

Article

The Differential Impact of Data Collection Methods and Language Background on English Tone Choice Patterns

Kevin Hirschi ^{1,*}  and Maria Kostromitina ²

¹ Department of Bicultural-Bilingual Studies, The University of Texas at San Antonio, San Antonio, TX 78249, USA

² Duolingo, Inc., Pittsburgh, PA 15206, USA

* Correspondence: kevin.hirschi@utsa.edu

Abstract

This study examines the impact of spoken data collection techniques and language background on falling, level, and rising tones. Elicited data from a Discourse Completion Task (DCT), structured speech from a collaborative oral assessment task, and naturalistic speech from a comprehensive corpus of inner-circle and Hong Kong English were analyzed for Discourse Intonation features, resulting in 2756 tone choices by 184 speakers. Multinomial logistic regression indicates that structured speech by L2 English learners and naturalistic speech by both inner circle and Hong Kong English speakers exhibited similar tone choice patterns. However, DCT responses by L2 English learners contained significantly fewer level tones and more rising tones. Qualitative analyses suggest that contrary to naturalistic studies, L2 learners use rising tones to focus their attention on the speaker during a request. L1 users, on the other hand, used a variety of tone choices that focus on language and mitigate directness. Overall, these results add further evidence that DCTs do not elicit speech that generalizes to naturalistic discourse. Structured tasks in which two L2 speakers interact mirror the rates of the inner circle and Hong Kong English speakers detected in this study.



Academic Editors: Elena Babatsouli and Erik R. Thomas

Received: 12 March 2025

Revised: 20 July 2025

Accepted: 5 August 2025

Published: 15 August 2025

Citation: Hirschi, K., & Kostromitina, M. (2025). The Differential Impact of Data Collection Methods and Language Background on English Tone Choice Patterns. *Languages*, 10(8), 193. <https://doi.org/10.3390/languages10080193>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: intonation; pragmatics

1. The Differential Impact of Data Collection Methods and Language Background on English Tone Choice Patterns

Research at the intersection of English pragmatics and pronunciation has long agreed that suprasegmental features of speech such as intonation constitute important discourse functions that are critical for communication. Intonation plays a fundamental role in spoken English discourse, helping to convey speaker intentions and status, define interlocutors' relationships, and express politeness (Brazil, 1997; P. Brown & Levinson, 1987; Pickering, 2018). While these functions of intonation may come naturally to first language (L1) English speakers, there exists great variation amongst learners and speakers of global Englishes in terms of different intonation patterns (Björkman, 2012; Pickering, 2009; Pickering & Wiltshire, 2000). Global varieties, including those in Kachru's (1992) outer circle, may establish their own norms and inter-varietal interaction may result in negative pragmatic consequences when listeners' expectations of intonation patterns are not met (Kostromitina, 2024; Pickering, 2009).

Studies have also found that intonation patterns in discourse may vary depending on register differences (Biber & Staples, 2014; Cheng & Warren, 2008) as well as the context in

which speech data is collected. While speech elicitation techniques such as Discourse Completion Tasks (DCTs) have been commonly employed for spoken research in pragmatics, this type of data may have limited generalizability because of the inauthentic discourse context. In other words, it may not be representative of intonation patterns displayed in real-world interactions that occur in globalized contexts. At the same time, collecting large, representative samples of naturalistic data for intonation analysis is time consuming and logistically challenging (Cheng et al., 2005).

Given the potential consequences of variation in intonation and the paucity of research that examines the impact of data collection technique on English discourse prosody amongst language users from different backgrounds, the current study sets to examine how one important feature of intonation—tone choices—varies based on speakers' language backgrounds and data collection methods. Specifically, it investigates the differences between elicited, structured (i.e., collaborative responses to a task), and naturally produced data across L1 and L2 English intonation in North America and Chinese speakers of English.

2. Discourse Functions of English Intonation

Intonation functions outlined in Chun (2002) include grammatical, attitudinal, indexical, discourse, and interactional functions (see also Levis & Wichmann, 2015). However, Brazil's (1997) Discourse Intonation framework suggests that intonation bears several functions in English discourse that aim at achieving communicative functions. In discussing intonation functions in discourse, Pickering (2018) suggested that intonation is pragmatic in nature—a point that is supported by earlier inquiries into intonation functions in speech acts (e.g., G. Brown et al., 2015). Brazil's Discourse Intonation is rooted in discourse analysis of English carried out in the UK, and has substantial evidence of applicability in other inner-circle English varieties (Pickering, 2018). In contrast to more widespread autosegmental-metrical (AM) approaches to intonation analysis (Ladd, 2008), Brazil's (1997) Discourse Intonation is specific to English, focused on pragmatic meaning rather than phonological representation, and considers both acoustic and discourse dimension when conducting analyses. A growing body of Discourse Intonation studies has demonstrated its ability to provide meaningful insights in qualitative analyses (e.g., Pickering, 2001, 2009), corpus analyses (e.g., Cheng & Warren, 2005, 2008; Cheng et al., 2005, 2008; Staples, 2015), and experimental/longitudinal data (e.g., Kostromitina, 2024; Kostromitina & Miao, 2024; Taguchi et al., 2021; Kang et al., 2021) produced by L1 speakers, L2 English learners, and certain outer circle varieties of English (e.g., Hong Kong English in Cheng & Warren, 2008, and related work).

Within Discourse Intonation, Pickering (2018) explained that 'common ground' (i.e., the shared knowledge or worldviews between interlocutors) affects speakers' choice of intonation patterns. Speakers make presumptions about listeners' existing knowledge during an interaction and adjust their tone choices accordingly, particularly on the final prominent syllable of a tone unit and especially at the end of a turn. Thus, rising (or referring) intonation is used when the speaker believes the listener already shares contextual knowledge; falling (or proclaiming) intonation is utilized when the speaker aims to present new information to expand this shared understanding. According to Brazil's (1997) model, fall and rise-fall tones serve multiple purposes, including a speaker's dominance in the local discourse of the conversation. This sense of dominance is defined narrowly, relating to how much they direct the flow of information, and not necessarily their socially imbued power. Rather, it describes a speaker's desire to share new information in a conversation or establish their authority. It can also do both at the same time. Rise and fall-rise tones, on the other hand, are used to seek solidarity with the listener and to indicate that they are

focused on what the listener has said and will say (i.e., passing the dominance along to the listener). Level tones, besides being used in combination with generic language routines, indicate lack of focus on the hearer or on the communicative message as well as speaker neutrality. In contrast to AM analyses, these tone choices do not neatly align with syllable stress or the end of a phrase in all cases, but instead rely on the combination of discourse and acoustic analyses. Furthermore, [Brazil's \(1997\)](#) level tone does not have a systematic equivalent in AM coding practices.

Several studies have demonstrated the preference for speakers to use different tones with different interlocutors, supporting the discourse-based functions within [Brazil's \(1997\)](#) Discourse Intonation framework. For example, [Cheng et al. \(2008, pp. 156–158\)](#) found more rising tones in interactions with supervisors (about 85% of tones were rising) than in informal office talk (about 67% of tones were rising) or in regular conversation (about 50% of tones were rising tones). They posit that the use of rising tones was likely to establish dominance in the discourse by the academic advisor, which was lessened in the other types of interactions. A precursor study on the same corpus found that rising tones were commonly used by the interlocutor with higher power status, but were also used by the other speakers in power imbalanced contexts ([Warren, 2004](#)).

3. L1 vs. L2 English Intonation Use

Contrastive studies have found that intonation use may vary between L1, L2, and global variety speakers of English. For learners, this may have to do with the relative difficulty of L2 intonation acquisition compared to other aspects of pronunciation ([Pickering, 2018](#)), as well as with the differences in how the L2 phonological system develops compared to the L1 system ([Ramírez-Verdugo et al., 2017](#)). Varying use of intonation between L1, L2, and global variety English speakers has been found across different pragmatic speech acts, including empathy in nurse-patient interactions ([Staples, 2015](#)), apologies ([Aijmer, 1997](#)), and refusals ([Kostromitina & Miao, 2024](#)). In [Hewings \(1995\)](#), L1 speakers of English mostly used rising and level tones in disagreements while L2 speakers resorted to falling tone choices. Hewings suggested that L1 speakers used this referring intonation to promote social integration with interlocutors. [Staples \(2015\)](#) found that L1 and internationally educated nurses, largely from the Philippines, differed in their use of tone choices to express sympathy towards their patients. In fact, international nurses in the study did not have a clear preference for a particular tone choice in this speech act while L1 nurses employed significantly more falling tones.

In a corpus-based study, [Ramírez-Verdugo \(2008\)](#) compared the use of tone choices and pitch accents in commands and requests produced by L1 and L2 English speakers who were teachers in Spain. Similarly to findings of previously mentioned studies, this examination also revealed that L1 speakers employed a wider variety of tones, including rising intonation, possibly to orient conversation towards classroom collaboration. In contrast, L2 speakers realized speech acts with mostly falling tones that implied ordering and commanding rather than encouraging. Even across L2 English speakers from various backgrounds, studies have found proficiency effects in the use of intonation with more proficient speakers' tone choice patterns resembling those of L1 speakers more closely ([Kang et al., 2021](#); [Taguchi et al., 2021](#)).

Research that focuses on different varieties of English have found great variation amongst tone choices. Examining Malaysian and Singaporean English speakers, [Goh \(2000\)](#) found that rising tones were rare but were particularly used when repeating information presented by a different interlocutor. In an investigation of International Teaching Assistant (ITA) speech of L1 Mandarin speakers in the North American context, [Pickering \(2001\)](#) found that speakers demonstrated a strong preference for falling tones, with several partici-

pants using very few rising tones and fewer than a quarter of the tones were level tones. A qualitative analysis indicated that all speakers used rising tones to remind listeners of common information, build rapport, and politely correct their students. [Cheng et al. \(2008\)](#) investigated tone choice amongst L1 English and Hong Kong English speakers, finding markedly similar frequencies of tone choice. However, there were slightly more level tones amongst Hong Kong English (HKE) speakers compared to L1 speakers (49.31% to 38.15%, respectively). Their qualitative analyses determined these differences due to hesitations and other “encoding difficulties” amongst HKE speakers ([Cheng et al., 2008](#), p. 141). Taken together, these studies provide evidence of the importance of intonation when L2 speakers interact with listeners of other varieties, and also highlight the value of examining intonation in actual language use through corpus linguistic techniques that collect data in the classroom or in conversation.

4. Effects of Speech Elicitation Instruments on Intonation Patterns Variation

In pragma-prosodic research, there exists a general divide between naturalistic and elicited data and its generalizability ([Kang et al., 2021](#); [Taguchi & Roever, 2017](#)). While naturalistic data is considered more representative of real-world spoken discourse, it is much harder to collect compared to elicited data which may offer more controlled albeit potentially less representative findings. Given this divide, scholars have attempted to find speech elicitation methods that are valid, convenient, and reasonably authentic. Discourse Completion Tasks (DCTs) have been a common, yet critiqued, method of data collection since they are adaptable from the written to the spoken modality and can be used to prompt elicitation of a speaker’s response. However, several researchers have argued that this method of data elicitation for speech analysis is limited as it does not reflect sequential organization of the conversation ([Golato, 2003](#); [Ikeda, 2017](#)).

As an alternative to DCTs, extended discourse tasks such as role plays and collaborative tasks have been increasingly used in a variety of speech studies. We refer to this type of data collection as *structured* speech because it contains linguistic and topic support through a prompt that includes a specific goal. Such structured tasks have an advantage because they can elicit a broader range of pragmatic features and are capable of eliciting longer speech samples and demonstrate sequential organization in spoken negotiations ([Félix-Brasdefer, 2004](#); [Ikeda, 2017](#)). Attempting to understand the validity of DCTs and structured tasks as approximations of natural speech, several studies ([Félix-Brasdefer, 2007](#); [Yuan, 2001](#)) examined data elicited with DCTs and structured tasks, such as role plays, comparing it with natural data comprising field notes and natural conversations. The results of these studies showed that structured speech data may be representative to a degree, although the repertoire of pragmatic strategies demonstrated in such data may be limited. Nevertheless, other studies have noted that role plays specifically may help to generate features that are observed in naturalistic spoken discourse including (re)negotiation of meaning, turn taking, and indirectness strategies ([Félix-Brasdefer, 2005](#)).

Crucially for intonational analysis in pragmatic research, [Kang et al. \(2021\)](#) also acknowledged the non-interactive nature of DCTs and their inability to provide longer speech samples needed for a more comprehensive analysis, especially in low imposition speech acts, suggesting that intonation patterns may be different in data collected with a dialogic task. Despite the concerns about data types expressed repeatedly by scholars studying pragmatic functions of intonation, there exists little evidence clarifying differences in representativeness of naturalistic and elicited discourse. Among a handful of studies comparing intonation patterns across types of spoken data, [Ramírez-Verdugo \(2008\)](#) examined tone choices of L1 and L2 English teachers in natural classroom discourse and scripted dialogues.

The study did not directly compare these two data sources but found that L2 teachers used mostly falling tone choices when giving instructions to their students.

5. Present Study

Existing intonation research at the interface of global Englishes, English learners, prosody and pragmatics has exemplified several methodological and conceptual gaps. First, related to instrument validity, it is unclear how the types of data used to study intonation (i.e., naturalistic, structured, and elicited speech) affect tone choice patterns. Second, while studies of intonation composition of discourse are emerging, we do not yet have a comprehensive representation of how tone choices are applied by L1, L2, and global English speakers using different data collection methods. To this end, the current study seeks to address this gap with the following research questions:

RQ1: To what extent do tone choice pattern preferences vary in spoken data depending on data collection method (i.e., elicited, structured, or naturalistic) and speaker language background?

RQ2: Which tone choices are used differently across data collection methods or speaker language backgrounds?

6. Materials and Methods

6.1. Corpora

Speech data were sourced from three datasets: (a) L1 Chinese speakers of English from the Corpus of Collaborative Oral Tasks (CCOT, Crawford & McDonough, 2021), (b) L1 English and Hong Kong English (HKE) speakers in the Hong Kong Corpus of Spoken English (Cheng et al., 2008), and (c) L1 English and L2 Chinese speakers in a DCT dataset collected by Kostromitina (2024). Parity amongst the participants and spoken data was sought in sampling from the datasets in terms of participant background (i.e., all were L1 Chinese) and all samples were in the academic context. However, it should be noted that Kostromitina (2024) and the CCOT primarily include L1 Mandarin speakers whereas the HKCSE was primarily collected in a Cantonese-speaking context. Furthermore, in Kostromitina (2024), participants were recruited in mainland China to complete an intervention online on an instructional platform at the high-intermediate/advanced proficiency level. In the CCOT, L1 Mandarin interlocutors were in the United States and had been tested below the threshold for university-level work to be enrolled in the language program, which was generally TOEFL IBT 70 or IELTS 6.0. Little information is provided in the HKCSE documentation about the language background of the participants, but only those who were marked as either L1 English or L1 Hong Kong Chinese English speakers were included in this study.

The three datasets represent a variety of data collection techniques that span from a specific speech act (i.e., requests) in Kostromitina (2024), to structured speech captured in collaborative oral assessment tasks in the CCOT, as documented in Crawford and McDonough (2021). In the HKCSE, however, prosodic codes were sampled from naturalistic recordings of a variety of academic conversations, including presentations by students, lectures, seminars, workshops and tutoring (See Table 1.2 in Cheng et al., 2008 for a breakdown of communicative events in the academic sub-corpus). See Table 1.

Table 1. Description of corpora.

Name	Description	Sample Used
Speech Act Experiment (Kostromitina, 2024)	Recordings of multi-turn video-based speech acts produced by L1 and L2 English undergraduate students	28 recordings of 34 L1 and 14 L2 English speakers completing DCTs, 1185 tone choices
Corpus of Collaborative Oral Tasks (CCOT) (Crawford & McDonough, 2021)	Intensive English Program students from various L1 backgrounds completing a collaborative task in pairs	62 recordings of 104 L1 Chinese learners, of English, 412 tone choices
Hong Kong Corpus of Spoken English (HKCSE) (Cheng et al., 2008)	Naturalistic recordings of L1 and HKE speakers in academic and business contexts	11 recording of 32 L1 and HKE speakers, 1159 tone choices sampled randomly

6.2. Tone Choice Coding

Analysis of tone choices was completed according to Brazil's (1997) description of termination prominence, which includes an instrumental and perceptual analysis of the pitch movement on final prominent syllable in a tone unit as it is prosodically contextualized within the utterance. Following the perceptual and instrumental analysis of Pickering (2001) and Kang et al. (2021), tone choice coding was assisted by Praat visualization of the fundamental frequency in limited cases of overlap and uncertainty for the CCOT and DCT data (Boersma & Weenink, 2020). In order to validate coding of tone choices, several steps were taken. For the DCT data (Kostromitina, 2024), a second coder was employed for 5% of the tone choices, resulting in an agreement rate of 92% during a first round of coding. For the CCOT dataset, the tone choice codes were examined by a second listener, resulting in 87% agreement. In both cases, agreement rates reached 100% after additional discussions. The HKCSE was coded by trained transcribers as described in Chapter 3 in Cheng et al. (2008).

The focus of the tone choice coding in the present study was limited to one tone choice per turn to focus on any effect of the short discourse elicited from the DCT in Kostromitina (2024). To approximate this focus in structured and naturalistic data, which in many cases had longer turns, tone choices in the utterance-final position prior to an interlocutor's turn were extracted in the CCOT and HKCSE. More than 11,000 tone choices were detected in the HKCSE academic subcorpus, so a tone choice sampling approach was taken in order to have relatively similar group sizes amongst the datasets. To accomplish a balanced sample, we created a Python script that extracts every seventh HKE speaker's tone choice but retains all L1 speaker tone choices, resulting in 1159 tone choices (634 HKE tone choices and 525 L1 tone choices). For interactions with overlap or incomplete tone units, the final complete tone unit was coded only if the interlocutor addressed this information.

6.3. Data Analysis

Tone choices from transcripts were tabulated and combined into a single data frame which contained the final tone choice of the conversational turn and other variables such as the data source, the conversation ID, and speaker language background. The data source was recoded into speech type so that tone choice observations from the DCT data were coded as *elicited*, those from the CCOT were coded as *structured*, and those from the HKCSE were coded as *naturalistic*. Tone choices were reduced from Brazil's five-level tone choices (rise, fall, level, rise-fall, fall-rise) to three levels (rise-fall was recoded as fall and fall-rise was recoded as rise) because of the rare cases of the compound tones. According to Brazil's model, compound tones indicate speaker dominance, where fall-rise is the non-dominant counterpart to the dominant rise tone, which is a referring tone. Rise-fall,

however, is the dominant proclaiming tone, whereas fall is the non-dominant form. The removal of dominance plays little importance to the pragmatic assumptions of the tones (Pickering, 2018).

For statistical modeling, the outcome variable was set as the simplified tone choice (fall, rise, level), with predictor variables of speech type and language background, and an interaction term. Statistical modeling was carried out using multinomial logistic regression using the *nnet* package (Ripley & Venables, 2016), which models the chances of observing the presence or absence of an outcome as influenced by predictors. The multinomial approach generalizes the more common binomial (i.e., 1 for presence and 0 for absence) to a nominal variable with three outcome possibilities. Similar to linear regression, predictors can be examined independently or with an interaction term that must be specified manually. The resulting *multinom* model was:

tonechoice ~ elicitedness + language_background + elicitedness: language_background.

Post hoc tests were conducted with the *contrasts* function in the *emmeans* package with a multivariate t distribution adjustment for multiple comparisons in logistic regression.

7. Results

The first research question investigated the extent to which tone choice patterns vary in speech depending on the data collection method (i.e., elicited or naturalistic). The tone choices were tabulated, and the results indicate a divergent path for L1 speakers as compared to L2 speakers. In all subsets, falling tones were the most frequent. There were 344 falling tones (53.00%) and 536 falling tones (50.19%) amongst L1 and L2 speakers, respectively, in the elicited (DCT) data. Rising tones were less frequent amongst L1 users for elicited (DCT) data, there were 99 rising tones (15.25%) but for L2 speakers there were 248 rising tones (46.27%). Level tones were used more by L1 speakers; there were 206 (31.74%), but only 19 (4.35%) amongst L2 speakers.

There were no L1 speakers in the structured dataset, but L2 speakers also used falling tones 249 times (60.44%), rising tones 57 times (13.83%), and level tones 106 times (25.73%). In the naturalistic data from the CCOT, there were 312 falling tones (49.21%) for L1 speakers and 317 falling tones (60.38%) for L2 speakers. For rising tones, there were 102 rising tones (16.09%) amongst L1 speakers and 70 (13.33%) produced by L2 speakers. Finally, in naturalistic speech, there were 220 level tones (34.70%) for L1 speakers and 138 (26.29%) for HKE speakers. See Table 2 for complete results.

Table 2. Tone choice frequencies and percentages by speech type and language background.

Dataset	Lang. Bkgr.	Fall		Rise		Level		Total	
		freq.	%	freq.	%	freq.	%	freq.	%
Elicited (DCT)	L1	344	53.00%	99	15.25%	206	31.74%	649	100%
	L2	269	50.19%	248	46.27%	19	3.54%	536	100%
Structured (Collaborative Task)	L1	-	-	-	-	-	-	-	-
	L2	249	60.44%	57	13.83%	106	25.73%	412	100%
Naturalistic (Corpus)	L1	312	49.21%	102	16.09%	220	34.70%	634	100%
	HKE	317	60.38%	70	13.33%	138	26.29%	525	100%

The frequencies and percentages exhibit remarkable similarities in falling tone choice preferences amongst both language backgrounds and datasets, ranging from 49.21% to 60.38%. However, the distribution of rising and level tones was similar amongst all variables except for L2 speakers in the elicited (DCT) dataset. For both L1 and L2 speakers in the

naturalistic and structured datasets, rising tones were used between 13.33% and 16.09% of the time, and level tones were employed between 25.73% and 34.70%. However, L2 speakers in the elicited (DCT) dataset used more rising tones (46.27%) and very few level tones (3.54%). To further understand the divergent pattern, the results were visualized in stacked proportional bar plots in Figure 1.

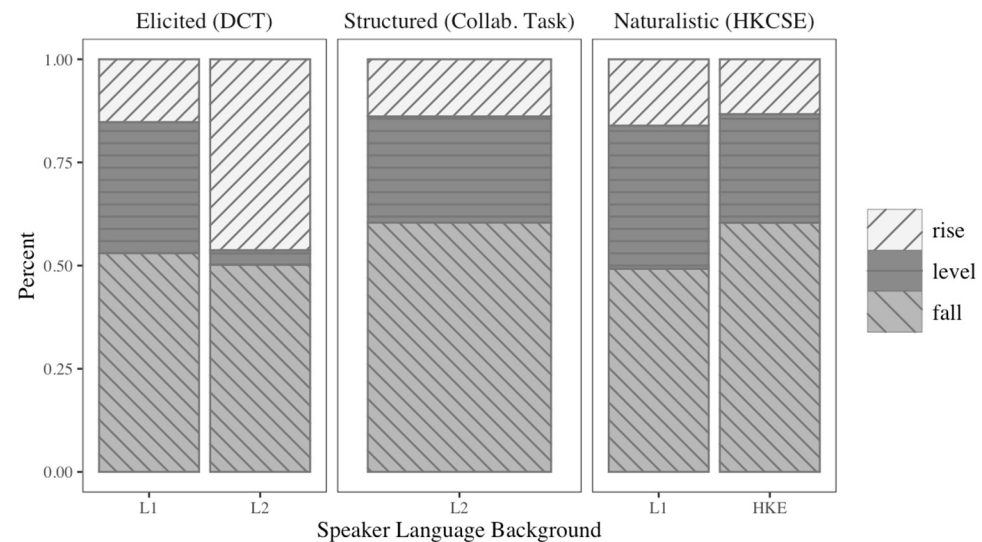


Figure 1. Tone choice by type of speech and language background.

The multinomial logistic regression was used to examine the predictive relationship between the language background (L1, L2/HKE) and data collection type (elicited, structured, or naturalistic) and tone choices. The summary results indicated significant differences between the reference category (rise L1 elicited) with level (0.73, SE = 0.12, $p < 0.001$) and level L2/HKE (−3.30, SE = 0.27, $p < 0.001$), but not with level tones in naturalistic data (0.04, SE = 0.17, $p = 0.834$). However, the main effects are relatively uninterpretable for multinomial logistic regression as our investigation is concerned with comparing tones across predictors. Interaction terms for level tones L2/HKE \times structured were significant (1.59, SE = 0.15, $p < 0.001$) as well as level L2/HKE \times naturalistic (3.21, SE = 0.33, $p < 0.001$).

A similar pattern emerged for the model significance when comparing the reference category (rise L1 elicited) to falling tones, where falling tones overall were significant (1.25, SE = 0.11, $p < 0.001$), and fall L2/HKE was significant (−1.16, SE = 0.14, $p < 0.001$). Naturalistic falling tones were not significant (−0.13, SE = 0.16, $p = 0.429$). The interaction terms, which are more interpretable as they compare rising tones by L1 speakers in elicited speech to falling tones by L2 speakers in structured speech, were significant (0.70, SE = 0.09, $p < 0.001$), as well as falling tones by HKE speakers in naturalistic speech (1.56, SE = 0.23, $p < 0.001$). The model overall explanatory value was McFadden's $R^2 = 0.07$, suggesting the model was able to explain 7% of the variance. See Table 3 for a complete model summary.

Post hoc analyses using the contrast function conducted significance tests on the rates of rising, falling, and level tones across language background and data collection type. The results are model-informed estimates of the difference between the predicted occurrence of each tone choice. The tests indicated no significant differences between the L1 elicited and L1 naturalistic instances of rising tones (Est = −0.01, $t_{10} = -0.04$, $p = 0.905$), level tones (Est = −0.03, $t_{10} = -1.13$, $p = 0.503$), or falling tones (Est = 0.04, $t_{10} = 1.36$, $p = 0.385$). These findings suggest that despite the smaller probability of rising and level tones in elicited (DCT) speech, and increased probability of falling tones in elicited (DCT) speech, these differences were due to chance. Therefore, these findings provide evidence for the validity

of using the DCT elicitation technique for L1 speakers as the multinomial logistic regression model could not detect a significant difference between these productions.

Table 3. Multinomial Logistic Regression with “Rise L1 elicited (DCT)” as reference category.

Predictors	Coefficient	SE	<i>t</i>	<i>p</i>	
level (Intercept)	0.73	0.12	5.99	<0.001	***
level L2/HKE	−3.30	0.27	−12.30	<0.001	***
level naturalistic	0.04	0.17	0.21	0.834	
level L2/HKE × structured	1.59	0.15	11.00	<0.001	***
level L2/HKE × naturalistic	3.21	0.33	9.80	<0.001	***
fall (Intercept)	1.25	0.11	10.90	<0.001	***
fall L2/HKE	−1.16	0.14	−8.08	<0.001	***
fall naturalistic	−0.13	0.16	−0.79	0.429	
fall L2/HKE × structured	0.70	0.09	8.14	<0.001	***
fall L2/HKE × naturalistic	1.56	0.23	6.88	<0.001	***

Note. *** $p < 0.001$.

Further post hoc analyses were conducted on the L2/HKE speech, which included three types of data collection techniques (elicited, structured, and naturalistic). The comparisons tests indicated a significant difference between L2 elicited and L2 structured rising tones (Est = 0.32, $t_{10} = 11.81$, $p < 0.001$) as well as L2 elicited and HKE naturalistic rising tones (Est = 0.33, $t_{10} = 12.59$, $p < 0.001$). No significant difference was detected between L2 structured and HKE naturalistic rising tones (Est = 0.01, $t_{10} = 0.22$, $p = 0.973$). A similar pattern emerged for level tones in which L2 elicited and L2 structured level tones were significantly different (Est = −0.22, $t_{10} = −9.66$, $p < 0.001$) and level tones in L2 elicited and L2 naturalistic were significantly different (Est = −0.23, $t_{10} = −10.93$, $p < 0.001$). There was no detectable difference between level tones when comparing L2 structured and L2 naturalistic speech (Est = −0.01, $t_{10} = −0.19$, $p = 0.979$). Finally, for falling tones, L2 elicited rates of falling tones were less likely than in L2 structured speech (Est = −0.10, $t_{10} = −3.17$, $p = 0.025$), and L2 elicited falling tones were less predicted than HKE naturalistic tones (Est = −0.10, $t_{10} = −3.36$, $p = 0.018$). Falling tones compared between L2 structured production were equivalent to HKE naturalistic speech (Est = 0.00, $t_{10} = 0.02$, $p = 0.999$). However, it should be noted that the estimates and significance level are much smaller for falling tones than rising or level tones. These findings suggest that tone choices roughly occur at the same rates between structured (i.e., collaborative task) speech in L2 production, but elicited speech through DCTs produce significantly fewer level tones and significantly more rising tones.

To further illuminate the differences, a qualitative approach was taken to extracting examples from the datasets, with a particular focus on rising and level tones in context where the L2 speaker is signaling a focus on their own speech. Rising tones, which were more frequent in elicited (DCT) speech by L2 learners, occurred often as a part of the request in a near-canonical production according to Brazil’s (1997) framework to build solidarity and indicate that the speaker is ready to listen to their interlocutor. Example 1 is from an L2 speaker in the elicited (DCT) dataset (capitalization indicates prominent syllables).

1. HI profESSor SMITH i was WONdering if YOU COULD WRITE a LETter of recom-menDation for ME to apPLY for a scholar SHIP ✍

However, in Example 2, an L2 naturalistic speech, a similar binary question at the end of a speaker's turn is produced with a level tone. This example takes place in CCOT conversation A050. This production is likely because of the complex multi-speaker dynamics at play in which the speaker is less interested in the interlocutor's response.

2. DOES it affect the LANguage or THE FORmat **HERE** ⇒

In Example 3, a structured speech from the collaborative task dataset, a sample from participant 06513 also illustrates a question at the end of a relatively long turn that would canonically have either a falling tone to indicate that the speaker is finished or a rising tone to focus their attention on the interlocutor. However, in this case, the limited time of the speaking tasks and pressure to express a sufficient amount of language for the assessment evaluators resulted in a level tone.

3. WE can reDUCE the RATES AND WHAT do you **THINK** ⇒

The non-canonical forms in naturalistic speech because of the production circumstances are therefore one of the likely explanatory factors behind the misalignment between prosody in elicited (DCT) speech and naturalistic speech found amongst both datasets for L2 speakers.

Additionally, increased pragmatic complexity may explain some of the differences in L1 and L2 tone patterns. In Example 4 from the elicited (DCT) dataset, an L1 speaker likely chose a level tone to not overly pressure the interlocutor into complying with the request. A falling tone might have been too finite for the context, and a rising tone may have implied too much pressure, leaving the speaker with a level tone as a strategy for focusing on the language rather than the interaction. See Example 4 for an L1 request elicited by a DCT.

4. HI mrs BLACK, i was WONdering if i could SOMEhow get in an english one oh FIVE this se**MES**ter ⇒

In contrast, L2 speakers, when faced with the DCT, reverted to the canonical rising tone indicating their attention on the interlocutor and their attempt to build solidarity with someone in power. It is also possible that their learning experience reduced the complexity of intonation in discourse to simply rising with questions and falling with statements, a theme commonly found in pedagogical materials. See Example 5.

5. CAN i CHANGE my CLASS ENGLISH one oh FIVE ↗ i wanna GO with my **FRIEND** ↗

In sum, qualitative analyses of data highlight two phenomena that explain the statistically significant differences between the rising and level tones of L2 elicited (DCT) speech and all other types of speech produced by both L1 and L2/HKE speakers, including those produced by the L1 speakers to the same prompts. These include speech in L2 elicited (DCT) tasks overly focused on shifting the attention to the interlocutor through the use of more rising tones than level tones, as well as the use of canonical forms rather than focusing on the language to mitigate the directness of the speech. The contrasts between the tones produced by L1 and L2 speech suggest an effect of the elicitation technique (i.e., DCT) that may reduce the naturalness of the speech.

8. Discussion

The study attempted to shed light on validity issues of data elicitation instruments in research at the intersection of pragmatics and prosody. Specifically, the study explored variation in tone choice patterns in spoken data depending on the data collection type (elicited, structured, and naturally produced) and on the language background of speakers. In order to address these questions, we made use of existing corpora of naturalistic and structured speech (HKCSE and CCOT) as well as researcher-collected datasets of elicited

data to examine how English learners employed falling, rising, and level tones in speech, limiting the focus to L1 Chinese speakers.

The results of the study revealed an intricate picture of tone choice variation in L1/HKE vs. L2 speech that varied due to both language background and data collection type. The tonal distribution was relatively similar in L1 English and HKE speech across naturalistic and elicited speech; about half of tone choices were falling tones, followed by level and few rising tones. In contrast, L2 speakers used very few level tones in elicited speech. The same was not exhibited in structured or naturalistic speech. The statistical comparisons indicated that to a large extent, L2 speakers responding to DCTs were the only major difference to any other type of speech. Similarly, previous studies have consistently noted that tone choices tend to be distributed differently in L1 Vs. L2 production in discourse, across speech acts and other pragmatic constructs (Pickering, 2009). However, these findings indicate largely similar patterns of tone choices in naturalistic and speech and structured speech amongst L1 and HKE speakers, suggesting that the data collection method may have impacted previous findings on differences between L1 and L2 speakers' tone choices.

In terms of L2 intonation use, parallel to current findings, previous studies have found that L2 English users tend to employ more rising tones in performing various speech acts (Kang et al., 2021; Taguchi et al., 2021). Romero-Trillo (2019) suggested that one reason behind this elevated use of rising tones (at least in requests and refusals) may be due to L2 learners' higher exposure to this feature (and thus mimicking its production) when interacting with L1 speakers, who tend to use more rising tones in L1-L2 communication. Previous studies also found that L1 English speakers consistently demonstrate falling intonation in their speech (Kostromitina, 2024). Finally, a higher proportion of level tones in L2 speech has been noted, for example, in Staples (2015) that described suprasegmental features of naturalistic nurse-patient interactions. In particular, L2 nurses' heightened use of level intonation was perceived as indifference and lack of compassion. On the other hand, however, other studies (e.g., Kostromitina, 2024) found an opposite pattern with L1 speakers using level tones more frequently than L2 speakers in elicited discourse, usually in conjunction with giving reasons for a refusal (i.e., with grounder modifications). While previous studies using the HKCSE found that level tones were used for different reasons, such as production constraints, amongst HKE speakers (Cheng et al., 2008), the present study did not find such differences, perhaps due to the extraction techniques. In other words, it is possible that the L2 and HKE speakers' disproportionally greater use of level tones in Staples (2015) and Cheng et al. (2008) were earlier in the utterance. In the present study, only one tone choice per turn was examined.

Brazil's (1997) Discourse Intonation framework provides further insights into the differences in tonal composition of L1 and L2 speech in elicited (DCT) data. In the framework, falling tones are usually associated with introducing new information and establishing common ground. It is expected that when making requests, refusals, and other speech acts, falling tones may be more common since the speaker is, in a way, providing new, previously unknown information to the listener in order to establish mutual understanding about the topic of the speech act. It appears that L2 speakers' use of rising tone choices may not reflect the expected informational structure of performing certain speech acts (Pickering, 2018).

Based on the results of the current study, it appears that intonation in L1 elicited discourse mirrors that of naturalistic speech whereas L2 English users' intonation seems to be more dependent on the data collection method used. Although previous pragmatic research is limited in terms of empirical comparisons between these two data

sources and their validity, we can draw on research from other areas of pragmatics to interpret these findings.

Previous research comparing elicited and naturalistic data collection methods in L2 pragmatics outside of prosody has indeed revealed differences between these two sources. While more controlled instruments like DCTs and role plays can provide data quickly, studies corroborate the present findings in indicating that these data often differ markedly from naturally occurring language use (Golato, 2003). Félix-Brasdefer (2007), for example, observed that DCTs produced only a subset of refusal strategies compared to role plays, which displayed higher frequencies of indirect refusals, more mitigation, and a wider range of strategies similar to natural conversation. These findings align with the current study's results showing that at least from the pragmatic perspective, L2 speakers tend to demonstrate different patterns in elicited versus naturalistic data, highlighting the limitations of elicited data for capturing authentic L2 pragmatic competence.

However, similar to the findings of the current study, the picture seems to be more nuanced for L1 speakers. The finding that L1 speakers tend to demonstrate similar patterns in elicited and naturalistic data aligns with previous studies in pragmatics. For example, Beebe and Clark Cummings (2006) argued that DCTs can provide authentic pragmatic norms even in artificial elicitation contexts. Félix-Brasdefer (2007) also found that data from open role plays represent a plausible approximation to natural speech in formal and informal requests. Previous studies similarly noted the effectiveness of open role plays in generating L1 pragmatic strategies that resembled those of naturalistic discourse, as noted above, including conversation openings and closings (Scarcella, 1979; Schegloff & Sacks, 1973) or indirectness (Félix-Brasdefer, 2005; Reiter, 2002).

9. Implications

The findings of this study have several important implications for researchers investigating pragma-prosodic intonation patterns in L1 and L2 users' speech. With regard to L1 English speech, the study seems to provide support to the continued use of elicited data collection techniques to examine the intonation-discourse link. However, we encourage researchers to use caution in overgeneralizing intonation patterns observed in elicited speech, particularly for L2 speakers. That is, researchers need to be mindful of the potential discrepancies between elicited and naturalistic data, especially when studying L2 speakers.

Researchers interested in examining pragma-prosodic speech features should consider implementing a multi-method approach, as advocated by Geeslin (2010), to leverage the strengths of both elicited and naturalistic data. This approach can help account for the differential effects of data collection methods on speech samples, providing a more comprehensive understanding of pragma-prosodic features of discourse. One approach to this could be to examine naturalistic speech data first, in order to detect prominent features of interest and examine their use, and then conduct follow-up studies for more focused investigations of these features, modifying pragmatic contexts as needed.

The results of this study also highlight potential benefits of structured speech data (such as data from collaborative tasks). Specifically, the findings suggest that when constructed with sufficient contextual information, role plays may offer advantages over natural data by allowing researchers to control for various sociolinguistic variables while still eliciting interactional data. However, researchers should be aware that the degree of *elicitedness* is a more fine-grained variable than previously thought.

10. Limitations

This study is limited in several ways. Namely, the application of Brazil's (1997) model to HKE speech merits further investigation as functions and features of prosody in a Global

English variety may diverge from inner-circle varieties in unexpected ways. Despite the similarity in tabulations in the present study between L1 English and HKE tone choices, the influence of a local tonal language such as Cantonese may have resulted in different pragma-prosodic mappings than those proposed by Brazil (1997). However, the findings of this study can also be viewed as initial evidence to further establish pragma-prosodic features particular to HKE, and a comparison of the more interpretive Discourse Intonation approach with the acoustical autosegmental-metrical approach may indicate important differences in realization of intonation as it relates to pragmatic meaning.

Second, other contextual variables, such as situational characteristics including the number of interlocutors, speaker power and distance, and topic, are likely to impact the tone choices. While holistic analyses may shed some initial light on pragma-prosodic feature use, future studies could aim to control for register in their analyses (i.e., specific contexts and situational characteristics of discourse) and, if possible, include other prosodic features beyond intonation. Additionally, differentiating between L1 backgrounds and examining the transfer of L1 intonation patterns in different languages could provide valuable insights, as well as examining the potential of variation of discourse and pragmatic functions of specific tone choices in global varieties of English, such as HKE.

11. Conclusions

The study examined differences in intonation patterns across elicited and naturalistic L1, HKE, and L2 English data, aiming to provide validity evidence for these data sources. While elicited data may not always be representative of intonation in L2 spoken discourse, it appears to be more reflective of naturalistic discourse for L1 speakers. This study underscores the complexity of investigating intonation patterns and pragmatic competence across different speaker groups and data collection methods. Based on the study results, we argue that by adopting a nuanced, multi-method approach and carefully considering the implications outlined above, researchers interested in studying pragma-prosodic speech features can enhance the validity and generalizability of their findings in this area. Future research will allow for further insight into the impact of data collection techniques as well as (in)validate the use of existing and novel spoken data elicitation techniques for speakers from various backgrounds.

Author Contributions: Conceptualization, K.H. and M.K.; methodology, K.H. and M.K.; validation, K.H. and M.K.; data curation, K.H. and M.K.; writing—original draft preparation, K.H. and M.K.; writing—review and editing, K.H. and M.K.; visualization, K.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Northern Arizona University (protocol code: 171113; date of approval: 22 February 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Dataset available on request from the authors.

Conflicts of Interest: Maria Kostromitina was employed by Duolingo, Inc. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Aijmer, K. (1997). *Conversational routines in English: Convention and creativity* (1st ed.). Routledge. [CrossRef]
- Beebe, L., & Clark Cummings, M. (2006). Natural speech act data versus written questionnaire data: How data collection method affects speech act performance. In S. Gass, & J. Neu (Eds.), *Speech acts across cultures: Challenges to communication in a second language* (pp. 65–88). De Gruyter Mouton. [CrossRef]
- Biber, D., & Staples, S. (2014). Exploring the prosody of stance: Variation in the realization of stance adverbials. In T. Raso, & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 271–294). John Benjamins Publishing Company. [CrossRef]
- Björkman, B. (2012). Questions in academic ELF interaction. *Journal of English as a Lingua Franca*, 1(1), 93–119. [CrossRef]
- Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* [Computer program]. Available online: <http://www.praat.org/> (accessed on 4 August 2020).
- Brazil, D. (1997). *The communicative value of intonation in English*. Cambridge University Press.
- Brown, G., Currie, K., & Kenworthy, J. (2015). *Questions of intonation* (1st ed.). Routledge. [CrossRef]
- Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge University Press. [CrossRef]
- Cheng, W., Greaves, C., & Warren, M. (2005). The creation of a prosodically transcribed intercultural corpus: The Hong Kong Corpus of Spoken English (prosodic). *ICAME Journal*, 29, 47–68.
- Cheng, W., Greaves, C., & Warren, M. (2008). *A corpus-driven study of discourse intonation: The Hong Kong corpus of spoken English (prosodic)* (Vol. 32). John Benjamins Publishing. [CrossRef]
- Cheng, W., & Warren, M. (2005). //CAN I help you: The use of rise and rise-fall tones in the Hong Kong Corpus of Spoken English. *International Journal of Corpus Linguistics*, 10(1), 85–107. [CrossRef]
- Cheng, W., & Warren, M. (2008). //→ONE country two SYStems/: The discourse intonation patterns of word associations. In A. Ädel, & R. Reppen (Eds.), *Corpora and discourse: The challenges of different settings* (pp. 135–153). John Benjamins Publishing Company. [CrossRef]
- Chun, D. (2002). *Discourse intonation in L2: From theory and research to practice*. John Benjamins Publishing Company. [CrossRef]
- Crawford, W., & McDonough, K. (2021). Introduction to the corpus of collaborative oral tasks. In W. Crawford (Ed.), *Multiple perspectives on learner interaction: The corpus of collaborative oral tasks* (pp. 7–16). De Gruyter Mouton. [CrossRef]
- Félix-Brasdefer, J. C. (2004). Interlanguage refusals: Linguistic politeness and length of residence in the target community. *Language Learning*, 54(4), 587–653. [CrossRef]
- Félix-Brasdefer, J. C. (2005). Indirectness and politeness in Mexican requests. In D. Eddington (Ed.), *Selected proceedings of the 7th Hispanic linguistics symposium* (pp. 66–78). Cascadilla Proceedings Project.
- Félix-Brasdefer, J. C. (2007). Natural speech vs. elicited data: A comparison of natural and role play requests in Mexican Spanish. *Spanish in Context*, 4(2), 159–185. [CrossRef]
- Geeslin, K. (2010). Beyond “naturalistic”: On the role of task characteristics and the importance of multiple elicitation methods. *Studies in Hispanic and Lusophone Linguistics*, 3(2), 501–520. [CrossRef]
- Goh, C. (2000). A discourse approach to the description of intonation in Singapore English. In A. Brown, D. Deterding, & E. L. Low (Eds.), *The English language in Singapore: Research on pronunciation* (pp. 35–45). Singapore Association for Applied Linguistics.
- Golato, A. (2003). Studying compliment responses: A comparison of DCTs and recordings of naturally occurring talk. *Applied Linguistics*, 24(1), 90–121. [CrossRef]
- Hewings, M. (1995). Tone choice in the English intonation of non-native speakers. *IRAL: International Review of Applied Linguistics in Language Teaching*, 33(3), 251–265. [CrossRef]
- Ikeda, N. (2017). *Measuring L2 oral pragmatic abilities for use in social contexts: Development and validation of an assessment instrument for L2 pragmatics performance in university settings* [Doctoral dissertation, University of Melbourne].
- Kachru, B. B. (Ed.). (1992). *The other tongue: English across cultures*. University of Illinois press.
- Kang, O., Kermad, A., & Taguchi, N. (2021). The effect of study abroad and proficiency on speech acts. *Journal of Second Language Pronunciation*, 7(3), 343–369. [CrossRef]
- Kostromitina, M. (2024). An exploratory study of intonational variation in L1 and L2 English speakers’ pragmatic production of high imposition requests and refusals. *Applied Pragmatics*, 6(1), 1–30. [CrossRef]
- Kostromitina, M., & Miao, Y. (2024). Listener perception of appropriateness of L1 and L2 refusals in English. *Studies in Second Language Learning and Teaching*, 14(2), 291–310. [CrossRef]
- Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.
- Levis, J. M., & Wichmann, A. (2015). English intonation–Form and meaning. In M. Reed, & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 139–155). Wiley. [CrossRef]
- Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly*, 35(2), 233–255. [CrossRef]
- Pickering, L. (2009). Intonation as a pragmatic resource in ELF interaction. *Intercultural Pragmatics*, 6(2), 235–255. [CrossRef]

- Pickering, L. (2018). *Discourse intonation: A discourse-pragmatic approach to teaching the pronunciation of English*. University of Michigan Press.
- Pickering, L., & Wiltshire, C. (2000). Pitch accent in Indian-English teaching discourse. *World Englishes*, 19(2), 173–183. [\[CrossRef\]](#)
- Ramírez-Verdugo, M. D., Jiménez Vilches, R., Rodríguez Merchan, B., & Aronsson, B. (2017). First and second language prosody: A study on speech production, perception and pragmatic features. In T. Navarro Tomas (Ed.), *Tendencias actuales en fonética experimental: Cruce de disciplinas en el centenario del Manual de Pronunciación Española* (pp. 268–270). UNED.
- Ramírez-Verdugo, D. (2008). A cross-linguistic study on the pragmatics of intonation in directives. In J. Romero-Trillo (Ed.), *Pragmatics and corpus linguistics* (pp. 205–233). De Gruyter Mouton. [\[CrossRef\]](#)
- Reiter, R. M. (2002). A contrastive study of conventional indirectness in Spanish: Evidence from Peninsular and Uruguayan Spanish. *Pragmatics*, 12(2), 135–151. [\[CrossRef\]](#)
- Ripley, B., & Venables, W. (2016). *nnet: Feed-forward neural networks and multinomial log-linear models*. The Comprehensive R Archive Network (CRAN).
- Romero-Trillo, J. (2019). Prosodic pragmatics and feedback in intercultural communication. *Journal of Pragmatics*, 151, 91–102. [\[CrossRef\]](#)
- Scarcella, R. (1979). On speaking politely in a second language. In C. Yorio, K. Perkins, & J. Schachter (Eds.), *On TESOL '79: The learner in focus* (pp. 275–287). TESOL.
- Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica*, 8, 289–327. [\[CrossRef\]](#)
- Staples, S. (2015). *The discourse of nurse-patient interactions: Contrasting the communicative styles of US and international nurses*. John Benjamins. [\[CrossRef\]](#)
- Taguchi, N., Hirschi, K., & Kang, O. (2021). Longitudinal L2 development in the prosodic marking of pragmatic meaning: Prosodic changes in L2 speech acts and individual factors. *Studies in Second Language Acquisition*, 44(3), 843–858. [\[CrossRef\]](#)
- Taguchi, N., & Roever, C. (2017). *Second language pragmatics*. Oxford University Press.
- Warren, M. (2004). A corpus-driven analysis of the use of intonation to assert dominance and control. In *Applied corpus linguistics* (pp. 21–33). Brill. [\[CrossRef\]](#)
- Yuan, Y. (2001). An inquiry into empirical pragmatics data-gathering methods: Written DCTs, oral DCTs, field notes, and natural conversations. *Journal of Pragmatics*, 33(2), 271–292. [\[CrossRef\]](#)

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.