases.

Specifically, the study employs the term frequency-inverse document frequency (TF-IDF) representation to

transform judicial reasoning texts into vectorized features. TF-IDF captures the relative importance of terms within individual judgments while reducing the influence of terms that are common across the corpus. This property makes TF-IDF suitable for identifying legal language that is comparatively emphasized across different types of custody cases.

Using TF-IDF representations, the analysis focuses on salient terms and noun phrases associated with custody determinations, including expressions related to child welfare, caregiving arrangements, parental responsibilities, residential stability, and procedural participation. This feature design prioritizes interpretability and provides a transparent basis for comparing judicial considerations across cases involving domestic and foreign spouses.

3.3 Interpretable NLP and Legal Domain

Given the contextual and normative nature of judicial reasoning in family law, this study adopts an interpretable legal NLP approach that emphasizes transparency and domain sensitivity. Rather than aiming to predict judicial outcomes, the analysis focuses on examining how legal considerations are articulated through language in written judgments.

Accordingly, feature-based and traceable NLP techniques are employed to allow direct examination of linguistic patterns and their association with substantive legal considerations. This approach enables analytical findings to be interpreted in relation to concrete legal concepts and adjudicative practices, which is particularly important in the analysis of child custody cases where judicial discretion is exercised through detailed reasoning.

By prioritizing interpretability over predictive performance, the proposed method supports systematic comparison of judicial language while preserving the legal context in which custody determinations are made.

4 Results

4.1 Lexical Differences in Judicial Considerations

The TF-IDF-based analysis reveals systematic differences in how specific judicial considerations are emphasized in child custody cases involving spouses from domestic and foreign backgrounds. Rather than indicating

divergent legal standards, these differences reflect variation in judicial focus at the lexical level within a shared adjudicative framework.

Tables reporting mean TF-IDF values show that terms associated with exit and appearance in court are more prominently emphasized in cases involving foreign spouses than in cases involving two domestic spouses. These lexical patterns indicate that courts more frequently address issues related to physical presence and procedural participation when adjudicating custody disputes involving foreign spouses.

Such differences suggest that judicial reasoning adapts to case-specific circumstances by foregrounding considerations that may affect the feasibility and execution of custody arrangements. Importantly, these patterns do not imply preferential or prejudicial treatment, but rather highlight how courts operationalize legal considerations through language in response to contextual factors.

4.2 Procedural Participation and Legal Vulnerability of Foreign Spouses

A closer examination of terms related to *exit* and *court appearance* provides insight into the procedural position of foreign spouses in custody litigation. As shown in the comparison of mean TF-IDF values, references to exit and appearance are substantially more salient in judgments involving foreign spouses than in those involving only domestic spouses.

Table 1 Comparison of Mean TF-IDF for “Exit” and “Court Appearance” Across Case Types

Mean TF-IDF Case Type	Exit	Court Appearance
Foreign Spouses	0.0371	0.0339
Domestic Spouses	0.0006	0.0208

Note: Higher TF-IDF values indicate greater lexical emphasis on the corresponding terms in judicial reasoning.

This emphasis reflects structural features of Taiwan’s legal and immigration systems. Following divorce, foreign spouses may lose their residency status, either voluntarily due to economic constraints or involuntarily as a result of legal requirements. As a consequence, foreign spouses are more likely to be absent from custody proceedings, leading to a higher incidence of default or

one-sided judgments.

While this pattern does not indicate judicial bias in decision-making, it does reveal a procedural vulnerability that may disadvantage foreign spouses in asserting custody claims. From a legal perspective, the observed linguistic emphasis underscores how procedural participation becomes a salient judicial concern when residency status and physical presence are uncertain.

4.3 Best Interests of the Child and Social Support for Caregiving

Another set of lexical differences identified through NLP analysis concerns terms related to housing and cohabitation. Comparisons of mean TF-IDF values show that references to renting and cohabitation are more prominent in custody judgments involving foreign spouses than in those involving two domestic spouses.

Table 2 Comparison of Mean TF-IDF Values for “Renting” and “Cohabitation” Across Case Types

Mean TF-IDF Case Type	Renting	Cohabitation
Foreign Spouses	0.0155	0.0515
Domestic Spouses	0.0050	0.0341

Note: TF-IDF values reflect the relative importance of terms related to housing arrangements and social support in judicial reasoning.

These patterns reflect differences in access to social and familial support following divorce. Domestic spouses are more likely to return to their family of origin or rely on established social networks, whereas foreign spouses may lack comparable support structures in Taiwan. Consequently, courts place greater emphasis on whether foreign spouses have stable housing arrangements, sufficient economic resources to rent accommodation, or access to mutual support networks through cohabitation.

From the perspective of judicial reasoning, such considerations are closely linked to the best interests of the child. By scrutinizing housing stability and caregiving support, courts seek to ensure that custody arrangements provide a stable living environment for minors. Although these inquiries may appear differential at the lexical level, they function to balance parental equality with child welfare considerations, effectively protecting the interests of both parents and the child as a third and particularly

vulnerable party.

5 Conclusion

This study applies interpretable legal natural language processing to examine how judicial considerations are articulated in child custody cases involving domestic and foreign spouses in Taiwan. By focusing on lexical-level differences in judicial reasoning rather than predictive outcomes, the analysis reveals how courts adjust their emphasis on legal considerations in response to case-specific contexts.

The findings indicate that while courts employ a shared legal framework across custody cases, differences emerge in the emphasis placed on procedural participation and caregiving support. These differences reflect structural and social conditions—such as residency stability and access to social support—rather than systematic inequality in judicial standards. In this sense, judicial language functions as a mechanism for balancing parental interests while safeguarding the best interests of the child.

From a methodological perspective, this study demonstrates the value of interpretable legal NLP as a bridge between legal doctrine and practical adjudication. By making judicial reasoning patterns observable, NLP-based analysis offers insights relevant not only to legal scholarship but also to social policy design in societies experiencing increasing cross-border family formation. As issues related to migration and family law continue to evolve, interpretable legal NLP provides a scalable and transparent approach for examining how legal systems respond to emerging social realities.

References

- [1] S. H. Hwang. Family policies in Taiwan: Development, implementation, and assessment. In *Handbook of Family Policies Across the Globe*, 2014.
- [2] W. S. Yang and Y.-Y. T. Liu. Gender imbalances and the twisted marriage market in Taiwan. *Springer Series on Demographic Methods and Population Analysis*, 2014.
- [3] D. N. Shyamala and H. J. Grace. Sentiment-based summarization of legal documents using natural language processing techniques. In *Effective*

Strategies for Combatting Social Engineering in Cybersecurity, 2024.

- [4] A. Radhika, N. K. Bhasin, R. Sabareesh, and I. I. Raj. Optimization of natural language processing models for multilingual legal document analysis. In Proceedings of the 2024 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS 2024), 2024.
- [5] Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel PreoŃiuc-Pietro, and Vasileios Lampos. Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. PeerJ Computer Science, 2016.