A Corpus-based Re-categorization for English and Chinese Hedges

Xiaole Liu[1,2], Roslina Binti Mamat[1], Salina Binti Husain[1]

[1] Faculty of Modern Languages and Communication, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

[2] School of Foreign Languages, Zhengzhou Business University, Zhengzhou, China

Correspondence: Xiaole Liu, Faculty of Modern Languages and Communication, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia.

Received: January 24, 2024 Accepted: July 22, 2025 Online Published: October 21, 2025

doi:10.5430/wjel.v16n2p143 URL: https://doi.org/10.5430/wjel.v16n2p143

Abstract

Hedge is an interesting linguistic phenomenon that plays a crucial role in cross-cultural communication. Exploring the differences in hedge use between Chinese and English discourse, as well as the reasons behind them, can further reveal the ideological significance reflected in the use of hedges and their role in the construction of discursive power. Based on two self-built corpora, comprising 122 speeches and statements from the Chinese and Malaysian Ministry of Foreign Affairs between the years 2023 and 2024, the study aims to establish a detailed and unified re-categorization for both English and Chinese hedges to enhance the operationalization of corpus linguistics for retrieving the usage characteristics of different types of hedges in Chinese and English discourse, thereby facilitating a cross-linguistic contrastive analysis. Based on Prince et al.'s (1982) and He's (1985) categorization, as well as Varttala's (2001) categorization of hedges from a cognitive lexical perspective, this study divides the existing four types of hedges – adaptors, rounders, plausibility shields, and attribution shields – into 10 specific sub-types based on their pragmatic functions. The finding indicates that, among the ten subcategories of hedges, the usage differences in eight subcategories between the Chinese and Malaysian corpora show significant differences. Therefore, the new subcategories of hedges presented in this article provide valuable insights for future scholars to conduct statistical analysis using corpus linguistics in contrast analysis between English and Chinese discourse.

Keywords: re-categorization for English and Chinese hedges, corpus linguistics, contrast analysis

1. Introduction

The term "hedge" was first introduced by Lakoff (1973), an American linguist, referring to "words whose job is to make things fuzzier or less fuzzy" (p. 471). Since then, an increasing number of linguists have studied "hedges". As a specific and common type of fuzzy language, "hedges" aroused worldwide concern among linguists from various countries.

Hedges are a fascinating linguistic phenomenon in many spoken and written languages, playing a vital role in effective communication. Hedges reflect speakers' linguistic strategies to convey and perform certain speech acts or to realize specific communicative intentions. Since "hedges" were first introduced by Lakoff in 1973, linguists have extensively researched "hedges." The research field has gradually expanded from theoretical research on semantics and logic to practical research such as pragmatics, discourse analysis, translation, cross-cultural communication, and language teaching. The selection of the research discourse has also expanded from doctor-patient communication discourse and academic discourse to advertisement discourse, news discourse, and courtroom trial discourse. However, according to the existing literature, there are relatively few studies on the comparative analysis of pragmatic strategies used in Chinese and English hedges across cross-cultural contexts.

Corpus linguistics is a research methodology that has evolved in recent decades to facilitate empirical studies of language variation and use. This methodology leads to findings that are significantly more generalizable, and valid than would be possible through other methods (Biber, 2009). The data analysis, conducted using a large corpus, not only addresses the limitations of previous studies that primarily focus on commonly used hedges but also enables a more comprehensive and objective examination of the characteristics of hedge use, thereby enhancing understanding of their nature and function.

Hedges are the most common and representative form of fuzzy language, characterized by their unique semantic features and rich pragmatic functions that other expressions cannot replace. The effective use of hedging language reflects the speaker's linguistic proficiency and significantly impacts the communication outcome. In recent years, comprehensive and in-depth comparative studies of the pragmatic strategies of hedging in Chinese and English, facilitated by large-scale corpora, have become essential for enhancing and advancing the theoretical framework of fuzzy linguistics and its applied research.

However, in the cross-linguistic contrastive study, some subcategories under the four existing classifications of hedges have notably different usage characteristics in Chinese and English discourses. For example, spokespersons from China and Malaysia use different subjects in front of the plausibility shields when expressing their subjective opinions (e.g., "I think," "I argue," "we insist") in political discourse. The differences in the types of subjects used by Chinese and Malaysian spokespersons reflect the differences in discourse

strategies and ideologies in their construction of discourse identity. Only by exploring the differences in the use of specific hedges and their reasons can the ideological significance be further exposed, as well as their role in the construction of discursive power. However, a comprehensive and unified classification of hedges *for English and Chinese discourse* that facilitates corpus linguistic retrieval searches and statistical analysis has yet to be established, hindering the possibility of conducting contrast analysis using corpus data. The current classification of hedges lacks a unified standard suitable for both Chinese and English discourse, and it does not provide a more detailed classification based on pragmatic functions.

The study aims to investigate a comprehensive and systematic re-categorization of English and Chinese hedges within the frameworks of the classifications proposed by Prince et al. (1982) and He (1985), grounded in a thorough analysis of their linguistic functions based on two self-built corpora. Therefore, the focus of this article is to answer the following research questions:

(1) Why re-categorize hedges under the frameworks of the classifications of Prince et al. (1982) and He (1985)?

(2) What are the new categories for English and Chinese hedges under the frame of the classifications of Prince et al. (1982) and He (1985)?

(3) How feasible is the new categorization of hedges in the contrast analysis of Chinese and English discourse?

2. Literature Review

The categorization of English and Chinese hedges has been a subject of extensive research, with scholars offering various frameworks to understand their linguistic functions. This section examines early categorizations of English hedges, with a primary focus on influential works. It delves into the pioneering efforts of scholars like Lakoff (1973), Zadeh (1972), and Myers (1989), who contributed to the development of hedge classification through both semantic and grammatical perspectives. These early studies laid the foundation for understanding hedges in English, highlighting their flexibility in modifying meaning and expression. Additionally, this section introduces the categorization of Chinese hedges, focusing on the contributions of Wu (1999) and He (1985), which parallel English categorization in some respects while reflecting the unique linguistic features of the Chinese language.

2.1 Early Categorization of English Hedges (Pre-1990s)

Lakoff (1973) did not propose a specific categorization of hedges. However, in his article *Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts*, he provided a table with various examples of hedges in English, which are shown in Table 1 below:

Table 1. Hedges Example List (Lakoff, 1973)

SOME HEDGES AND RELATED PHENOMENA	
sort of	in a real sense
kind of	in an important sense
loosely speaking	in a way
more or less	in a manner of speaking
on the ___ side (tall, fat, etc.)	details aside
roughly	so to say
pretty (much)	a veritable
relatively	a true
somewhat	a real
rather	a regular
mostly	virtually
technically	all but technically
strictly speaking	practically
essentially	all but a
in essence	anything but a
basically	a self-styled
principally	he calls himself a
particularly	in name only
par excellence	actually
largely	really
for the most part	he, as much as
very	-like
especially	-ish
exceptionally	can be looked upon as
quintessential(ly)	can be viewed as
literally	pseudo-
often	crypto-
more of a ___ than anything else	(he's) another (Caruso/Lincoln/Babe Ruth/ ...)
almost	_____ is the ___ of ____ (e.g.,. America is the Roman Empire of the modern world, Chomsky is the DeGaulle of Linguistics, etc.)
typically/typical	
as it were	in a sense

From the examples and definitions provided, we can deduce Lakoff's general approach to categorizing hedges: the linguistic forms of

hedges, according to Lakoff, encompass not just single words but also phrases, predicate constructions, and even affixes.

In 1972, Zadeh categorized hedges into two major types in his paper, *A Fuzzy Set: Theoretical Interpretation of Linguistic Hedges*. The first type directly modifies fuzzy words but does not apply to precise words. For example, one can say "very strong" but not "very circular" because "strong" is a fuzzy term, while "circular" is precise. This category includes terms like "more or less," "sort of," "much," and "slightly." The second category of fuzzy terms explains how they apply to fuzzy words, such as "essentially," "technically," "regularly," and "strictly speaking".

Zadeh (1972) also proposed a grammatical perspective, categorizing hedges into four types: adjectives and adverbs (e.g., "always," "usually," "almost"), suffixes (e.g., "-ly," "-ish"), manner adverbs indicating hesitation and uncertainty, and degree adverbs (e.g., "so…that," "as though"). Additionally, sentence structures expressing politeness, such as "I think" and "I believe," were recognized as hedges.

Zuck and Zuck (1985) expanded the definition, arguing that hedges also include modal auxiliary verbs (e.g., "would," "could," "should"), semi-auxiliary verbs (e.g., "seem," "appear"), verbs (e.g., "consider," "speculate"), adverbs and adverbial phrases, certain adjectives (e.g., "possible," "plausible"), non-specific nouns, and indefinite pronouns.

Myers (1989) pointed out that hedges primarily consist of modal verbs, modifiers (e.g., "presumably," "likely"), non-fact verbs (e.g., "suggest," "provide"), hesitation and degree markers (e.g., "it is likely that"), and personal and impersonal attribution markers (e.g., "I would say," "as reported").

2.2 Late Categorization of English Hedges (Post-1990s)

Building on previous research, Hyland (1998) categorized hedges from a semantic and morphological perspective into two main types: lexical and strategic. Figure 1 shows that although this classification does not cover all fuzzy constraints in English, it provides a solid foundation for further systematic research.

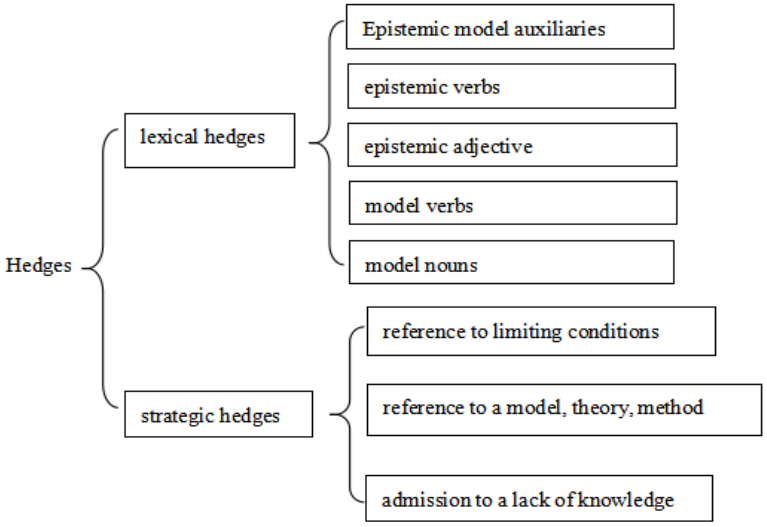

Figure 1. Classification of Hedges from a Semantic Perspective (Hyland, 1998)

Hyland (1998) further divided hedges in academic writing pragmatically, considering the motivations behind their use. He identified two types of hedges: content-oriented and reader-oriented. Content-oriented hedges aim to mitigate the relationship between propositional content and a non-linguistic mental representation of reality, essentially hedging the correspondence between what is said and what is thought to be true. In contrast, reader-oriented hedges focus on the interactional effects of statements, highlighting the relationship between the writer and reader. Content-oriented hedges are further divided into writer-oriented hedges, which minimize the writer's responsibility, and accuracy-oriented hedges, which aim for precision. Accuracy-oriented hedges address the relationship between propositions and reality, as shown in Figure 2.

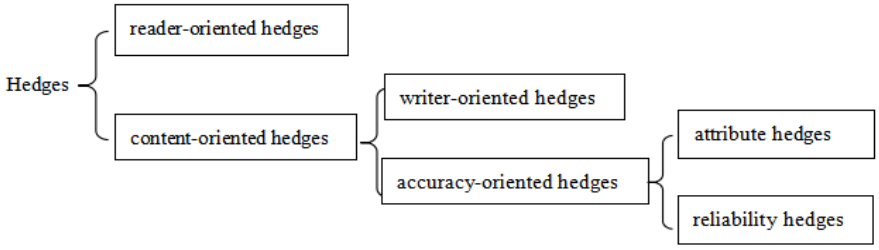

Figure 2. Classification of Hedges from the Pragmatic Perspective (Hyland, 1998)

Although Hyland's classification is comprehensive, it has limitations due to the semantic uncertainty and complexity of hedges' linguistic functions. Nonetheless, Hyland's framework offers valuable insights into understanding hedges in academic discourse and serves as a reference for further research in other contexts.

Varttala (2001) also contributed significantly to hedges classification, providing a list of 253 hedges. His categorization, which focuses on lexical hedges, is detailed in Table 2 below. Varttala's classification does not include non-lexical types, such as sentence structures or suffixes, since both he and Hyland considered lexemes the most important form of hedges in academic writing.

Table 2. Classification of Hedges (Varttala, 2001)

Classification	Examples
modal auxiliaries	can, could, may, might, must, should, will, would
epistemic verbs	argue, believe, suggest, conceive, assume, anticipate, point, imply
epistemic adverbs	possibly, frequently, approximately, highly, typically
epistemic adjectives	doubtful, potential, likely, plausible, apparent, suggestive
epistemic nouns	suggestion, possibility, assumption, argument, claim, proposition

Hyland (1996) reported that over 85% of hedges in academic papers are lexical, with non-lexical forms constituting less than 15%. He underscored the importance of epistemic hedges as a fuzzy strategy in academic discourse. Salager-Meyer (1994), in her analysis of written medical English, categorized hedges into five types, incorporating linguistic and medical sociological perspectives:

(1). Shields: Modal auxiliaries, semi-adjuncts, adverbs of possibility, and corresponding adjectives (e.g., might, probably).

(2). Approximators: Terms that limit the ambiguity of quantity, extent, frequency, and time (e.g., almost, frequently, about).

(3) Author's doubt and direct involvement: Expressions reflecting the author's uncertainty or personal involvement (e.g., I believe).

(4) Emotionally-charged intensifiers: Hedges that intensify the author's emotional stance (e.g., extremely, highly).

(5) Compound hedges: Phrases with multiple hedges (e.g., it may suggest that it would seem unlikely).

Salager-Meyer's (1994) categorization aligns with Prince et al. (1982), although their terminological use and classification approach differ. The most detailed and influential classification to date is that of Prince et al. (1982), who analyzed hedges in doctor-to-doctor conversations. Their approach, based on Lakoff's (1973) focus on hedges' bi-directional functions, is presented in the tree diagram below:

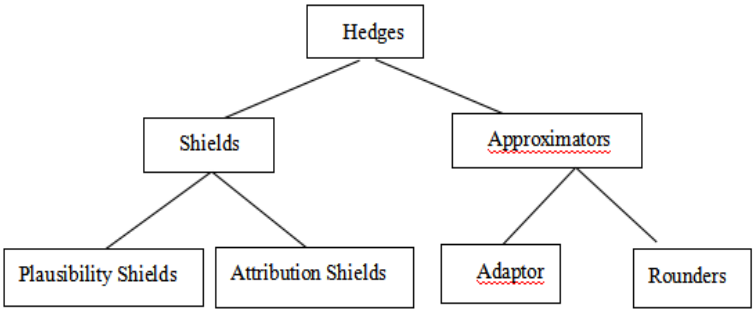

Figure 3. The Classification of Hedges (Prince et al., 1982)

Other scholars (e.g., Clemen, 1997; Holmes, 1982, 1984; House & Kasper, 1981; Hübler, 1983; Millan, 2008) have also classified hedges based on semantics, lexicality, syntactic structure, and pragmatics. Due to differing perspectives and interpretations, a standardized classification of hedges remains elusive. Previous scholars have classified hedges from various perspectives, including lexical, grammatical, and pragmatic. Although hedges have not yet formed a unified classification, the classification of hedges based on pragmatic function by Prince et al. (1982) has been among the most influential and widely accepted ones. This preference is also why the author uses Prince et al.'s (1982) classification of English hedges as the basis for conducting a unified classification of English and Chinese hedges in this study.

2.3 Categorization of Chinese Hedges

In China, Wu (1999) described the term "hedge" as a 'fuzzy restrictive element.' From a semantic perspective, Wu (1999) classified fuzzy restrictive expressions in Chinese into four categories based on their grammatical function:

(1) Adverbs and adjectives: The first category includes words such as the adverbs "很" (very), "实际上" (actually), and the adjectives "十足的" (quite), which modify the intensity or scope of an expression.

(2) Suffixes: The second category consists of suffixes that indicate a slight degree or approximation, such as the suffixes "-ish", seen in words like "带点绿色的" (greenish) or "有点甜的" (sweetish), which convey a sense of partiality or approximation.

(3) Subordinate clauses: The third category involves the use of subordinate clauses to express the degree of an ambiguous adjective or adverb, as in constructions such as "如此…以至于…" (so… that…) or similar structures used to describe varying degrees of quality.

(4) Subjective expressions: The fourth category includes phrases or clauses that reflect subjective opinions or perspectives, such as “我认为他很自大” (I think he is very arrogant), “在他的印象里，她非常自负” (According to his impression, she is very proud), and “我估计他可能活不久了” (I reckon he may not be long for this world).

In addition, Wu (1999) further categorized hedges based on the type of word they modify. One category consists of elements that only modify fuzzy words, such as “非常好” (very good) and “基本正确” (basically correct). The other category includes elements that can modify both fuzzy words and precise words, such as “将近黄昏” (nearly dusk) or “接近完美” (closely perfect), as well as precise terms like “将近十点” (close to ten o’clock) and “将要结束” (approaching the end).

He (1985), drawing on the framework of Prince et al. (1982), provided further elaboration on the classification of hedges. According to He, approximators serve to modify the content of the discourse, affecting its truthfulness and scope. Approximators can modify the intended meaning or provide a scope of variation for the discourse, aligning with semantic categorization.

He (1985) also subdivided approximators into two types: adaptors and rounders.

(1) Adaptors: Adaptors play a significant role in speech communication, as they allow the speaker to express ideas that are close to being accurate but not entirely true, thus making the discourse sound more polite or acceptable. This approach helps avoid overly arbitrary statements. Examples in English include terms like “some,” “very,” “kind of,” “a little bit,” “almost,” “somewhat,” “more or less,” and “really.” In Chinese, equivalent terms include “很” (very), “有点儿” (a little/bit), “或多或少” (more or less), “相当” (quite), “某种程度上” (to some degree), and “几乎” (almost).

(2) Rounders: Rounders are used when the speaker does not provide precise figures due to uncertainty or the impossibility of giving an exact number. These expressions provide a general range or approximation to guide the listener’s understanding. Examples in English include terms such as “around,” “approximately,” “roughly,” and “over,” as well as expressions like “something between X and Y.” In Chinese, similar terms include “大约” (around), “大概” (approximately), “左右” (roughly), “超过” (over), and “在…之间” (something between X and Y).

According to He (1985), shields, on the other hand, do not affect the truth conditions of a sentence but instead reflect the speaker’s tone and attitude, placing them within the pragmatic category. He (1985) categorized shields into two subtypes: plausibility shields and attribution shields.

(1) Plausibility shields express the speaker’s speculation or hesitation, often used when the speaker is not entirely confident in the truth of a statement or is reluctant to make a definitive assertion. Examples of plausibility shields in English include expressions like “I think,” “I wonder,” “I suspect,” “probably,” and “as far as I can tell.” In Chinese, similar expressions include “我认为” (I think), “我估计” (I guess), “我怀疑” (I suspect), “可能” (probably), and “就我而言” (as far as I can tell).

(2) Attribution shields refer to a third party’s viewpoint or opinion, indirectly conveying the speaker’s attitude. These shields are frequently used in news reporting or when the speaker wants to distance themselves from the expressed opinion, adding an element of objectivity. Examples in English include phrases like “it is said that…,” “according to someone’s estimates,” “someone says that…,” and “as it is well known.” In Chinese, equivalent expressions include “据说” (it is said that…), “据某人估计” (according to someone's estimate), “众所周知” (as it is well known), and “据某人所说” (someone says that…).

These hedges provide a nuanced understanding of how fuzzy and vague expressions are employed in both Chinese and English, reflecting the speaker’s attitude, uncertainty, or desire to soften the impact of their statements.

Research on the classification of Chinese hedges began relatively late and remains scarce. The most influential study to date is by He (1985), who, drawing on the classification criteria for English hedges proposed by Prince et al. (1982), made several additions and provided a detailed explanation. However, his classification of hedges fails to account for the differences between English and Chinese hedges, limiting their depth and applicability.

3. Methodology

3.1 Research Design

The corpus-based re-categorization of English and Chinese hedges will be conducted using both quantitative and qualitative methods, with a focus on their linguistic functions and communicative roles. Firstly, a quantitative method will be adopted to collect and process the corpus data. Then, the author will analyze the results and discussion to check the feasibility of new categories of English and Chinese hedges under the guidance of Prince et al.’s (1982), He’s (1985) categorization, as well as Varttala’s (2001) categorization of hedges from a cognitive lexical perspective.

3.2 Research Procedures

The first step is to re-categorize the English and Chinese hedges under the guidance of Prince et al.’s (1982) and He’s (1985) categorization, as well as Varttala’s (2001) categorization from a cognitive lexical perspective. The second step is to build the corpus required for the article, design reasonable annotations, and then use the annotation software to process the corpus data manually. After the annotation is completed, research tools will be used to retrieve, compile, and analyze the results, specifying the feasibility of new categories of English and Chinese hedges in the contrast analysis between Chinese and Malaysian political discourse.

3.3 Research Tools

3.3.1 Tokenization Software

Tokenization in English is typically achieved by ensuring that a space follows each word. However, tokenization in Chinese is much more complicated and important because there are no spaces between words in Chinese text, unlike in English.

The tokenization software used in the Chinese text of this article is Corpus Word Praser, a free lexical analysis system developed by Xiao Hang at the Institute for Applied Research in Linguistics, Ministry of Education. It is specifically designed for the Chinese language, with primary functions including Chinese tokenization and lexical annotation. Figure 4 shows a sample text of the Chinese data after tokenization using the Corpus Word Praser software in this study:

携手 创建 国际 调解 院 打造 和平 解决 国际 争端 新 平台
——秦刚 部长 在 国际 调解 院 筹备 办公室 成立 仪式 上 的 致辞
尊敬 的 各位 来宾，
女士 们，先生 们，朋友们：
大家 好！
很 高兴 出席 国际 调解 院 筹备 办公室 成立 仪式。筹备 办公室 的 成立 是 过去 几 年 大家 不懈 努力 的 硕果，标志 着 国际 调解 院 进入 实质性 创建 的 新 阶段。我 代表 中国 外交部 表示 热烈 祝贺，并 感谢 大家 的 辛勤 努力 和 卓越 贡献。
女士 们，先生 们，朋友们，
中国 国家主席 习近平 指出，当前，世界 之 变、时代 之 变 和 历史 之 变 正 以 前所未有 的 方式 展开。和平、发展、合作、共赢 是 大势所趋。但 某些 国家 逆 历史 潮流，大搞 单边 主义、强权政治，破坏 国际 法治。我们 更加 需要 坚定 地 维护 以 联合国 为 核心 的 国际 体系、以 国际法 为 基础 的 国际 秩序、以 联合国 宪章 宗旨 和 原则 为 基础 的 国际 关系 基本 准则。本着 这 一 精神，中国 倡议 建立 国际 调解 院，得到 许多 国家 积极 响应。
国际 调解 院 将 是 世界 上首 个 专门 以 调解 方式 解决 国际 争端 的 政府 间 国际 法律 组织，将 超越 司法 和 仲裁 你 输 我 赢 的 局限性，旨 在 实现 争端 各方 的 合作 共赢，对 促进 国际 和平 安全 发展 和 国际 秩序 稳定 具有 重要 意义。国际 调解 院 是 践 行 联合国 宪章 规定 的 和平 解决 国际 争端 原则 的 重要 实践，将 丰富 和 发展 国际 争端 解决 的 机制 和 方式；也是 我们 发起 国 向 国际 社会 提供 的 一 项 全球 法治 公共 产品，将 推动 全球 治理 体系 的 不断 发展 完善。我们 相信，国际 调解 院 必将 充分 发挥 其 灵活、经济、便捷 的 独特 优势，为 各国 和平 解决 国际 争端 提供 新 的 选择。
女士 们，先生 们，朋友们，
未来 的 筹备 工作 任务 更 艰巨，使命 更 光荣。需要 我们 集 众 智、汇 群 力、增 互 信，共同 打造 一个 权威 可信、高效 专业 的 国际 调解 机构。我 提 四 点 建议：
一 要 恪守 相互 尊重 的 宗旨。要 尊重 各国 主权，尊重 各国 核心 利益 和 合理 关切，推动 当事 方 平等 对话、建立 互 信，不 搞 胁迫 和 施压，多 做 弥合 分歧 的 事情，通过 协商 找到 解决 方案。
二 要 发挥 共商 共建 的 合力。要 面向 世界 各国，对 所有 国家 开放，不仅 包括 发展中国家，也 欢迎 发达国家 加入。建立 国际 调解 院 是 我们 大家 的 事情，我们 要 商量着 办，求同存异，携手 共建。
三 要 秉持 开放 包容 的 精神。要 兼容并蓄，要 充分 吸纳 各 大 文明 和 各 大 法系 优 长，将 国际 调解 院 打造成 文明 互 鉴 的 范例，更 好 造福 人类 和平 事业。
四 要 实现 均衡 发展 的 目标。要 着力 实现 地域、性别 等 平衡，增强 发展中国家 的 参与 度，培养 发展中国家 的 调解 员，推动 国际 争端 解决 机制 向 更加 公平 普惠 的 方向 发展。
我 相信，在 我们 的 共同 努力 下，国际 调解 院 必将 成功 设立 和 运作，“一国两制”下 的 香港 也 必将 充分 发挥 独特 的 地位 和 优势，为 国际 调解 院 的 建立、发展 贡献 力量！
谢谢 大家！

Figure 4. Sample Text of Tokenization in Chinese Corpora

3.3.2 Annotation Software

The Beijing Foreign Studies University (BFSU) Qualitative Coder was chosen as a qualitative data annotator for this study, which was developed by Prof. Xu Jiajin and Jia Yunlong of BFSU in 2011. The BFSU Qualitative Coder can annotate texts according to research needs and count the words and word frequency of each subclass of the annotation system. First, according to the research requirements, the annotation code was designed, as shown in Table 3, and the encoding list was completed in the prescribed format of the BFSU Qualitative Coder. After completing the above work, the text can be imported into the BFSU Qualitative Coder for annotation. The Chinese and English annotation interfaces are shown in Figures 5 and 6, respectively.

Table 3. Annotation Code and Examples of Annotation

Types of Hedges		Annotation Code	Examples of Annotation
Rounder	Quantity-varying rounders	<Rounder-A>	Chinese: <Rounder-A>大约</Rounder-A> English:<Rounder-A>around</Rounder-A>
	Frequency-varying rounders	<Rounder-B>	Chinese:<Rounder-B>一直</Rounder-B> English:<Rounder-B>always</Rounder-B>
	Enumeration-varying rounders	<Rounder-C>	Chinese:<Rounder-C>等</Rounder-C> English:<Rounder-C>and so on</Rounder-C>
Adaptor	Reduced-ambiguity adaptors	<Adaptor-A>	Chinese:<Adaptor-A>很</Adaptor-A> English:<Adaptor-A>quite</Adaptor-A>
	Restricted-proposition adaptors	<Adaptor-B>	Chinese:<Adaptor-B>严格来说</Adaptor-B> English:<Adaptor-B>technically</Adaptor-B>
Plausibility shield	Subjective plausibility shields	<Plausibility-A>	Chinese:<Plausibility-A>我相信</Plausibility-A> English:<Plausibility-A>I believe</Plausibility-A>
	Modal plausibility shields	<Plausibility-B>	Chinese:<Plausibility-B>能够</Plausibility-B> English:<Plausibility-B>can</Plausibility-B>
	Speculative plausibility shields	<Plausibility-C>	Chinese:<Plausibility-C>好像</Plausibility-C> English:<Plausibility-C>seem to</Plausibility-C>
Attribution shield	Attribution shields with sources	<Attribution-A>	Chinese:习近平 <Attribution-A>说</Attribution-A> English:Lee <Attribution-A>insisted</Attribution-A>
	Attribution shields without sources	<Attribution-B>	Chinese:<Attribution-B>据报道</Attribution-B> English:<Attribution-B> It is reported that </Attribution-B>

Figure 5. Annotation Interface of Chinese Corpus

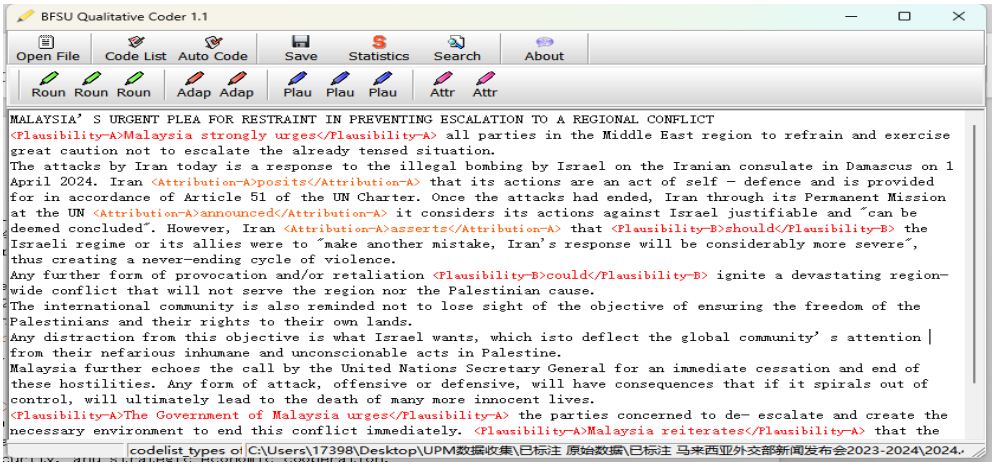

Figure 6. Annotation Interface of English Corpus

3.3.3 Retrieval Software

In this study, AntConc 4.3.1 was chosen as the retrieval software. It is a free software developed by Laurence Anthony, a British graduate of Waseda University in Japan. With its concise interface, ease of operation, and comprehensive set of functions, this software has become widely used in corpus linguistics research. Figures 7 and 8 show the operation interfaces when AntConc 4.3.1 is used to retrieve the Chinese rounder “一直” and the English rounder “always”.

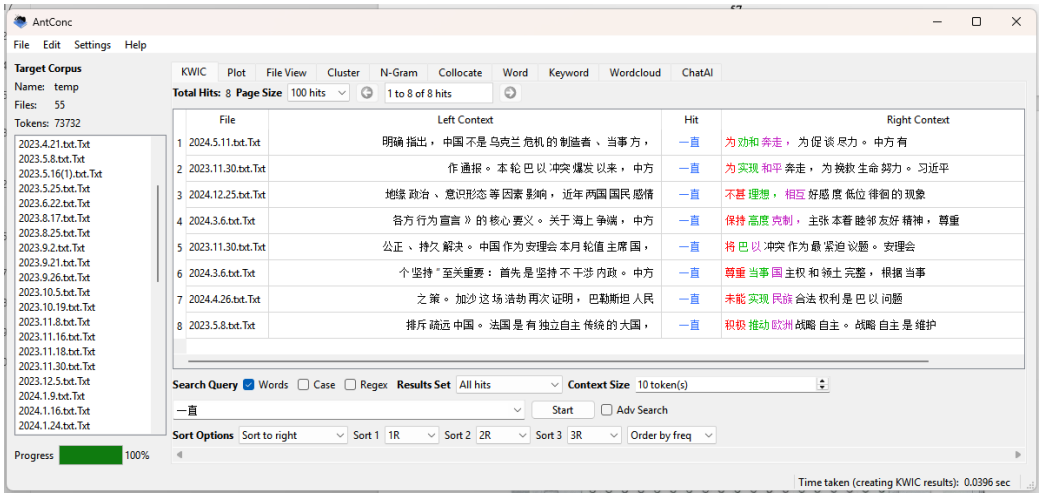

Figure 7. Retrieving Chinese Rounder “一直” in AntConc4 .3.1

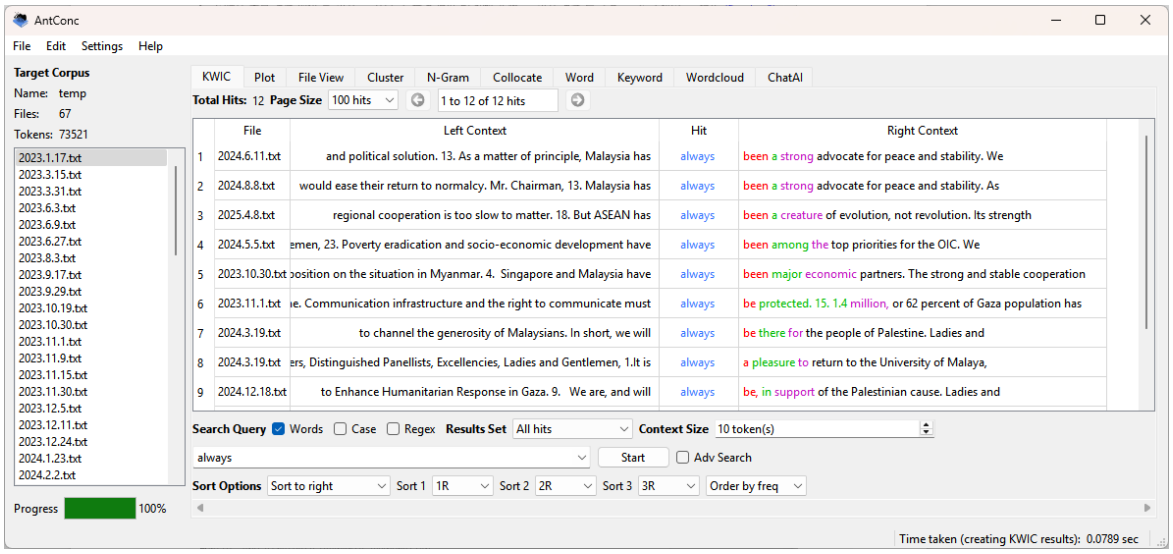

Figure 8. Retrieving English Rounder “always” in AntConc 4.3.1

3.3.4 Significance Testing Tool

In this study, the Log-likelihood Ratio Calculator, free software developed by Prof. Xu Jiajin of the Research Center for Foreign Language Education at BFSU, is used to test for significant differences. Figure 9 below shows the operation interface of the Log-likelihood Ratio Calculator.

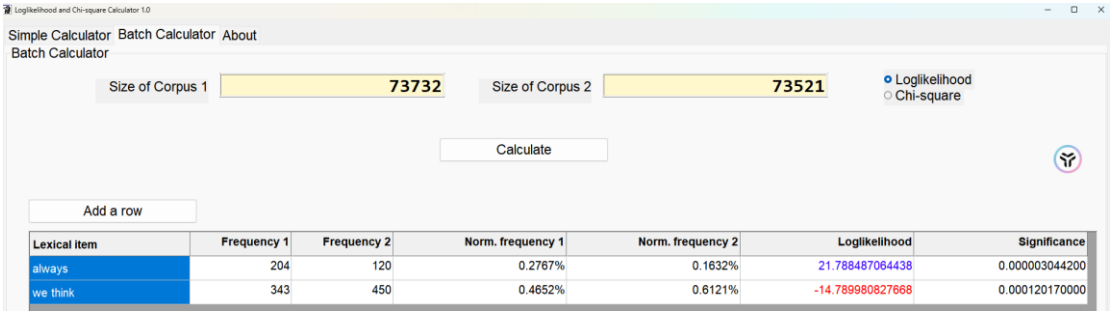

Figure 9. Example of Operation Interface of the Log-likelihood Ratio Calculator

3.4 Data Collection

Two corpora of speeches and statements from the Chinese and Malaysian Ministry of Foreign Affairs have been built, which are:

(1) The corpus of the discourse of speeches and statements from the Chinese Ministry of Foreign Affairs (“Chinese Corpus” for short): 55 speeches and statements from the Ministry of Foreign Affairs of China from the two years of 2023 and 2024, covering about 73,732 words. Source from the website of the Ministry of Foreign Affairs of China: https://www.fmprc.gov.cn/

(2) The corpus of the discourse of speeches and statements from the Malaysian Ministry of Foreign Affairs (“Malaysian Corpus” for short): 67 speeches and statements from the Ministry of Foreign Affairs of Malaysia from the two years of 2023 and 2024, covering about 73,521 words. Source from the website of the Ministry of Foreign Affairs of Malaysia: https://www.kln.gov.my/web/guest/press-releases

4. Results and Discussion

Based on Prince et al.’s (1982) and He’s (1985) categorization, as well as the classification of hedges provided by Varttala (2001) from a cognitive lexical perspective, the existing four types of hedges were further subdivided into ten subcategories, as shown in Figure 10. Next, to test whether the new classification of hedges is feasible for the contrast analysis between Chinese and English discourse, the author combined two self-built corpora of Chinese and English political discourse to conduct retrieval, statistics and analysis of the ten subcategories of hedges.

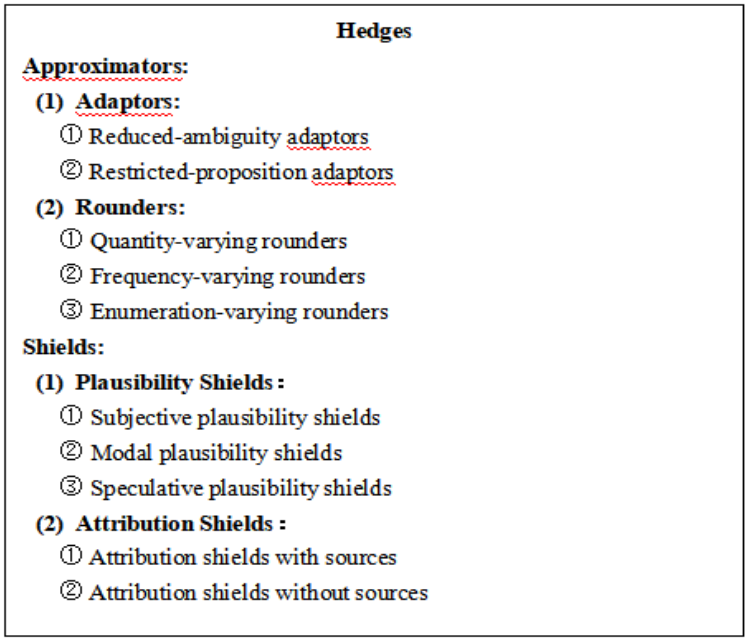

Figure 10. Re-categorization of Hedges

Additionally, according to the research methods of corpus linguistics, to facilitate comparison among corpora of different sizes, the author must carry out "frequency normalization" processing. Based on the size of the corpus and the actual frequency, normalized calculations can be made per thousand words, per ten thousand words, and so on. The data of this article was calculated based on the occurrence frequency per ten thousand words (the original frequency divided by the total number of words in the corpus and then multiplied by ten thousand to obtain the average occurrence frequency per ten thousand words). That is:

$$\text{Normalized frequency of (per ten thousand words)} = \frac{\text{raw frequency}}{\text{the total number of words in the corpus}} \times 10\,000$$

4.1 Three Subcategories of Plausibility Shields

Plausibility shields express the speaker's opinion and attitude toward a particular subject. Plausibility shields can be further subdivided into three types: subjective plausibility shields, modal plausibility shields, and speculative plausibility shields.

(1) Subjective plausibility shields indicate the speaker's attitude toward the proposition, including the "first-person subject + cognitive verb" structure shown in Table 4.

Table 4. Examples of the Structure of "first-person subject + cognitive verb."

English	Chinese
I think, I guess	我认为，我想，我觉得
I suppose	我主张
I consider, I argue	我相信
we agree	我们愿意
we insist	我们坚持

The prepositional phrase reflects the speaker's attitude toward the proposition as well. Table 5 shows some examples.

Table 5. Examples of the Prepositional Phrase

English	Chinese
as far as I can tell, in my opinion	依我看来
as we just said	正如我们之前所说的
from our perspective, in our view	依我们看来
according to my understanding	据我们理解

(2) Modal plausibility shields use modal verbs to convey subjective attitudes, including suggestions, wishes, and possibilities. Table 6 shows some examples.

Table 6. Examples of Modal Plausibility Shields

English	Chinese
may, might	可能
can, could	可以，会
should, ought	应当
must	必须

(3) Speculative plausibility shields refer to verbs, adverbs, and adjectives that express "possibility." Table 7 shows some examples.

Table 7. Examples of Speculative Plausibility Shields

Speculative Plausibility Shields in English	Speculative Plausibility Shields in Chinese
appear to	显得
tend to	倾向于
possibly, probably, likely, perhaps	大抵，大概，可能
seem to	看起来，似乎，好像

To compare the similarities and differences in the use of plausibility shields by spokespersons of the Ministry of Foreign Affairs of China and Malaysia when expressing opinions, the author conducted a statistical analysis of the frequencies of the three subcategories of plausibility shields in the Chinese corpus and the Malaysian corpus. As shown in Table 8 (normalized frequency per 10,000 words), the frequency of the three subcategories of plausibility shields in the Malaysian corpus is lower than that in the Chinese corpus. The author tested whether these differences were significant using the Log-likelihood Ratio Calculator and found that the log-likelihoods of the frequencies of three subcategories of plausibility shields in the two corpora were 19.26, 73.11, and -0.20, respectively (shown in Figure 13). Among them, the significant difference values of subjective plausibility shields and modal plausibility shields are both less than 0.001, showing significant differences between Chinese and Malaysian discourses. The use of modal plausibility shields in the two corpora shows the most significant difference, with the log-likelihood of frequency 73.11, which indicates that the frequency of modal plausibility shields in the Chinese corpus is much higher than that in the Malaysian corpus.

Table 8. Comparison of the Frequency of Three Subcategories of Plausibility Shields in Chinese Corpus and Malaysian Corpus

Three Subcategories of Plausibility Shields	Corpus of speeches and statements of the Ministry of Foreign Affairs of China	Corpus of speeches and statements of the Ministry of Foreign Affairs of Malaysia
Raw frequencies of Subjective Plausibility Shields	313	212
normalized frequency of Subjective Plausibility Shields	42	29
Raw frequencies of Modal Plausibility Shields	509	271
normalized frequency of Modal Plausibility Shields	69	37
Raw frequencies of Speculative Plausibility Shields	2	3
Normalized frequency of Speculative Plausibility Shields	0	0

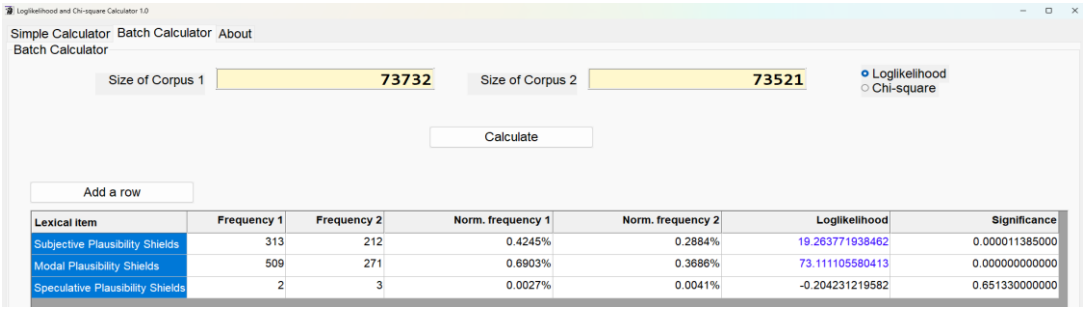

Figure 11. Analysis of Significant Differences in the Frequencies of Three Subcategories of Plausibility Shields in Chinese Corpus and Malaysian Corpus

4.2 Two Subcategories of Attribution Shields

Attribution shields express the speaker's attitude toward something by quoting a third person's opinion. The source of information refers to the person from whom the speaker quoted the words. In speeches issued by the Ministry of Foreign Affairs, when spokespersons quote the remarks, views, or opinions of third parties, they typically need to provide the sources. The author divided attribution shields into two subcategories: attribution shields with sources and attribution shields without sources.

(1) Attribution shields with sources typically follow the "third-person subject + transitive verb" structure. Examples are shown in

Table 9.

Table 9. Examples of the Structure of "third-person subject + transitive verb."

English	Chinese
Mr. Smith said	王菲说
Jane insisted	李峰坚持，小明认为
the president announced	习近平呼吁

The type of third-person paraphrase marked with a prepositional phrase is shown in Table 10.

Table 10. Examples of "third-person paraphrase marked with a prepositional phrase."

English	Chinese
according to Mr. Li	依据李明所说
as Obama puts it	按照奥巴马所说
in Peter's eyes	就王莉来看
from Mary's point of view	从玛丽的角度来看

(2) Attribution shields without sources refer to the paraphrased form where the third-person subject is omitted. Examples are shown in Table 11.

Table 11. Examples of Attribution Shields without Sources

Attribution Shields without Sources in English	Attribution Shields without Sources in Chinese
it is said that it says that	据悉，据说
it is reported that, as reported	有报道称
it is assumed that	据推测
based on	依据
according to the statistics	据统计
according to a survey	据调查

Through corpus analysis, the author found that the frequency of attribution shields used by spokespersons of the Ministry of Foreign Affairs of Malaysia is significantly higher than that used by spokespersons of the Ministry of Foreign Affairs of China. This difference deserves in-depth contrast analysis. Table 12 shows the raw frequencies and normalized frequencies (per 10,000 words) of the two subcategories of attribution shields in the two corpora of the Ministry of Foreign Affairs of China and Malaysia.

In addition, the author compared the following groups of data: a comparison of the frequency of attribution shields with sources in the corpora of the Ministry of Foreign Affairs of China and Malaysia and a comparison of the frequency of attribution shields without sources in the speech corpora of the Ministry of Foreign Affairs of China and Malaysia. A contrast analysis of the differences between these two groups of data is shown in Figure 12. The results show that the frequency likelihood ratio of attribution shields without sources is 0.30, which is higher than 0.001, indicating no significant difference. However, the frequency likelihood ratio of attribution shields with sources is 0.00, less than 0.001, indicating a significant difference between the frequencies of attribution shields with sources in the two corpora.

Table 12. Comparison of the Frequency of Two Subcategories of Attribution Shields in the Two Self-built Corpora of Chinese and Malaysian Political Discourse

Two subcategories of Attribution Shields	Corpus of speeches of the Ministry of Foreign Affairs of China	Corpus of speeches of the Ministry of Foreign Affairs of Malaysia
raw frequencies of Attribution Shields with Source	257	952
normalized frequency of Attribution Shields with Source	35	129
raw frequencies of Attribution Shields without Source	1	3
normalized frequency of Attribution Shields without Source	0	0

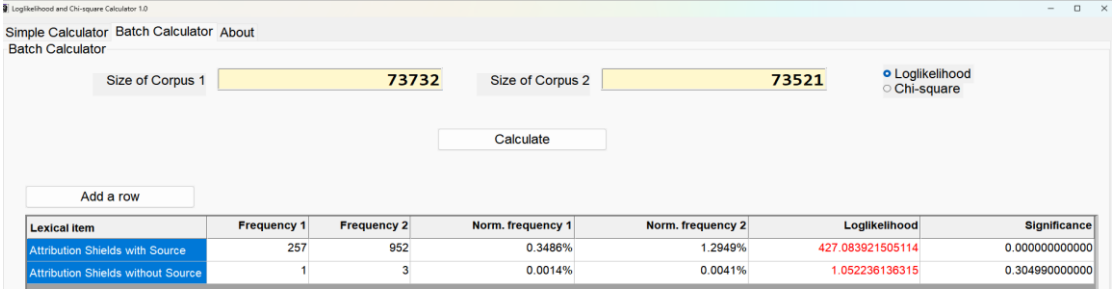

Figure 12. Analysis of Significant Differences in the Frequencies of Two Subcategories of Attribution Shields in Chinese Corpus and Malaysian Corpus

4.3 Two Subcategories of Adaptors

Adaptors indicate semantic differences in degree, providing nuanced modifications to expressions. The author further divided adaptors into reduced-ambiguity adaptors and restricted-proposition adaptors based on their different pragmatic functions.

(1) Reduced-ambiguity adaptors involve directly modifying adjectives or adverbs to clarify membership affiliation and mitigate ambiguity. Table 13 provides examples.

Table 13. Examples of Reduced-ambiguity Adaptors

Reduced-ambiguity Adaptors in English	Reduced-ambiguity Adaptors in Chinese
very, much, pretty, rather, highly	相当，格外，非常，极其，太，过于,很，挺
slightly, sort of, kind of, somewhat, a little bit	稍微，有点儿
almost, nearly	几乎，差一点

(2) Restricted-proposition Adaptors impose specific criteria that limit the scope of a proposition, thereby enhancing its relevance and providing a more nuanced interpretation. Table 14 provides some examples.

Table 14. Examples of Restricted-proposition Adaptors

Restricted-proposition Adaptors in English	Restricted-proposition adaptors in Chinese
broadly speaking	严格来说
simply speaking	泛泛来说
basically	基本上
in a sense	从某种意义上说
to a certain extent	在一定程度上
virtually	事实上

The author conducted statistical and contrast analyses on the frequency of the two subcategories of adapters in the Chinese and Malaysian corpora. The results show that the reduced-ambiguity adaptors were used 349 times in the Chinese Corpus, with a normalized frequency of 47 times per 10,000 words. In contrast, they were used 24 times in the Malaysian Corpus, with a normalized frequency of only 3 times per 10,000 words. Restricted-proposition adaptors were used only once in the Chinese Corpus, with a normalized frequency of 0 per 10,000 words. They are used 60 times in the Malaysian Corpus, with a normalized frequency of 8 per 10,000 words, as shown in Table 15.

The log-likelihood ratio of the raw frequencies of reduced-ambiguity adaptors in the self-built corpora is 338.04, and the log-likelihood ratio of the raw frequencies of restricted-proposition adaptors is 74.52. The significant difference values between the two groups are less than 0.001, indicating significant differences in the use of reduced-ambiguity adaptors and restricted-proposition adaptors between China's and Malaysia's political discourse. Moreover, 338.04 is much higher than 74.52. Therefore, the frequency difference in reduced-ambiguity adaptors between Chinese and Malaysian political discourse is significant.

Table 15. Comparison of the Frequency of Two Subcategories of Adaptors in the Two Self-built Corpus of Chinese and Malaysian Political Discourse

Two subcategories of Adaptors	Corpus of speeches of the Ministry of Foreign Affairs of China	Corpus of speeches of the Ministry of Foreign Affairs of Malaysia
Raw frequencies of Reduced-ambiguity Adaptors	349	24
Normalized frequency of Reduced-ambiguity Adaptors	47	3
Raw frequencies of Restricted-proposition Adaptors	1	60
Normalized frequency of Restricted-proposition Adaptors	0	8

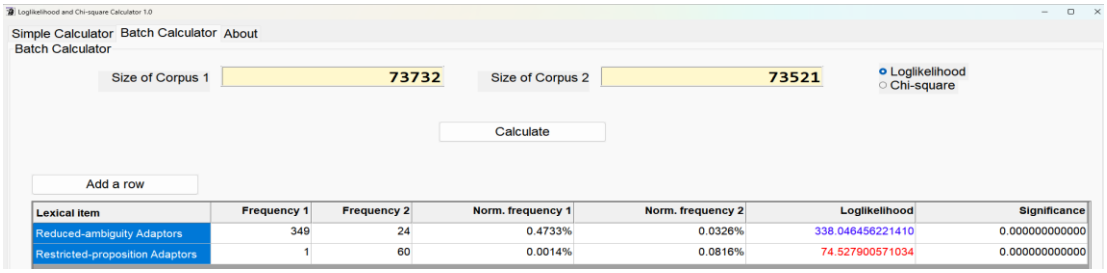

Figure 13. Analysis of Significant Differences in the Frequencies of Two Subcategories of Adaptors in Chinese Corpus and Malaysian Corpus

4.3 Three Subcategories of Rounders

Rounders limit the range of variation in concepts such as quantity and frequency. The author further divides rounders into three

subcategories: quantity-variating rounders, frequency-varying rounders, and enumeration-varying rounders.

(1) Quantity-varying rounders can be further subdivided into numerical and non-numerical types. Numerical type: By placing these terms before or after a specific quantity, the exact value becomes vague, but the estimation of the quantity can vary depending on the individual and the context. Examples of these can be found in Table 16.

Non-numerical type: These terms are used to make the modified word ambiguous in terms of quantity or reference. Examples of these can be seen in Table 17.

(2) Frequency-varying rounders are used to make the frequency with which an action occurs vague. Examples of these can be found in Table 18.

(3) Enumeration-varying rounders are used for incomplete enumeration. Examples are illustrated in Table 19.

Table 16. Examples of Numerical Types of Quantity-varying Rounders

Numerical Type of Quantity-varying Rounders in English	Numerical Type of Quantity-varying Rounders in Chinese
about 5 years	**大约** 5 岁
around 3 o'clock	3 点**左右**
nearly 30 meters	**将近** 30 米
approximately 70%	**约** 70%

Table 17. Examples of Non-numerical Type of Quantity-varying Rounders

Non-numerical Type of Quantity-varying Rounders in English	Non-numerical Type of Quantity-varying Rounders in Chinese
Many, lots of, much, a number of,	很多，大量，许多
most	绝大多
some, several, a little, a few	有一些,
few, little	少量
quite a bit	相当多
thousands of	成千上百

Table 18. Examples of Frequency-varying Rounders

Frequency-varying Rounders in English	Frequency-varying Rounders in Chinese
usually, often	时常
sometimes	有时
always	一直,总是
seldom, rarely	很少，不常
never	从未
frequently	经常

Table 19. Examples of Enumeration-varying Rounders

Enumeration-varying Rounders in English	Enumeration-varying Rounders in Chinese
anything like that, or something else,	诸如此类, 之类的，什么的
and things like that	
and so on	等, 等等

Through contrast analysis of the frequency of the three subcategories of rounders, the results showed that the quantity-varying rounders were used 167 times in the Chinese Corpus, with a normalized frequency of 23 times per 10,000 words; they were used 102 times in the Malaysian Corpus, with a normalized frequency of 14 times per 10,000 words. Frequency-varying rounders were used 392 times in the Chinese corpus, with a normalized frequency of 53 per 10,000 words. In contrast, they were used only 44 times in the Malaysian corpus, with a normalized frequency of 6 per 10,000 words. Enumeration-varying rounders were used 169 times in the Chinese corpus, with a normalized frequency of 23 per 10,000 words; they are used only once in the Malaysian corpus, with a normalized frequency of 0 per 10,000 words, as shown in Table 20.

Through the Log-likelihood Rate calculator tests, the log-likelihood ratio of the raw frequencies of quantity-varying rounders, frequency-varying rounders, and enumeration-varying rounders in the self-built corpus are 15.67, 318.20, and 222.92 respectively. The significant difference values of the three subcategories of rounders are all less than 0.001, indicating significant differences between China and Malaysia in all of them. Moreover, 338.04 and 222.92 are much higher than 15.67. Therefore, the frequency differences in frequency-varying rounders and enumeration-varying rounders between Chinese and Malaysian political discourses are significantly different.

Table 20. Comparison of the Frequency of Three Subcategories of Rounders in Chinese Corpus and Malaysian Corpus

Three subcategories of Rounders	Corpus of speeches of the Ministry of Foreign Affairs of China	Corpus of speeches of the Ministry of Foreign Affairs of Malaysia
Raw frequencies of Quantity-varying Rounders	167	102
Normalized frequency of Quantity-varying Rounders	23	14
Raw frequencies of Frequency-varying Rounders	392	44
Normalized frequency of Frequency-varying Rounders	53	6
Raw frequencies of Enumeration-varying Rounders	169	1
Normalized frequency of Enumeration-varying Rounders	23	0

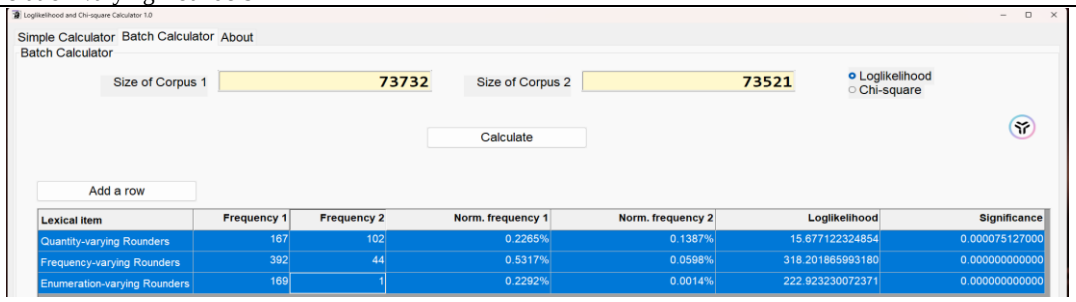

Figure 14. Analysis of Significant Differences in the Frequencies of Three Subcategories of Rounders in Chinese Corpus and Malaysian Corpus

5. Conclusion

The findings suggest that among the ten subcategories, eight show significant differences in frequency between Chinese and Malaysian political discourse, specifically subjective plausibility shields, modal plausibility shields, attribution shields with source, reduced-ambiguity adaptors, restricted-proposition adaptors, quantity-varying rounders, frequency-varying rounders, and enumeration-varying rounders.

The results provide evidence that the new categories of Chinese and English hedges proposed under the frameworks of the classifications of Prince et al. (1982) and He (1985) are highly operable when using corpus linguistics to analyze Chinese and English discourse and can provide a new perspective for future contrast analysis between Chinese and English discourse. For example, when conducting a critical discourse analysis of political discourse, the author can further analyze the pragmatic strategies and ideological differences between spokespersons of the Ministry of Foreign Affairs of China and Malaysia in the construction of pragmatic identities by contrastive analysis of the differences in the choice of personal subject types when using subjective plausibility shields. In addition, the author can analyze the emotional differences between the spokespersons of the Ministry of Foreign Affairs of China and Malaysia on specific news events by counting the differences in the magnitude of modal verbs when choosing modal plausibility shields.

In short, establishing a new categorization suitable for Chinese and English hedges provides a promising method for future contrast analysis of Chinese and English discourses. The new categorization plays a crucial role in revealing cultural values, diplomatic concepts, and national positions reflected in the differences in the use of hedges in political discourse through critical discourse analysis. Future scholars should pay more attention to the pragmatic strategies and ideological differences reflected in the use of each subcategory of hedges.

Acknowledgments

Not applicable.

Authors' contributions

The authors contributed equally to the study.

Funding

Not applicable.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

Azizah, D. (2021). Hedges function in masculine and feminine feature's language: A pragmatics analysis. *Journal of Pragmatics Research, 3,* 59-69. https://doi.org/10.18326/jopr.v3i1.69-69.

Biber, D. (2009). *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, 1-34.

Brown, P., & Levinson, S. (1987). *Politeness: Some Universals in Language Usage.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511813085

Channel, J. M. (1994). *Vague Language*. Oxford: Oxford University Press, 214-218.

Clemen, G. (1997). The concept of hedging. In R. Markkanen & H. Schroder (Eds.), *Hedging and Discourse*. Walter de Gruyter, 235-248. https://doi.org/10.1515/9783110807332.235

Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics* (4th ed.). Oxford: Blackwell Publishers, 227. https://doi.org/10.1002/9781444302776

Fraser, B. (2010). Pragmatic competence: The case of hedging. *Studies in Pragmatics, 9*, 15-34. https://doi.org/10.1163/9789004253247_003

He, Z. (1985). Hedges and speech communication. *Foreign Languages*, *5*, 27-31.

Holmes, J. (1982). Expressing doubt and certainty in English. *RELC Journal*, *2*, 9-29. https://doi.org/10.1177/003368282013002O2

Holmes, J. (1984). Modifying illocutionary force. *Journal of Pragmatics, 8*, 345-365. https://doi.org/10.1016/0378-2166(84)90028-6

House, J., & Kasper, G. (1981). Politeness markers in English and German. In F. Coulmas (Ed.), *Conversational Routine*. Mouton, 157-185. https://doi.org/10.1515/9783110809145.157

Hüler A. (1983). *Understatements and Hedges in English (Pragmatics and Beyond IV: 6).* John Benjamins. https://doi.org/10.1075/pb.iv.6

Hyland, K. (1996). Writing without conviction? Hedging in science research articles. *Applied Linguistics, 4,* 433-453. https://doi.org/10.1093/applin/17.4.433

Hyland, K. (1998). *Hedging in Scientific Research Articles*. John Benjamins Publishing Company, 163-187. https://doi.org/10.1075/pbns.54

Kopple, W. (1985). Some exploratory discourse on meta-discourse. *College Composition and Communication, 1*, 82-93. https://doi.org/10.58680/ccc198511781

Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy Concepts. In P. Peranteau et al. (Eds.), *Papers from the English Regional Meeting*. Chicago Linguistic Society. Retrieved from https://link.springer.com/article/10.1007/BF00262952

Liu, J. (2020). A pragmatic analysis of hedges from the perspective of politeness principle. *Theory and Practice in Language Studies, 12*, 1614-1619. http://doi.org/10.17507/tpls.1012.15

Meyer, G. (1989). The pragmatics of politeness in scientific articles. *Applied Linguistics, 10*, 1-35. https://doi.org/10.1093/applin/10.1.1

Millan, E. L. (2008). Epistemic and approximative meaning revised: using hedges,boosters, and approximators when writing research in different disciplines. In S. Burgess & P. Martin- Martin (eds.) *English as an Additional Language in Research Publication and Communication*. Peter Lang, Bern, Germany, 65-82.

Prince, E. F., Frader, J., & Bosk, C. (1982). On hedging in physician-physician discourse. In R. J. Pietro (ed.). *Linguistics and professions*. Norwood, 83-97.

Rosch, E. H. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and Categorization*. Erlbaum and Associates, 27-48. https://doi.org/10.4324/9781032633275-4

Salager-Meyer. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purpose, 2*, 149-170. https://doi.org/10.1016/0889-4906(94)90013-2

Su, Y. (2002). Classification and function of English and Chinese hedges, *Journal of Guangzhou University (Social Science Edition), 4*, 29-32.

Varttala, T. (2001). *Hedging in Scientifically Oriented Discourse: Exploring Variation According to Disciplines and Intended Audience*. The University of Tampere.

Verschueren, J. (1999). *Understanding Pragmatics*. Arnold, 193.

Wu, T. (1999). *Fuzzy Linguistics*. Shanghai Foreign Language Teaching Press.

Yule, G. (1996). *Pragmatics*. Oxford University Press.

Yule, G. (2010). *The Study of Language* (4th ed.). Cambridge University Press, 148.

Zadeh, L. A. (1972). A fuzzy-set-theoretic interpretation of linguistic hedges. *Journal of Cybernetics, 2*, 4-34. https://doi.org/10.1080/01969727208542910

Zuck, J. G., & Zuck, L. V. (1985). Hedging in news writing. In A. M. Comu et al. (Eds.), *Beads or bracelets? How do we Approach LSP? Selected Papers From the 5th European Symposium on LSP*. Oxford University Press, 172-180.