

Article

Not peer-reviewed version

---

# Terminological Variation in Baseball Chinese Across the Taiwan Strait: A Document-Based Study of Standardization and Specialist Communication

---

[Bingcheng Chen](#) and [Yuksai Nam](#) \*

Posted Date: 19 May 2026

doi: 10.20944/preprints202605.1220.v1

Keywords: cross-strait variation; baseball terminology; specialist language; terminology variation; terminology standardization; Chinese sports terminology



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Terminological Variation in Baseball Chinese Across the Taiwan Strait: A Document-Based Study of Standardization and Specialist Communication

Bingcheng Chen and Yuksai Nam \*

School of Physical Education, Xiamen Institute of Technology

\* Correspondence: lanyuqian@xit.edu.cn

## Abstract

This article examines cross-Strait variation in Chinese baseball terminology through a document-based comparison of two primary sources: the terminology appendix contained in the Chinese Taipei Baseball Association's baseball rules and the China Baseball Association's Basic Terminology of Baseball standard. Based on Supplementary Dataset S1, a cleaned 363-entry English-Chinese comparison dataset, the study investigates how baseball terms differ across the Strait in documentary coverage, lexical designation, expression style, and communicative relevance. The analysis identifies 214 directly comparable entries with renderings on both sides. Of these, 101 are classified as convergent or near-convergent, while 113 show lexical divergence. A further 149 entries do not enter the directly comparable subset. The findings show that cross-Strait baseball terminology is shaped by more than isolated word-level difference. Taiwan-side terms often preserve compact and conventionalized forms used in baseball practice, whereas Mainland standardized forms frequently display a more explicit and institutionally codified style. The article argues that such variation should not be treated simply as inconsistency, but as specialist-language variation shaped by different historical, institutional, and communicative conditions. On this basis, the article suggests a graded, communication-oriented approach that tolerates low-sensitivity variants, cross-references moderate-sensitivity terms, and coordinates high-sensitivity rule terms for umpiring, commentary, translation, and instruction.

**Keywords:** cross-strait variation; baseball terminology; specialist language; terminology variation; terminology standardization; Chinese sports terminology

---

## 1. Introduction

### 1.1. Baseball Across the Taiwan Strait: Sport, Exchange, and Terminology

Baseball is a highly codified sport in which specialist vocabulary plays an important role in rules, umpiring, coaching, commentary, statistics, and player development. Across the Taiwan Strait, baseball is not only a historical or cultural point of comparison; it is also a field of practical sporting, educational, and communicative exchange. As cross-Strait baseball exchanges, youth events, and related collaborations continue to expand, terminology becomes more than a linguistic detail. It is one of the basic means through which coaches, players, umpires, teachers, and organizers communicate, and it is also one of the points at which misunderstanding and uneven interpretation may arise.

These difficulties are not merely incidental matters of wording. Like other sport-specific lexical systems, baseball terminology does not circulate in a uniform form once it moves across languages, institutions, and speech communities. Research on sports lexis has shown that internationally diffused sports often generate borrowing, loan translation, local adaptation, and competing naming conventions, especially when English-origin terms enter new linguistic and institutional

environments (Bergh & Ohlander, 2017; Milić, 2013). In this sense, baseball terminology can be studied not only as a set of labels, but also as part of a broader system of specialist communication shaped by sport history, usage tradition, and standardization practice.

### *1.2. Sports Terminology and Cross-Strait Lexical Variation*

Sport is also a productive site for language research because it combines formal rule-based terminology with highly conventionalized community usage. Ferguson's (1983) classic study of sports announcer talk showed that sports discourse develops distinctive register features of its own. More recent work on sports vocabulary further shows that sport-specific terms are shaped by the communicative settings in which they circulate, including reporting, commentary, translation, and local discourse communities (Bergh & Ohlander, 2017; Hiel & Zenner, 2023; Milić, 2013). Sport history and global-history scholarship likewise suggest that modern sports should be understood as transnationally circulating systems whose meanings, institutions, and vocabularies are repeatedly localized in new settings (Taylor, 2013). Baseball is especially revealing in this respect because it combines a dense rule system with a long history of international transmission and local adaptation.

Among the studies closer to the present topic, Tao examined cross-Strait sports terminology, paying particular attention to baseball and basketball as well as to issues of word formation and translation (Tao, 2013). Although that study is not built around a full documentary dataset of baseball terms, it helps establish an important point: differences in sports terminology across the Strait are patterned rather than accidental, and they are tied to different linguistic habits and histories of (Tao, 2013). More broadly, studies of sports language and sports lexis show that sporting domains provide useful sites for observing how borrowing, local alternatives, and community usage become stabilized in specialist discourse (Bergh & Ohlander, 2017; Ferguson, 1983; Hiel & Zenner, 2023).

A second strand of scholarship comes from research on cross-Strait lexical and terminological difference more generally. This body of work shows that divergence across the Strait is not limited to baseball, but forms part of a broader pattern in which historically related language forms develop different naming habits and preferred lexical structures over time. Studies of Mainland and Taiwan Mandarin have examined vocabulary differences, near-synonym variation, conventional expressions, and intralingual negotiation across the Strait, showing how differences may emerge in form, meaning, usage, and communicative function (Liu & Liles, 2025; Liu & Chen, 2022; Song, 2025; Wang & Huang, 2018; Zhou & Zhou, 2019). That broader background matters here because it situates baseball terminology within a larger pattern of differentiated terminology formation.

### *1.3. Baseball History, Taiwan Baseball Development, and Terminology Background*

What makes baseball terminology especially interesting, however, is that it is also tied to the history of the sport itself. In the case of Taiwan, that history cannot be separated from the island's long exposure to Japanese baseball culture. Baseball was introduced and developed in Taiwan under Japanese colonial rule, and this early institutional setting shaped the sport's local development in ways that differed from its later development on the Mainland. Lin and Lee (2007) show that baseball was promoted under Japanese colonialism not merely as a leisure activity, but also as a political and educational medium through which the colonial government sought social integration. This means that baseball in Taiwan entered schools, public culture, and organized sporting life through a distinctive colonial route rather than through later international sport exchange alone.

The social meaning of baseball in Taiwan also changed over time. Wang (2009) argues that Taiwanese baseball became entangled with Japanese colonization, Chinese Nationalist rule, ethnic stratification, and the emergence of Taiwanese identity. In this sense, baseball on the island became more than an imported sport; it became a site where history, identity, and social memory were repeatedly negotiated. Yu's historical account similarly treats Taiwanese baseball as a sport whose development must be understood in relation to wider social, cultural, and political settings (Yu, 2007). Taken together, these studies suggest that baseball in Taiwan cannot be understood simply as a later

local adoption of an international sport. Instead, it developed through a distinctive historical route, and that route left its mark on baseball culture more broadly.

Viewed in this historical context, differences in baseball terminology across the Strait can also be read as part of the different historical pathways through which the sport took root and acquired local meaning. Japanese influence forms part of the historical pathway through which baseball terminology developed in Taiwan and helps illuminate some of the naming differences visible across the Strait.

#### 1.4. Theoretical Orientation

Building on these strands of scholarship, the sociocognitive approach to terminology offers a particularly suitable theoretical orientation for the present study. Temmerman's work does not begin from the assumption that specialist terminology should consist of one concept neatly matched to one correct designation. Instead, it treats terms as historically shaped, socially embedded, and cognitively organized. From this perspective, variation is not simply a form of noise or failure; it may reflect the different ways in which communities stabilize, transmit, and work with specialist knowledge over (Temmerman, 2000). For a study of cross-Strait baseball terminology, this shift in perspective is especially helpful. It allows lexical difference to be read not merely as inconsistency, but as evidence of partially different traditions of naming, usage, and institutionalization.

At the same time, Cabré's communicative perspective provides a valuable supporting framework by emphasizing that terms are not only conceptual units, but also linguistic and communicative ones. If the sociocognitive approach helps explain why variation exists, the communicative perspective helps explain why some forms of variation carry greater practical weight (Cabré, 2003; 1999). Freixa's work on denominative variation also supports the present study because it shows that multiple designations for the same concept are common in specialist language and may arise from dialectal, functional, discursive, interlinguistic, and cognitive causes (Freixa, 2006).

Recent terminology studies have further shown the value of using empirically grounded comparison to examine specialized language. Cognitive and knowledge-organization approaches to terminology emphasize that specialized terms are connected to conceptual structure, domain knowledge, and communicative use (Faber, 2012; Hjørland, 2023). For example, Haddad Haddad and Montero-Martínez (2019) use contrastive terminological analysis to examine English-Arabic climate-change terminology and show how specialist terms may reveal conceptual and translational patterns. Reimerink and co-researchers (2024) similarly demonstrate that specialized domains and subdomains may differ in how related concepts are linguistically expressed, and that such differences are relevant to terminology description. Work on glossary construction also highlights the importance of explicit terminological organization and transparent conceptual information in specialist resources (Maculan et al., 2023). Although these studies examine domains other than sport, they support the present paper's broader assumption that specialist terms should be studied in relation to conceptual organization, linguistic form, and communicative context.

From an institutional perspective, standardization helps frame the role of the Mainland source in this paper. Because one of the two primary sources is a formal terminology standard, research on standardization is relevant to how that document is interpreted. Such research has repeatedly shown that standards are selective rather than exhaustive: they codify preferred forms for particular purposes rather than simply recording every expression circulating in practice (Ayres-Bennett & Bellamy, 2021; Yin, 2021). Research on language norms has also emphasized that explicit codification interacts with geographical, historical, functional, stylistic, and social forms of variation (Amorós-Negre & Costa-Carreras, 2022). The China Baseball Association standard should therefore not be treated as a complete lexical map of all baseball usage in Mainland China. In this study, standardization is understood as one institutional response to variation rather than as the disappearance of variation itself.

#### 1.5. The Present Study

What is still lacking is a document-based term-by-term comparison of how a specialist sports vocabulary diverges across the Taiwan Strait, where shared written Chinese coexists with different translation pathways, standardization practices, and usage traditions. The present study addresses this gap through a document-based comparison of two primary sources: the terminology appendix contained in the Chinese Taipei Baseball Association rulebook and the China Baseball Association's official terminology standard, both of which are described in detail in Section 2.1. The first is used as the main comparison list because it provides the broadest English-term inventory available within the present study and therefore defines the scope of comparison. The second serves as the Mainland standard reference. Rather than functioning as a record of every baseball term in circulation, it presents a screened and codified set of preferred forms. This source design makes it possible to move beyond anecdotal contrast and examine cross-Strait terminology difference in a more systematic way, while also recognizing that the two documents occupy different positions within the documentary record.

Drawing on this literature, the present article investigates cross-Strait baseball terminology as a problem of variation, specialist language, and institutional standardization. Based on Supplementary Dataset S1, it focuses on how terminology diverges, how these differences can be described analytically, and why they matter for baseball-related communication in Chinese. More specifically, the article addresses three questions:

1. How do baseball terms diverge across the Taiwan Strait in terms of coverage, lexical choice, and expression style?
2. What historical and linguistic factors help explain these patterns of divergence?
3. What implications do such differences have for terminology planning and practical communication in baseball contexts?

By answering these questions, the study aims to contribute to three related fields: the analysis of specialist vocabulary in sport, the study of terminology variation in Chinese, and the practical discussion of how baseball terminology can be organized without losing sight of actual usage.

## 2. Materials and Methods

This study uses qualitative document analysis together with structured term-by-term comparison to examine cross-Strait variation in baseball terminology. The analysis is based on documentary materials in which baseball terms have already been selected, organized, and codified. This makes document analysis especially suitable, since it allows recurrent patterns of terminology use and differentiation to be examined within the textual and institutional settings in which they appear (Bowen, 2009). Similar contrastive terminology studies have also shown that document- or dataset-based comparison can be used to examine how specialist terms are shaped by conceptual organization, translation practice, and communicative context (Haddad Haddad & Montero-Martínez, 2019; Reimerink et al., 2024).

The comparison is informed by a sociocognitive view of terminology, which treats terms as historically shaped and socially situated rather than as fixed labels in a fully closed system. In this context, the purpose of the study is not to determine a single correct naming system, but to examine how baseball concepts are expressed, stabilized, and differentiated across the Strait through two related yet non-identical documentary traditions (Temmerman, 2000).

### 2.1. Primary Sources

The study draws on two primary documentary sources.

The first source is the Chinese Taipei Baseball Association rulebook appendix, specifically Appendix 3, "Proposed Standardized Chinese Translations of Baseball Terminology." Hereafter, this source is referred to as the CTBA appendix. In this study, it serves as the main source for constructing Supplementary Dataset S1 because it provides the broadest set of English baseball entries and

Taiwan-side Chinese renderings used for cross-Strait comparison (Chinese Taipei Baseball Association, 2024).

The second source is the China Baseball Association's Basic Terminology of Baseball (T/CBAA 0001–2020). Hereafter, this source is referred to as the CBA standard. This document serves as the Mainland standard reference. As stated in its introduction, the standard was developed through the collection, screening, and redefinition of commonly used baseball terms. It is therefore treated here as a screened foundational terminology standard rather than as a complete record of all baseball expressions in circulation (Chinese Baseball Association, 2020).

**Table 1.** Primary sources and their roles in the study.

Source	Full title	Year	Document type	Role in this study
Source 1	Chinese Taipei Baseball Association rulebook appendix, Appendix 3, "Proposed Standardized Chinese Translations of Baseball Terminology"	2024	Appendix-based terminology list	Main comparison list
Source 2	Basic Terminology of Baseball (T/CBAA 0001–2020)	2020	Official terminology standard	Mainland standard reference

## 2.2. Dataset Construction and Comparison Procedure

The comparison was built in three stages.

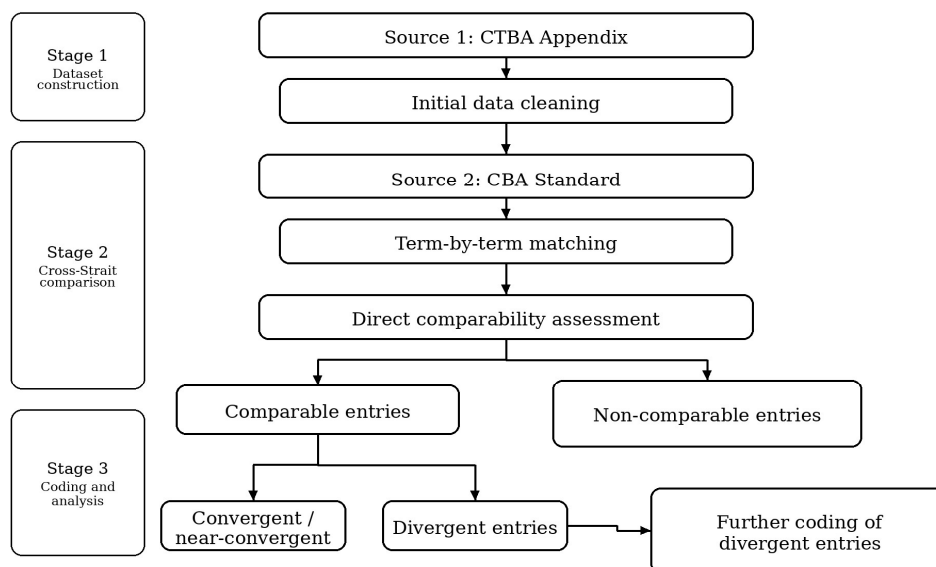
First, all baseball entries from Source 1 were extracted and recorded in Supplementary Dataset S1 as the basis of analysis. These entries are organized by English headwords together with Taiwan-side Chinese renderings where available. Because Source 1 provides the broadest entry list available within the present research design, it establishes the scope of the dataset. On this basis, a cleaned dataset of 363 baseball entries was compiled. Two entries contain English terms only and do not have Chinese renderings on either side. They were retained in Supplementary Dataset S1 for numbering transparency but were excluded from direct lexical comparison.

Second, each entry in that list was checked against Source 2 to determine whether a corresponding standardized Mainland Chinese term could be identified. Where such a form was available, it was recorded as the Mainland counterpart. Where no directly corresponding item could be located in the selected standard, the entry was treated as falling outside the standard's direct coverage within the present dataset. This distinction is important because the study does not equate non-matching with non-existence; it indicates only that no corresponding entry was identified in the selected standard.

Third, the two sources were aligned term by term to produce Supplementary Dataset S1. This manually constructed dataset then served as the basis for the staged coding process described below, including the identification of cross-Strait comparability, lexical divergence, expression-style tendency, and likely communicative relevance.

The comparison matrix and coding workbook were constructed and managed in Microsoft Excel.

The staged workflow from source selection to comparability assessment and later-stage coding is summarized in Figure 1.



**Figure 1.** Staged workflow for constructing and coding the cross-Strait baseball terminology dataset. The figure presents the three procedural stages used in the study: dataset construction from Source 1, cross-Strait matching with Source 2, and coding and analysis of comparable, non-comparable, convergent, and divergent entries.

### 2.3. Analytical Dimensions

The analysis is organized around three dimensions: coverage difference, lexical divergence, and expression-style variation. These dimensions were chosen because they capture cross-Strait terminology difference at three related levels: the extent to which entries enter the shared documentary space, the way comparable concepts are named, and the manner in which those names are linguistically shaped. In that sense, the dimensions are not mechanically adopted from a single theoretical model. Rather, they arise from the structure of the dataset itself and from the aims of the study, while remaining broadly informed by a sociocognitive view of terminology, communicative approaches to specialist language, and the institutional logic of standardization (Cabr e, 1999, 2003; Temmerman, 2000).

The first dimension is coverage difference. This refers to the extent to which the entries recorded in Supplementary Dataset S1 can be matched with corresponding standardized Mainland forms in Source 2. In the present study, coverage is therefore understood in a documentary and dataset-based sense: it measures how far the selected Mainland standard enters the dataset, rather than the full range of baseball terminology used in practice. This dimension is especially important because the study does not compare two documents of identical scope. One source provides the main set of entries for comparison, while the other is a screened standard.

The second dimension is lexical divergence. This concerns cases in which the same English baseball concept is rendered differently across the Strait. Such differences may appear in personnel labels, umpiring terms, rule-related vocabulary, or technical expressions. At this level, the analysis is concerned with designation: where naming converges, where it diverges, and what kinds of divergence recur most consistently across the dataset.

The third dimension is expression-style variation. Here the focus shifts from lexical non-identity alone to broader differences in form and presentation. Some terms are shorter, more compressed, and more conventionalized in specialist use; others are fuller, more explicit, and closer to instructional or institutional language. This dimension is included because terminology difference in the present dataset is not exhausted by choosing different words for the same concept. It also involves different ways of shaping specialist meaning for different users, contexts, and communicative purposes.

Taken together, these three dimensions provide the basis for the staged coding and analysis that follow. Coverage difference establishes the documentary relationship between the two sources; lexical divergence identifies where naming differs within the directly comparable subset; and expression-style variation captures how those differences are further patterned in linguistic form. Read together, they provide a framework for examining how baseball terminology is selected, named, and stabilized across the Strait.

#### 2.4. Coding Framework

To make the comparison systematic, the study applies a coding framework to the term-by-term comparison matrix provided as Supplementary Dataset S1. The purpose of the coding is not to rank terms as correct or incorrect, but to distinguish among different layers of cross-Strait terminology difference that would otherwise collapse into a simple same/different contrast. In the present study, the coding framework serves as an analytical bridge between documentary comparison and interpretation.

The framework is layered. It begins with the most basic question—whether a Taiwan-side rendering and a corresponding Mainland standardized form can be identified for a given entry. Only after that does it move to the question of cross-Strait comparability. Once an item enters the directly comparable subset, the analysis then considers whether the two sides converge lexically or diverge, what kind of divergence is involved, whether a clear expression-style tendency can be observed, and whether the difference is likely to carry greater communicative consequence in practice.

In this study, convergent entries refer to cases in which the Taiwan-side and Mainland-side renderings are identical or differ only in script form. Near-convergent entries refer to cases in which the two renderings are not fully identical but show close lexical correspondence or functional equivalence in the dataset. By contrast, divergent entries refer to cases where the two renderings differ in lexical designation, expression style, conceptual packaging, or communicative sensitivity. For example, 妨礙 / 妨碍 is treated as convergent because the difference is mainly script-based. Cases where the two renderings are not identical but remain functionally equivalent are treated as near-convergent. Terms such as 捕手 / 接手 or 代打 / 替补击球员 are treated as divergent because they involve different designations or expression-style patterns.

This layered design is important because not every variable applies equally to every term. For example, a term that does not have a corresponding Mainland standard match cannot enter the directly comparable subset; in such cases, categories such as identity status, difference type, expression-style tendency, and communication-risk level are not applicable. Likewise, difference type applies only where a comparable item is lexically divergent, and expression-style tendency is used only where a clear directional contrast—such as greater compactness on one side or greater explicitness on the other—can reasonably be identified. The presence of uncoded or not-applicable cells therefore reflects the structure of the framework rather than incomplete coding.

In broader terms, the framework was developed from the needs of the dataset itself. It allows the study to distinguish documentary presence, lexical comparability, divergence pattern, stylistic tendency, and likely communicative relevance without forcing all items into a single undifferentiated category. This is particularly useful in a study informed by a sociocognitive view of terminology, since the aim is to observe how terms are selected, stabilized, and differentiated across related documentary traditions rather than to measure distance from a single ideal naming system (Temmerman, 2000).

The coding framework was applied through explicit decision rules. Table 2 provides a condensed coding framework, while Appendix A provides the full coding manual with detailed decision rules and examples from the terminology dataset. Later-stage categories such as difference type, expression-style tendency, and communicative sensitivity were used only where applicable, as indicated by the N/A code.

**Table 2.** Condensed coding framework for cross-Strait baseball terminology comparison.

Coding level	Main categories / codes	Purpose	Example from the dataset
Documentary presence	T1/T0; M1/M0	Identifies whether Taiwan-side and Mainland-side renderings are present in the source documents.	Homerun→全壘打/本壘打
Direct comparability	S1/S0	Distinguishes entries that can enter direct cross-Strait comparison from those that cannot.	Catcher→捕手/接手
Convergence status	C1/C0	Distinguishes convergent or near-convergent entries from divergent entries.	Balk→投手犯規/投手犯規; Pinch-hitter→代打/替補击球员
Difference type	D1-D5	Classifies major forms of divergence, including lexical designation, compactness/explicitness, register, rule sensitivity, and statistical/technical naming.	Obstruction→妨礙跑壘/阻擋
Expression-style tendency	E1/E2/E0	Identifies whether one side shows a clear tendency toward compactness or explicitness.	壘審/司壘裁判員
Communicative sensitivity	R1-R3	Marks whether a difference is likely to matter in practice, especially in rule explanation, translation, or instruction.	打數/自由击球数; 妨礙跑壘/阻擋
Not applicable	N/A	Used where later-stage coding is not applicable, especially for non-comparable entries.	English-only entries or entries without a Mainland standard match

Two final points should be noted. First, the categories are cumulative rather than competitive: a term may be comparable, lexically divergent, more compact on one side, and higher in communicative sensitivity at the same time. Second, the communication-risk distinction is intentionally modest. It does not claim to predict misunderstanding in every real context; rather, it marks those cases in which divergence is more likely to matter in rule explanation, commentary, teaching, or translation.

Used in this way, the coding framework helps the study move beyond impressionistic observation. It makes it possible to distinguish terms that remain outside the shared comparison space from those that are comparable but divergent, and to further separate routine lexical variation from differences more likely to matter in practice.

To improve coding reliability, the coding process combined sports-domain expertise and linguistic/terminological expertise. Initial term-by-term alignment and baseball-specific interpretation were conducted by a coder with expertise in baseball and physical education. The coding was then reviewed by a second coder with expertise in terminology, specialist-language analysis, and cross-Strait lexical variation. Borderline cases, especially those involving convergence or near-convergence, expression-style tendency, and rule-sensitive divergence, were discussed by the two coders and resolved through reference to the coding manual, the two source documents, and Supplementary Dataset S1.

The full coded dataset is provided as Supplementary Dataset S1 to make the coding decisions transparent and inspectable. This audit trail is important because several later-stage categories, such

as expression-style tendency and communicative sensitivity, involve judgment based on explicit decision rules.

### 3. Findings

#### 3.1. Structure of the Comparison Dataset and Coverage Results

Table 3 summarizes the structure of the updated comparison dataset. The cleaned matrix contains 363 English baseball entries, with Taiwan-side and Mainland-side Chinese renderings recorded where available. The full matrix, including comparability status, convergence status, and later-stage coding, is provided in Supplementary Dataset S1.

**Table 3.** Condensed coding framework for cross-Strait baseball terminology comparison.

Indicator	Count	Denominator	Percentage
Total baseball entries in main comparison list	363	363	100.0%
Entries with Taiwan-side Chinese renderings	361	363	99.4%
Entries with corresponding Mainland standardized renderings	214	363	59.0%
Directly comparable subset	214	363	59.0%
Convergent or near-convergent comparable terms	101	214	47.2%
Divergent comparable entries	113	214	52.8%
Entries not entering the directly comparable subset	149	363	41.0%
English-only entries with no Chinese rendering on either side	2	363	0.6%

\* Note. The figure of 214 refers only to the number of items in the present dataset for which a corresponding Mainland standardized form could be identified in Source 2. The table therefore reports coverage within Supplementary Dataset S1, not the full lexical range of baseball terminology in Mainland China (Chinese Baseball Association, 2020).

The table shows two main patterns. First, the Taiwan-side source provides near-complete coverage for the cleaned comparison dataset, with Chinese renderings available for 361 of the 363 entries. Second, the Mainland standard overlaps with 214 entries, or 59.0% of the dataset. This means that a substantial portion of the list enters direct comparison, but a sizeable group remains outside the directly comparable subset. Within the 214 comparable entries, the split is relatively balanced but slightly favors divergence: 101 entries are convergent or near-convergent, while 113 show lexical divergence.

These results reflect the different documentary roles of the two sources. The CTBA appendix provides the broader comparison list, while the CBA standard records a selected and codified set of preferred Mainland forms. For this reason, the 149 entries outside the directly comparable subset should not be interpreted as terms absent from Mainland baseball discourse. They indicate only that no corresponding Mainland standardized form was identified in the selected standard, or, in two cases, that the entry contains English only.

This distinction is important for the rest of the analysis. Only the 214 entries with renderings on both sides can be compared directly for convergence, divergence, expression style, and communicative sensitivity. Entries outside this shared comparison space are retained for dataset transparency but are not treated as lexical divergences.

#### 3.2. Distribution of Lexical Divergence Types

Within the 214 directly comparable entries, 113 entries are coded as divergent in Supplementary Dataset S1. To make the pattern of divergence more transparent, each divergent entry was assigned

a primary difference type. This coding does not imply that each term has only one possible interpretation; rather, it identifies the most salient form of difference for analytical reporting.

As Table 4 shows, the largest group of divergent entries involves different lexical designations. These are cases where the two sides use different established labels for the same baseball concept, without the difference being primarily a matter of length, register, or rule sensitivity. Examples include 捕手 / 接手 for Catcher, 飛球 / 高飞球 for Fly ball, and 全壘打 / 本壘打 for Homerun. These cases suggest that lexical divergence is not limited to marginal terms but appears in familiar and central baseball vocabulary.

**Table 4.** Primary types of lexical divergence in the directly comparable subset.

Primary difference type	Code	Count	Percentage of divergent entries
Different lexical designation	D1	69	61.1%
Compact conventional form vs. explicit standardized form	D2	12	10.6%
Register or institutional style difference	D3	3	2.7%
Rule-sensitive divergence	D4	16	14.2%
Statistical or technical naming divergence	D5	13	11.5%
Total divergent entries		113	100.0%

A second group involves contrasts between compact conventional forms and more explicit standardized forms. Examples include 野手 / 守场员 for Fielder, 代打 / 替补击球员 for Pinch-hitter, and 代跑 / 替补跑垒员 for Pinch-runner. In these pairs, the Taiwan-side rendering is often shorter and more conventionalized in baseball usage, while the Mainland standardized form tends to make the role or function more explicit.

A smaller but analytically important group involves register or institutional style difference, especially in terms related to roles or umpiring. For example, 主審 / 司球裁判员 for Plate umpire reflects not only different lexical choices but also different degrees of formalization. The Mainland form is more institutionally explicit, while the Taiwan-side form is more compact and conventional in baseball practice.

The dataset also contains 16 rule-sensitive divergences. These are particularly important because the terms are closely tied to rule explanation, umpiring judgment, or real-time interpretation. Examples include 妨礙跑壘 / 阻挡 for Obstruction, 封殺行為 / 封杀 for Force play, and 突襲投球 / 急投 for Quick-return pitch. Such cases are not necessarily more frequent than ordinary lexical differences, but they carry greater analytical weight because they may affect how rules are explained or interpreted.

Finally, 13 divergent entries involve statistical or technical naming. Examples include 打數 / 自由击球数 for Times at bat, 打擊率 / 安打率 for Batting average, and 自責率 / 责任失分率 for Pitcher's earned run average. These differences show that cross-Strait variation also appears in the statistical language of baseball, where precision and consistency are important for record-keeping, commentary, and instruction.

Taken together, these results show that lexical divergence in the dataset is patterned rather than random. Most divergent entries involve different lexical designations, but a substantial minority involve explicitness, rule sensitivity, or technical naming. This supports the view that cross-Strait baseball terminology differs not only in individual word choice, but also in the way specialist meaning is conventionalized, formalized, and made communicatively available across the two documentary traditions.

### 3.3. Expression-Style Tendency and Communicative Sensitivity

Supplementary Dataset S1 also codes the 113 divergent entries for expression-style tendency and communicative sensitivity. These categories help distinguish lexical difference from broader differences in how baseball meaning is packaged and how much practical consequence a difference may carry.

Table 5 shows that more than half of the divergent entries do not display a clear compactness/explicitness contrast. These cases are divergent mainly because the two sides use different lexical labels, as in 裁決 / 判定 for Adjudged, 捕手 / 接手 for Catcher, and 內野高飛球 / 內場高飛球 for Infield fly. In such cases, the difference is primarily lexical rather than stylistic.

**Table 5.** Expression-style tendency among divergent entries.

Expression-style tendency	Code	Count	Percentage of divergent entries
No clear expression-style tendency	E0	62	54.9%
Taiwan-side form is more compact or conventionalized	E1	28	24.8%
Mainland form is more explicit or definition-oriented	E2	23	20.4%
Total divergent entries		113	100.0%

At the same time, a substantial minority of divergent entries show a recognizable expression-style tendency. In 28 cases, the Taiwan-side form is coded as more compact or conventionalized. Examples include 野手 / 守场员 for Fielder, 飛球 / 高飛球 for Fly ball, and 違規投球 / 不合法投球 for Illegal pitch. In these cases, the Taiwan-side term often appears shorter or more established in baseball practice, while the Mainland form gives more explicit semantic information.

In 23 cases, the Mainland form is coded as more explicit or definition-oriented. Examples include 主審 / 司球裁判员 for Plate umpire, 壘審 / 司壘裁判员 for Base umpire / field umpire, and 犧牲觸擊 / 触击牺牲打 for Sacrifice bunt. These examples suggest that Mainland standardized terminology sometimes favors forms that spell out institutional role, action, or function more overtly. The pattern is not universal, but it is sufficiently recurrent to support the observation that expression style is one dimension of cross-Strait baseball terminology variation.

Table 6 indicates that most divergent entries are coded as low in communicative sensitivity. These are cases where different terms are unlikely to cause major difficulty among experienced baseball users because the context or shared baseball knowledge is likely to make the reference clear. Examples include 捕手 / 接手 for Catcher, 飛球 / 高飛球 for Fly ball, and 截止比賽 / 中止比賽 for Called game.

**Table 6.** Communicative sensitivity among divergent entries.

Communicative sensitivity	Code	Count	Percentage of divergent entries
Low communicative sensitivity	R1	69	61.1%
Moderate communicative sensitivity	R2	28	24.8%
High communicative sensitivity	R3	16	14.2%
Total divergent entries		113	100.0%

By contrast, 28 entries are coded as moderately sensitive because they may require clarification for learners, translators, commentators, or non-specialists. Examples include 代打 / 替補击球员 for Pinch-hitter, 代跑 / 替補跑垒员 for Pinch-runner, and 打數 / 自由击球数 for Times at bat. These terms are not necessarily rule-dangerous, but the difference in form may reduce immediate clarity for users who are familiar with only one terminology tradition.

The 16 high-sensitivity entries are especially important because they are closely connected with rule interpretation, umpiring, or real-time judgment. Examples include 促請裁決 / 申述 for Appeal,

封殺行為 / 封杀 for Force play, 妨礙跑壘 / 阻挡 for Obstruction, and 突襲投球 / 急投 for Quick-return pitch. These cases carry greater communicative weight because baseball rules often require fast and precise interpretation. In such contexts, terminology variation may affect how rules are explained, translated, or understood during play.

Taken together, the expression-style and communicative-sensitivity coding shows that lexical divergence is not uniform. Some differences are mainly routine alternatives in naming, while others involve differences in explicitness, institutional style, or rule-sensitive interpretation. This supports the broader argument that cross-Strait baseball terminology should be analyzed not only by whether terms differ, but also by how they differ and how much those differences are likely to matter in practice.

### 3.4. Rule-Sensitive Divergence and Umpiring Relevance

A closer look at the high-sensitivity items shows that communicative sensitivity is closely connected with rule-related terminology. In Supplementary Dataset S1, all 16 entries coded as high in communicative sensitivity are also coded as rule-sensitive divergences. These entries are not the largest group in numerical terms, but they are analytically important because they involve terms used in rule explanation, umpiring judgment, or real-time interpretation.

The entries in Table 7 cluster around several rule-related areas. Some concern appeal, protest, and judgment procedures, such as Appeal and Protesting game. Others concern illegal or irregular actions, such as Illegal pitch and Illegally batted ball. A further group relates to base running and put-out conditions, such as Force play, Retouch, Caught stealing, and In jeopardy. There are also entries connected with ball status and play continuation, such as Delayed deadball and Hit by pitch.

**Table 7.** High-sensitivity rule-related divergent entries.

English term	Taiwan rendering	Mainland rendering	Difference type	Communicative sensitivity
Appeal	促請裁決	申述	Rule-sensitive divergence	High
Force play	封殺行為	封杀	Rule-sensitive divergence	High
Forfeited game	褫奪比賽	弃权比賽	Rule-sensitive divergence	High
Illegal / Illegally	違規	不合法	Rule-sensitive divergence	High
Illegal pitch	違規投球	不合法投球	Rule-sensitive divergence	High
In jeopardy	負險狀態	出局危險	Rule-sensitive divergence	High
Obstruction	妨礙跑壘	阻挡	Rule-sensitive divergence	High
Quick-return pitch	突襲投球	急投	Rule-sensitive divergence	High
Retouch	再觸壘	再踏壘	Rule-sensitive divergence	High
Protesting game	提訴比賽	抗议比賽	Rule-sensitive divergence	High
Caught stealing	盜壘被刺	偷壘被杀	Rule-sensitive divergence	High
Force double play	順封出局	双封杀	Rule-sensitive divergence	High
Reverse force double play	逆封出局	封触双杀	Rule-sensitive divergence	High
Delayed deadball	緩時死球	延迟死球	Rule-sensitive divergence	High
Hit by pitch	觸身死球	投球中身	Rule-sensitive divergence	High
Illegally batted ball	違規擊球	不合法击球	Rule-sensitive divergence	High

The clearest example remains the distinction between Interference and Obstruction. In the comparison dataset, Interference is coded as convergent or near-convergent because both sides render it as 妨礙 / 妨碍. By contrast, Obstruction is coded as divergent and high in communicative

sensitivity because the Taiwan-side rendering 妨礙跑壘 and the Mainland-side rendering 阻挡 package the rule concept differently. This matters because the two English rule terms are closely related but not identical. If Chinese renderings do not keep the concepts sufficiently distinct, additional explanation may be needed in teaching, translation, commentary, or umpiring contexts.

Overall, the rule-sensitive entries show that the practical significance of terminology variation does not depend only on frequency. Some differences may be numerically limited but communicatively important because they occur in rule-governed contexts where precise interpretation matters. This supports the need to distinguish routine lexical divergence from higher-sensitivity terminology differences in cross-Strait baseball communication.

## 4. Discussion

### 4.1. Coverage Difference as a Documentary and Institutional Phenomenon

The coverage pattern reported in Section 3.1 should be understood primarily as a documentary and institutional phenomenon. The comparison dataset shows that the CTBA appendix provides a broad English-Chinese terminology list, while the CBA standard overlaps with 214 entries in that list. The remaining M0 entries should not be read as evidence that the corresponding baseball concepts are absent from Mainland baseball discourse. Rather, M0 indicates non-coverage in the selected standard used for this study.

This point is important because the CBA standard is not designed to reproduce every term that may circulate in coaching, commentary, translation, equipment description, training routines, or informal baseball use. As a standardizing document, it performs a selective institutional function: it identifies and codifies a foundational set of preferred Mainland forms. The 149 entries outside the directly comparable subset therefore reflect, at least in part, the standard's institutional choice to prioritize foundational terminology rather than to provide an exhaustive baseball lexicon.

Supplementary Dataset S1 makes this distinction explicit by marking M0 entries separately from divergent entries. This prevents the analysis from confusing non-coverage in the selected standard with terminological difference between the two sides. In other words, coverage difference establishes the boundary of direct comparison. Only entries with renderings on both sides can be examined for convergence, divergence, expression-style tendency, and communicative sensitivity.

From a terminology perspective, this finding supports the view that standardization is selective rather than exhaustive. Standards codify preferred forms for specific institutional purposes; they do not necessarily reproduce every expression circulating in actual practice (Ayres-Bennett & Bellamy, 2021). The CBA standard should therefore be read as one institutional response to terminology variation, not as a complete lexical map of Mainland baseball discourse. This point is central to the interpretation of the whole study: cross-Strait baseball terminology differs not only because terms are named differently, but also because the two documentary traditions define and codify the relevant lexical field differently from the outset.

### 4.2. Lexical Divergence and Community-Based Naming Traditions

The distribution of lexical divergence types shows that cross-Strait baseball terminology is not characterized by a single pattern of difference. Among the 113 divergent entries, the largest group consists of different lexical designations. This suggests that many differences are not primarily matters of formality, explicitness, or rule sensitivity, but of alternative naming traditions. Pairs such as 捕手 / 接手, 飛球 / 高飞球, and 全壘打 / 本壘打 show how the same baseball concept may be stabilized through different Chinese lexical choices.

This pattern is consistent with a sociocognitive view of terminology, which treats terms as historically situated and socially embedded rather than as fixed labels attached to concepts in an abstract system (Temmerman, 2000). From this perspective, divergence does not necessarily indicate conceptual disorder. It may reflect the fact that related baseball communities have developed

different but internally recognizable ways of naming the same roles, actions, or statistical categories. Freixa's account of denominative variation further supports this interpretation by showing that multiple designations for the same concept can arise from regional, functional, cognitive, and interlinguistic factors (Freixa, 2006).

The results also show that lexical divergence is distributed across different areas of baseball vocabulary. It appears in player roles, umpiring terms, rule-sensitive expressions, and statistical terminology. The distribution suggests that cross-Strait difference is not confined to peripheral expressions. Instead, it appears in core areas of baseball discourse, including familiar role labels and rule-related terms. The dataset therefore supports the argument that cross-Strait baseball terminology has developed through partially overlapping but non-identical naming traditions.

At the same time, the findings do not imply that all divergent terms create serious communication problems. Many divergent pairs are likely to remain manageable for experienced users because the surrounding baseball context helps identify the intended concept. The importance of the pattern lies less in immediate misunderstanding than in what it reveals about terminology stabilization. Terms circulate through rules, coaching, commentary, training materials, translation, and habitual use. Over time, different communities may stabilize different designations for the same concept, especially when they have different histories of baseball transmission and institutional codification.

This helps explain why cross-Strait baseball terminology should not be treated simply as a list of mismatched words. The more useful interpretation is that the two sides preserve related but distinct naming traditions. These traditions share a common conceptual field, often anchored in English baseball terminology, but they do not always select the same Chinese designations or the same degree of explicitness. Lexical divergence is therefore part of the broader process through which specialist knowledge is localized, conventionalized, and institutionalized in different Chinese-language settings.

#### 4.3. *Expression Style and the Packaging of Specialist Meaning*

The findings on expression-style tendency show that cross-Strait variation is not limited to lexical choice alone. Although more than half of the divergent entries do not show a clear compactness/explicitness contrast, a substantial number do. In these cases, the difference lies not only in what term is chosen, but also in how much semantic information the term makes explicit.

The contrast between compact conventional forms and explicit standardized forms is especially visible in pairs such as 代打 / 替补击球员, 代跑 / 替补跑垒员, 主审 / 司球裁判员, and 壘审 / 司垒裁判员. The Taiwan-side forms tend to be shorter and more conventionalized in baseball practice, while the Mainland-side forms often spell out the role, function, or institutional position more overtly. This does not mean that one form is necessarily better than the other. Rather, the two forms appear to serve partly different communicative orientations: one favors compactness and established field usage, while the other favors explicitness and standard-oriented clarity.

The historical dimension of this compactness is especially visible in Taiwan-side E1 terms that correspond closely to Japanese baseball kanji compounds. Morphologically, many of these E1 forms are two-character Japanese baseball kanji-compound structures, whereas the corresponding Mainland standardized forms often expand into longer descriptive phrases. For example, 代打 corresponds to Japanese daida, a compact baseball term for 'pinch hitter', while 主审 and 壘审 correspond to the Japanese umpiring terms 主审 and 壘审. Kelly's Japanese-English glossary of baseball terms lists 代打 (daida) as 'pinch hitter', 主审 as 'head umpire', 球审 as 'home plate umpire', and 壘审 as 'base umpires' (Kelly, 2018). These forms show how compact kanji compounds can function as established baseball terminology. In the Taiwan-side data, forms such as 代打, 主审, and 壘审 are therefore not merely short expressions; they are historically sedimented forms that are consistent with Taiwan's Japanese-influenced baseball pathway (Wang, 2009; Yu, 2007). This

helps explain why such terms remain conventionalized in Taiwan even when the Mainland standard prefers more explicit forms such as 替补击球员 or 司球裁判员.

This pattern is important for specialist-language analysis because terms are not only labels for concepts; they are also linguistic forms shaped for particular users and communicative contexts. Cabré's communicative theory of terminology is useful here because it treats terms as conceptual, linguistic, and communicative units at the same time (Cabré, 1999, 2003). A compact form may work efficiently among experienced baseball users, but a more explicit form may be easier for learners, translators, or readers encountering the concept through formal instruction. Expression style therefore affects how specialist knowledge is packaged and accessed.

The findings also suggest that standardization may encourage a more explicit naming style in some cases. Mainland standardized forms such as 司球裁判员, 司垒裁判员, 替补击球员, and 替补跑垒员 are semantically transparent and institutionally legible, even if they may sound less compact than their Taiwan-side counterparts. By contrast, Taiwan-side terms such as 主審, 壘審, 代打, and 代跑 reflect a more compressed terminology tradition that is likely to be familiar within established baseball communities.

This expression-style difference should not be overstated. Supplementary Dataset S1 shows that not all divergent entries follow this pattern, and many differences are mainly lexical rather than stylistic. Still, the recurring contrast between compactness and explicitness helps explain why some cross-Strait term pairs feel more different than others. It also shows that terminology variation involves not only the selection of different words, but also different strategies for making specialist meaning concise, conventional, explicit, or institutionally standardized.

#### 4.4. Rule-Sensitive Divergence and Communicative Consequence

The high-sensitivity items identified in Section 3.4 show that some terminology differences carry greater practical weight than others. Numerically, rule-sensitive divergences are not the largest group of divergent entries, but they are important because they occur in contexts where precise interpretation matters. Baseball rules often require immediate judgment, explanation, and application. In such contexts, terminology is not merely descriptive; it helps organize how players, coaches, umpires, translators, commentators, and learners understand the action taking place.

The clearest high-sensitivity case is the distinction between Interference and Obstruction. In official baseball rules, interference and obstruction are related but distinct concepts: offensive interference involves an act by the team at bat that hinders, obstructs, impedes, or confuses a fielder attempting to make a play, whereas obstruction refers to a fielder who, while not in possession of the ball and not fielding it, impedes the progress of a runner (Office of the Commissioner of Baseball, 2024). The Taiwan-side rendering 妨礙 for Interference preserves a broad meaning of 'interference' or 'hindrance', while 妨礙跑壘 for Obstruction narrows that meaning by explicitly adding the base-running component. By contrast, the Mainland standard renders Interference as 妨碍 but Obstruction as 阻挡, thereby creating a sharper lexical separation between the two rule concepts. This difference in conceptual packaging is why Obstruction is treated as high in communicative sensitivity. For umpires and rule interpreters, the risk is not simply that the two sides use different words; it is that the Taiwan-side pair keeps both concepts within the semantic field of 妨礙, while the Mainland pair separates them through 妨碍 versus 阻挡. In cross-regional rule explanation, umpire training, translation, or live commentary, this divergence may require explicit clarification so that users do not collapse the defensive act of obstructing a runner into the broader category of interference.

Other high-sensitivity examples show similar issues. Terms such as Force play 封殺行為 / 封杀, Retouch 再觸壘 / 再踏垒, Hit by pitch 觸身死球 / 投球中身, and Quick-return pitch 突襲投球 / 急投 are tied to rule explanation and umpiring judgment. Differences in these terms may not

necessarily cause misunderstanding among experienced baseball users, but they are more likely to require clarification in cross-regional teaching, translation, rule explanation, or umpiring-related communication.

This is why communicative sensitivity was included as an interpretive coding category. The purpose is not to predict misunderstanding in every actual situation. Rather, it marks cases where terminology difference is more likely to matter because the term is connected with rule-governed action. A low-sensitivity difference may simply represent alternative naming, while a high-sensitivity difference may affect how a rule is explained, learned, or applied. In this sense, the analysis supports a selective approach to terminology work: not all divergences require the same level of attention.

This also has implications for terminology standardization and terminology presentation. If all lexical differences are treated as equally problematic, standardization becomes too broad and blunt. A more communication-oriented approach would prioritize rule-sensitive terms, especially those used in rulebooks, umpire training, coaching materials, translation, and cross-regional competition. For less sensitive terms, it may be more useful to document regional variants and explain their correspondence rather than trying to eliminate variation altogether.

#### *4.5. Broader Contribution and Implications for Terminology Work*

This study shows that even in a highly codified sport such as baseball, terminology does not necessarily converge into a single uniform system across the Taiwan Strait. Instead, the dataset reveals a layered pattern in which shared baseball concepts are represented through different degrees of documentary coverage, lexical convergence, lexical divergence, expression-style tendency, and communicative sensitivity. This layered pattern is important because it shows that cross-Strait baseball terminology cannot be adequately understood through a simple same/different contrast.

The contribution of the study is both empirical and conceptual. Empirically, it provides a clearly bounded document-based comparison of 363 English baseball entries and their Taiwan-side and Mainland-side Chinese renderings. The cleaned dataset in Supplementary Dataset S1 makes the dataset transparent and allows the coding decisions to be inspected. Conceptually, the study supports an understanding of specialist terminology as a system shaped by usage tradition, institutional codification, historical transmission, and communicative need. This is especially relevant in the Chinese context, where shared written forms coexist with different regional naming traditions and standardization practices.

Although the empirical focus is baseball, the contrast between compact conventional usage and explicit institutional standardization is relevant to the management of cross-regional terminology in other specialist domains.

The findings also suggest a more realistic approach to terminology planning. If all cross-Strait differences are treated as equally problematic, terminology standardization risks becoming too broad and insensitive to actual usage. A more selective approach is preferable. Terms that bear directly on rule interpretation, umpiring, coaching, translation, and formal instruction deserve greater attention because ambiguity in these areas may have practical consequences. By contrast, low-sensitivity lexical alternatives that remain intelligible within ordinary baseball communication may not require the same degree of intervention.

The study also has implications for how baseball terminology is presented in dictionaries, teaching materials, rulebooks, and translation resources. In cross-regional contexts, it may be useful to present standardized forms together with widely used regional variants. Such a presentation would acknowledge actual variation while still helping users identify equivalence across terminology systems. This would be especially useful for learners, translators, commentators, and participants in cross-Strait baseball exchanges.

Terminology standardization, then, is better understood not as the elimination of all variation, but as the selective organization of variation. Its purpose should be to support clearer communication where misunderstanding is more likely to matter, while also recognizing that some forms of variation are stable, conventional, and manageable within specialist communities. For baseball across the

Strait, terminology work is most useful when it remains close to actual communicative needs and attends not only to lexical form, but also to how terms are used, taught, interpreted, and standardized.

## 5. Conclusions

This study examined terminological variation in Baseball Chinese across the Taiwan Strait through a structured document-based comparison of two primary sources: the CTBA appendix and the CBA's Basic Terminology of Baseball standard. Using a cleaned 363-entry comparison dataset, the analysis showed that cross-Strait terminology difference is not confined to isolated lexical substitutions. It appears in documentary coverage, lexical designation, expression style, and communicative sensitivity, and these dimensions together provide a more precise account of how specialist language varies within a shared written Chinese framework.

The first research question asked how baseball terms diverge across the Taiwan Strait in terms of coverage, lexical choice, and expression style. The results show that the two documentary sources overlap only partially. Of the 363 entries in the comparison dataset, 214 enter the directly comparable subset, while 149 remain outside direct lexical comparison because no corresponding Mainland standardized form was identified in the selected standard or because the entry contains English only. Within the comparable subset, 101 entries are convergent or near-convergent and 113 show lexical divergence. The divergent entries include different lexical designations, contrasts between compact conventional forms and explicit standardized forms, register or institutional style differences, rule-sensitive divergences, and statistical or technical naming differences. Cross-Strait baseball terminology is therefore neither fully unified nor simply fragmented; it is structured through overlapping but non-identical naming systems.

The second research question asked what historical and linguistic factors help explain these patterns. The findings suggest that terminology variation is shaped by the interaction of historical transmission, community usage, and institutional codification. Taiwan-side terminology often reflects a longer and more deeply conventionalized baseball tradition, including Japanese-influenced kanji-compound forms such as 代打, 主審, and 壘審. These compact expressions remain recognizable because they have circulated through baseball practice and instruction over time. Mainland standardized terminology, by contrast, often favors more explicit and institutionally transparent forms, such as 替补击球员, 司球裁判员, and 司垒裁判员. The contrast is thus not merely a difference between shorter and longer terms, but a difference in how baseball knowledge has been packaged, transmitted, and standardized in related but distinct documentary and institutional traditions.

The third research question asked what implications these differences have for terminology planning and practical baseball communication. The findings show that not all divergent terms carry the same communicative weight. Many low-sensitivity variants are likely to remain manageable for experienced users because context and baseball knowledge make the intended concept clear. Moderate-sensitivity terms may require cross-reference or explanation for learners, translators, commentators, and users moving between regional terminology systems. High-sensitivity rule terms require closer attention because they are tied to rule interpretation, umpiring judgment, instruction, translation, and real-time communication. In this respect, the distinction between Interference and Obstruction is especially important: it shows how terminological packaging can affect the clarity with which related but distinct rule concepts are understood.

These answers support a selective and communication-oriented approach to terminology work. Instead of treating every lexical difference as a problem requiring unification, terminology planning should respond to different levels of communicative sensitivity. Routine variants can be documented as equivalent regional forms. Terms that may create difficulty for learners or translators should be cross-referenced in glossaries, teaching materials, dictionaries, and translation resources. Rule-sensitive terms should receive the strongest coordination through rulebooks, umpire training materials, official translations, and cross-regional competition resources. Such an approach respects

established usage traditions on both sides while giving priority to clarity in the contexts where misunderstanding would matter most.

The study therefore contributes both empirically and conceptually. Empirically, it provides a clearly bounded comparison dataset of 363 English baseball entries and their Taiwan-side and Mainland-side Chinese renderings, with coding information made available in Supplementary Dataset S1. Conceptually, it demonstrates that specialist terminology should be analyzed not only by whether two forms are the same or different, but also by how they differ and how much those differences matter in practice. This supports a sociocognitive and communicative understanding of terminology as a system shaped by use, history, institutional selection, and communicative need.

The study remains documentary in scope. It captures how baseball terminology is selected, represented, and codified in two key written sources, not the full range of terminology use in live play, coaching, commentary, umpiring, or player interaction. Future research could extend this work by examining actual language use in baseball settings, including broadcast discourse, instructional materials, umpire training materials, coaching documents, and translation practice. Such work would strengthen the link between documentary terminology and communicative practice.

Although the empirical focus is baseball, the underlying issue is relevant to other specialist domains. The tension between compact conventional usage and explicit institutional standardization is likely to appear wherever shared concepts circulate across different regional, historical, or institutional traditions. Cross-Strait baseball terminology therefore offers a useful case for understanding how specialist language develops, varies, and can be coordinated without assuming that all variation must be eliminated. Its value lies not only in documenting difference, but also in showing how such difference can be analyzed and managed according to practical communicative needs.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Supplementary Dataset S1, a cleaned dataset of 363 English baseball terms and their Taiwan-side and Mainland-side Chinese renderings, together with coding information used in the analysis.

**Author Contributions:** Conceptualization, B.C. and Y.N.; methodology, B.C. and Y.N.; data curation, B.C. and Y.N.; formal analysis, B.C. and Y.N.; writing—original draft preparation, B.C. and Y.N.; writing—review and editing, B.C. and Y.N.. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The dataset generated and analyzed in this study is available as Supplementary Dataset S1.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

CBA	China Baseball Association
CTBA	Chinese Taipei Baseball Association

## Appendix A

### Appendix A.1

Table A1 provides the full coding manual used in the study. The condensed version appears in Table 2 in the main text.

**Table A1.** Full coding manual for cross-Strait baseball terminology comparison.

Coding dimension	Code / label	Description and decision rule	Example from the dataset
Taiwan-side rendering presence	T1	A Taiwan-side Chinese rendering is present for the English headword. Code T1 when the Taiwan-side column contains a Chinese rendering.	Homerun -> 全壘打
Taiwan-side rendering absence	T0	No Taiwan-side Chinese rendering is present. Code T0 when the Taiwan-side column is blank or when the entry is English-only.	English-only residual entries
Mainland standard match	M1	A corresponding Mainland standardized rendering is identified in the CBA standard. Code M1 when the Mainland-side column contains a standardized rendering corresponding to the English headword.	Homerun -> 本壘打
Mainland standard non-match	M0	No corresponding Mainland standardized rendering is identified in the selected CBA standard. This does not mean the term is absent from Mainland baseball usage.	Entries outside direct coverage of the selected standard
Cross-Strait comparability	S1	The item has both Taiwan-side and Mainland-side Chinese renderings and can enter direct comparison. Code S1 only when both sides provide Chinese renderings for the same English headword.	Catcher -> 捕手 / 接手
Non-comparability	S0	The item does not enter the directly comparable subset. Code S0 when one or both Chinese renderings are unavailable.	English-only entries; entries without a Mainland standard match
Convergent / near-convergent	C1	The two renderings are identical, script-equivalent, closely corresponding, or functionally equivalent. Code C1 when the two renderings show strong lexical or functional correspondence, even if they are not completely identical in form.	Interference -> 妨礙 / 妨礙; Balk -> 投手犯規 / 投手犯規

Divergent	C0	The two renderings differ lexically, stylistically, or conceptually beyond script difference or close functional correspondence. Code C0 when the two sides use clearly different designations or expression styles.	Catcher -> 捕手 / 接手; Pinch-hitter -> 代打 / 替补击球员
Different lexical designation	D1	The two sides use different established lexical labels for the same baseball concept. Apply when the difference is mainly a choice of different lexical items rather than length, register, or rule sensitivity.	捕手 / 接手
Compact conventional form vs. explicit standardized form	D2	One rendering is shorter and more conventionalized, while the other is more explanatory or definition-oriented. Apply when the contrast is mainly between compact specialist usage and fuller explanatory standardization.	代打 / 替补击球员
Register or institutional style difference	D3	The two renderings differ in degree of formality, official style, or institutional wording. Apply when one form sounds more like conventional field usage and the other more like formal institutional terminology.	主審 / 司球裁判员
Rule-sensitive divergence	D4	The difference concerns a rule-related concept where wording may affect explanation or interpretation. Apply when the term is closely tied to rule interpretation, umpiring, or real-time judgment.	妨礙跑壘 / 阻挡
Statistical or technical naming divergence	D5	The difference concerns statistical, scoring, or technical baseball terminology. Apply when the term belongs mainly to statistics, scoring, or technical measurement.	打數 / 自由击球数
Taiwan-side compactness	E1	The Taiwan-side form is shorter, more compressed, or more conventionalized in baseball practice. Apply when the Taiwan-side rendering is clearly more compact than the Mainland rendering.	壘審; 代打
Mainland-side explicitness	E2	The Mainland form is more explicit, descriptive, or institutionally codified. Apply when the Mainland rendering gives more	司壘裁判员; 替补击球员

		semantic detail or resembles a formal definition.	
No clear expression-style tendency	E0	No obvious compactness/explicitness contrast is identified. Apply when the difference is lexical but not clearly stylistic.	捕手 / 接手
Low communicative sensitivity	R1	The difference is unlikely to impede understanding among experienced baseball users. Apply when both forms are recognizable and context is likely to resolve the difference.	全壘打 / 本壘打
Moderate communicative sensitivity	R2	The difference may require clarification for learners, translators, commentators, or non-specialists. Apply when the terms differ enough to reduce immediate clarity outside experienced baseball communities.	打數 / 自由击球数
High communicative sensitivity	R3	The difference may affect rule explanation, umpiring, or real-time communication. Apply when the term is rule-sensitive and ambiguity may have practical consequences.	妨礙跑壘 / 阻挡
Not applicable	N/A	The category is not applicable to the entry. Use N/A when the item is not directly comparable or when a later-stage coding category does not apply.	English-only entries; entries without Mainland standard match

## References

1. Amorós-Negre, C., & Costa-Carreras, J. (2022). PRESENTATION of SPECIAL ISSUE: Variation and Change in Language Norms in Languages. *Languages*, 7(2), 145. <https://doi.org/10.3390/languages7020145>
2. Ayres-Bennett, W., & Bellamy, J. (2021). Introduction. In J. Bellamy & W. Ayres-Bennett (Eds.), *The Cambridge Handbook of Language Standardization* (pp. 1–24). Cambridge University Press. <https://doi.org/10.1017/9781108559249.001>
3. Bergh, G., & Ohlander, S. (2017). Loan translations versus direct loans: The impact of English on European football lexis. *Nordic Journal of Linguistics*, 40(1), 5–35. <https://doi.org/10.1017/S0332586517000014>
4. Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
5. Cabré, T. (1999). *Terminology: Theory, methods and applications* (J. C. Sager, Ed.; J. A. DeCesaris, Trans.; Vol. 1). John Benjamins Publishing Company. <https://doi.org/10.1075/tlrp.1>
6. Cabré, T. (2003). Theories of terminology: Their description, prescription and explanation. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 9(2), 163–199. <https://doi.org/10.1075/term.9.2.03cab>

7. Chinese Baseball Association. (2020). *Basic Terminology of Baseball*. <http://baseball.sport.org.cn/gfgg/qita/2020/0420/316349.html>
8. Chinese Taipei Baseball Association. (2024). *Proposed Standardized Chinese Translations of Baseball Terminology*. [http://www.ctba.org.tw/file/rules\\_2024.pdf](http://www.ctba.org.tw/file/rules_2024.pdf)
9. Faber, P. (Ed.). (2012). *A Cognitive Linguistics View of Terminology and Specialized Language*: DE GRUYTER. <https://doi.org/10.1515/9783110277203>
10. Ferguson, C. A. (1983). Sports announcer talk: Syntactic aspects of register variation. *Language in Society*, 12(2), 153–172. <https://doi.org/10.1017/S0047404500009787>
11. Freixa, J. (2006). Causes of denominative variation in terminology: A typology proposal. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 12(1), 51–77. <https://doi.org/10.1075/term.12.1.04fre>
12. Haddad Haddad, A., & Montero-Martínez, S. (2019). The ‘Carbon Capture’ Metaphor: An English-Arabic Terminological Case Study. *Languages*, 4(4), 77. <https://doi.org/10.3390/languages4040077>
13. Hiel, Q., & Zenner, E. (2023). English and Dutch terms in Belgian Dutch soccer reporting: A mixed-methods approach. *Soccer & Society*, 24(7), 1010–1026. <https://doi.org/10.1080/14660970.2023.2250664>
14. Hjørland, B. (2023). Terminology. *KNOWLEDGE ORGANIZATION*, 50(2), 111–127. <https://doi.org/10.5771/0943-7444-2023-2-111>
15. Lin, C.-Y., & Lee, P.-C. (2007). Sport as a medium of national resistance: Politics and baseball in Taiwan during Japanese colonialism, 1895–1945. *The International Journal of the History of Sport*, 24(3), 319–337. <https://doi.org/10.1080/09523360601101311>
16. Liu, F., & Liles, X. (2025). Regional variation in conventional expressions: A comparative study of Mandarin in Mainland China and Taiwan. *Lingua*, 326, 104035. <https://doi.org/10.1016/j.lingua.2025.104035>
17. Liu, J., & Chen, Y. (2022). A Comparable Corpus-Based Study of Cross-Strait Varieties: The Case of 注意 Zhùyì and 專心 Zhuānxīn. *Theory and Practice in Language Studies*, 12(3), 558–564. <https://doi.org/10.17507/tp1s.1203.16>
18. Maculan, B. C. M. D. S., Mesquita, M. S. L., & Falcão, L. C. D. J. (2023). A Glossary for Knowledge Organization Systems Terminology. *KNOWLEDGE ORGANIZATION*, 50(1), 23–33. <https://doi.org/10.5771/0943-7444-2023-1-23>
19. Milić, M. (2013). The influence of English on Serbian sports terminology. *ESP Today*, 1, 65–79.
20. Office of the Commissioner of Baseball. (2024). *Official Baseball Rules*. MLB.Com. <https://mktg.mlbstatic.com/mlb/official-information/2024-official-baseball-rules.pdf>
21. Reimerink, A., León-Araúz, P., & Cabezas-García, M. (2024). Phraseology and Culture in Terminological Knowledge Bases: The Case of Pollution and Environmental Law. *Languages*, 9(3), 84. <https://doi.org/10.3390/languages9030084>
22. Song, G. (2025). Intralingual translation as performance of identity: Linguistic negotiation in Mandarin across the Taiwan strait. *Social Sciences & Humanities Open*, 12, 101961. <https://doi.org/10.1016/j.ssaho.2025.101961>
23. Tao L. (2013). 两岸体育术语的对比分析——主要以棒球、篮球为例 [A comparative analysis of sports terminology across the Strait: Mainly baseball and basketball]. *语文学刊*, (19), 87–88; 119.
24. Taylor, M. (2013). Editorial – sport, transnationalism, and global history. *Journal of Global History*, 8(2), 199–208. <https://doi.org/10.1017/S1740022813000181>
25. Temmerman, R. (2000). *Towards New Ways of Terminology Description: The sociocognitive approach* (Vol. 3). John Benjamins Publishing Company. <https://doi.org/10.1075/tlrp.3>
26. Wang, S. (2009). Taiwanese Baseball: A Story of Entangled Colonialism, Class, Ethnicity, and Nationalism. *Journal of Sport and Social Issues*, 33(4), 355–372. <https://doi.org/10.1177/0193723509349938>
27. Wang, X., & Huang, C.-R. (2018). From Near Synonyms to Power Relation Variations in Communication: A Cross-Strait Comparison of “Guli” and “Mianli.” In J.-F. Hong, Q. Su, & J.-S. Wu (Eds.), *Chinese Lexical Semantics* (Vol. 11173, pp. 155–166). Springer International Publishing. [https://doi.org/10.1007/978-3-030-04015-4\\_13](https://doi.org/10.1007/978-3-030-04015-4_13)
28. William. W. Kelly. (2018). *A Japanese-English glossary of baseball terms*. [https://www.professional.wwkelly.net/wp-content/uploads/HT\\_Glossary\\_2018-1211.pdf](https://www.professional.wwkelly.net/wp-content/uploads/HT_Glossary_2018-1211.pdf)

29. Yin, J. (2021). Chinese terminology studies: Practical explorations and theoretical contributions. *Chinese Semiotic Studies*, 17(1), 71–88. <https://doi.org/10.1515/css-2021-0003>
30. Yu, J. (2007). *Playing in Isolation: A History of Baseball in Taiwan*. U of Nebraska Press.
31. Zhou, J., & Zhou, S. (2019). A Study on Differences Between Taiwanese Mandarin and Mainland Mandarin in Vocabulary. *Proceedings of the 3rd International Conference on Culture, Education and Economic Development of Modern Society (ICCESE 2019)*. Proceedings of the 3rd International Conference on Culture, Education and Economic Development of Modern Society (ICCESE 2019). <https://doi.org/10.2991/iccese-19.2019.48>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.