



Comparing spoken discourse marker use in NS and NNS speech: A corpus-based study using ChatGPT

Soojin Ahn

(Incheon National University)

Ahn, Soojin. 2025. Comparing spoken discourse marker use in NS and NNS speech: A corpus-based study using ChatGPT. *Linguistic Research* 42(Special Edition): 261-289.

This study investigates the frequency and functional use of spoken discourse markers (DMs) by native speakers (NS) and non-native English speakers (NNS) using spoken monologue data from the ICNALE corpus. The dataset includes English learners from nine different L1 backgrounds at two proficiency levels, along with NS speakers. Employing an AI-assisted corpus analysis approach using ChatGPT and regex-based pattern matching, the study identifies 32 spoken discourse markers categorized into 14 functional types. Results reveal clear developmental trends: while NNS-low learners showed narrow and formulaic use, NNS-high learners demonstrated more frequent and flexible use, approaching NS patterns. However, both NNS groups underused pragmatically rich markers (e.g., *well*, *I mean*, *you know*) and overused basic ones like *I think*, *so* and *and*. NSs showed greater variety and more implicit functional use. The study highlights the value of explicit instruction targeting underused DMs and shows how generative AI can support scalable, context sensitive discourse analysis. Findings offer implications for EFL pedagogy and the integration of AI tools into applied linguistics research. (Incheon National University)

Keywords spoken discourse markers, corpus analysis, ChatGPT, NNS (non-native speakers), NS (native speakers)

1. Introduction

DMs are essential elements in spoken discourse, serving both textual and interpersonal functions by signaling discourse structure, managing turn-taking, and expressing speaker stance (Cuenca and Marín 2012; Fraser 2015). Lexical items such as *you know*, *I mean*, *well*, *so*, and *but* contribute to fluency, coherence, and pragmatic

appropriateness in interaction (Aijmer 2002; Neary-Sundquist 2014). While NS use DMs fluidly across a range of functions, NNS often display restricted usage, limited functional diversity, or deviant patterns such as underuse or overuse of certain markers (Müller 2005; Sankoff et al. 2009). These discrepancies can disrupt natural interaction and signal gaps in pragmatic competence, which is a crucial component of communicative competence (Ishihara and Cohen 2022). According to Bachman (1990), pragmatic competence refers to the ability to use language appropriately in social contexts, which includes sociolinguistic and functional knowledge, both of which are reflected in appropriate DM use.

Despite the growing awareness of the role DMs play in spoken interaction, they remain underrepresented in language curricula and often escape explicit instruction in EFL contexts (Bardovi-Harlig and Hartford 2005; Römer 2022). While native speakers acquire pragmatic norms implicitly through extensive exposure, language learners face challenges in internalizing the nuanced functions of DMs, especially in spontaneous conversation. Consequently, learner speech can sound mechanical, overly formal, or pragmatically inappropriate (Ebrahimi and Xodabande 2023). This mismatch between learners' formal accuracy and pragmatic naturalness underscores the need for pedagogical approaches that target real-time discourse features.

A number of contrastive studies using spoken corpora have explored differences in DM use across NS and NNS speech, revealing insights into how DMs function across varying proficiency levels, L1 backgrounds, and communicative contexts (Shimada 2011; Liu 2017). However, many of these studies continue to rely on earlier computational methods, such as POS tagging and n-gram filtering, which are computationally demanding, requiring substantial processing power, memory, time, and storage, and may not capture the context sensitive nature of DMs (Neary-Sundquist 2014). These methods also struggle to identify multi-functional uses of the same DM across contexts, limiting their usefulness for pragmatic analysis.

This study addresses these limitations by adopting a corpus-based approach enhanced by AI-driven analysis. Using large-scale datasets of spoken English from both NS and NNS speakers of diverse L1 backgrounds, this research leverages the capabilities of generative AI, particularly ChatGPT, to facilitate context sensitive identification and classification of DMs. Unlike traditional computational methods, ChatGPT allows for rapid and intuitive filtering of DMs from natural speech data without the need for complex preprocessing, such as syntactic parsing or customized

tag sets.

By demonstrating how ChatGPT can be used to analyze DMs more efficiently and pragmatically, this study proposes an innovative and accessible methodology for the study of spoken pragmatics. It not only contributes to the field of corpus-based pragmatics but also offers practical implications for EFL instruction by helping educators better understand and teach the dynamic use of DMs. The aim of this study is to compare the use of DMs in NS and NNS speech using a ChatGPT-assisted corpus analysis, highlighting patterns of usage, functional diversity, and pedagogical implications. In doing so, it introduces a novel interdisciplinary approach that combines generative AI, corpus linguistics, and second language acquisition.

2. Background

2.1 Functional pragmatic approach to discourse marker classification

This study adopts a functional pragmatic approach to discourse marker analysis, drawing primarily on the frameworks of Carter and McCarthy (2006) and Fraser (2010), which emphasize the role of DMs as discourse-organizing devices rather than semantically fixed expressions. In this view, DMs are categorized by their pragmatic function in spoken interaction, such as sequencing, reformulation, or stance marking, rather than by syntactic or lexical form.

Carter and McCarthy's (2006) model, which classifies DMs based on their contribution to discourse flow (e.g., managing sequencing, reformulation, epistemic stance, and speaker alignment), serves as the foundation for the present study. This model has been widely applied in corpus-based research due to its operationalizability and relevance to spoken English. Similarly, Fraser's (2010) typology of pragmatic markers, especially those expressing epistemic stance or signaling topic management, provides further nuance to the categorization process.

Given the wide variation in DM classifications across frameworks (Blakemore 2002), this study acknowledges that any functional typology is necessarily interpretive. However, the chosen frameworks (Carter and McCarthy 2006; Fraser 2010) are well-established in applied corpus linguistics and pragmatic research, lending theoretical and methodological consistency to the current analysis. Moreover, the

operational definitions developed in this study (see Table 2) were cross validated with actual corpus data, ensuring the applicability of theoretical categories to naturally occurring speech.

While some discourse markers like *so* can serve multiple functions (e.g., sequencing, summarizing, topic shift, hedging), this study limits its classification to three consistently observable categories: sequencing, resultative, and resuming. This decision reflects both theoretical precedent and the nature of the the International Corpus Network of Asian Learners of English (ICNALE) data, which consists of monologic, explanatory speech. Functions like hedging or interpersonal negotiation, though relevant in dialogic contexts, were rare or difficult to distinguish reliably in this dataset and were therefore excluded.

Importantly, although the ICNALE corpus consists of monologues, prior research (Aijmer 2002; Aijmer and Simon-Vandenberg 2006) has shown that core aspects of pragmatic competence, such as managing topic flow, signaling epistemic stance, and structuring discourse, remain observable even in non-interactive contexts. Thus, monologic data remain valid for exploring the developmental patterns and functional diversity of DM use among language learners.

2.2 NNS patterns and developmental trends

NNSs often display distinct patterns in their use of DMs, typically characterized by a reliance on a narrow set of high-frequency, general-purpose markers, that is, DMs that perform a range of basic functions such as linking ideas, indicating sequence, or initiating turns, but lack the interpersonal or context sensitive nuance of more interactional expressions, such as *and*, *so*, *okay*, and *I think*. In contrast, more pragmatically complex items like *well*, *you know*, or *I mean* are often underused by learners (Muller 2005; Taguchi and Kim 2018; Huang 2019).

This restricted use reflects not only a limited functional repertoire but also a tendency toward formulaic deployment. Learners may insert familiar markers without fully understanding their discourse role, treating them as placeholders or imitating native-like rhythm (Ishihara and Cohen 2022). For instance, *so* may be used indiscriminately to begin a sentence, even when it does not logically connect with the previous clause. Such mechanical use contrasts with NSs, who deploy DMs flexibly

for a range of functions including reformulation, contrast, stance marking, and topic management.

Proficiency level also plays a critical role. Beginner learners tend to rely heavily on simple additive or causal markers (*and, but, so*), while more advanced learners begin to use a broader set of DMs, including evaluative (*actually, to be honest*) and elaborative (*in fact, in other words*) forms (Romero-Trillo 2002; Ebrahimi and Xodabande 2023). Nevertheless, even among higher-proficiency learners, certain pragmatic functions remain underdeveloped, particularly those associated with interpersonal nuance or discourse level cohesion.

Sociolinguistic factors such as exposure to natural input and peer interaction also shape DM development. Studies show that learners immersed in informal or peer driven environments may approximate NS usage patterns more closely (Diskin 2017; Huang 2019). Likewise, bilingual or heritage speakers often exhibit richer DM repertoires than foreign language learners, influenced by identity and acculturation (Kim 2012). These findings highlight that pragmatic development is not solely a function of formal instruction but also of communicative context and social alignment.

2.3 Novel contributions of the present study

Large-scale spoken corpora, such as the LINDSEI, MICASE, or BASE corpora, have provided valuable data for tracking DM frequency and function. However, despite their richness, studies using these corpora often rely on labor intensive and rigid methods such as manual annotation or n-gram filtering, which limit the efficiency and contextual accuracy required for pragmatic analysis (Neary-Sundquist 2014; Brezina et al. 2018). Moreover, many previous studies focused on relatively small or specialized datasets when comparing NS and NNS usage, limiting the generalizability of findings.

Recent developments in AI-based language modeling offer promising alternatives. Tools like ChatGPT possess the capability to interpret discourse context and speaker intent, crucial factors for identifying the subtle and multi-functional use of DMs. While prior research has demonstrated the benefits of integrating corpus linguistics with learner pragmatics, few studies have leveraged AI to address these limitations or explore new pedagogical applications.

This study makes two key contributions. First, it extends the scope of NS-NNS comparison by analyzing DM usage using the ICNALE, a large-scale spoken corpus that has been underutilized in prior DM research. Unlike many previous studies limited to smaller or less standardized datasets, ICNALE provides a more robust empirical foundation for investigating cross-linguistic and proficiency-related variation. A more detailed description of this corpus follows in Section 3.1. Second, it introduces a novel methodology by applying ChatGPT to analyze the frequency and functional roles of DMs. This AI-enhanced approach enables context sensitive, efficient, and scalable analysis of pragmatic features that are often overlooked in traditional corpus-based methods (see Section 3.3 to get more detail).

By combining robust data with innovative tools, this study bridges the gap between computational pragmatics and applied language teaching, offering a framework that is both empirically rich and pedagogically relevant. The study addresses the following research question:

Q. How do native speakers and EFL learners differ in the frequency and functional use of spoken DMs in the ICNALE corpus?

3. Method

3.1 Data

The spoken data used in this study consists of a corpus of informal monologues from ICNALE, featuring EFL learners at different proficiency levels and a native speaker corpus as a baseline. Proficiency levels were aligned with the Common European Framework of Reference for Languages (CEFR), with B2 representing NNS-high and A2 representing NNS-low. These levels were determined by transforming TOEIC/TOEFL or vocabulary size test scores into CEFR bands. The ICNALE Spoken Monologue includes four levels of CEFR: A2, B1-1, B1-2, and B2.

This dataset comprises 60 second speech samples on two predetermined topics: (1) the importance of part-time jobs for college students, and (2) whether smoking should be banned in restaurants. ICNALE includes 4,400 monologues totaling approximately 73 hours of audio, collected via phone from both NSs and EFL learners

across various Asian L1 backgrounds: Korean, Indonesian, Chinese, Japanese, Thai, Taiwanese, Malay, Pakistani, and Filipino.

The analysis in this study is based on the ICNALE Spoken Corpus, which, although not composed of naturally occurring conversation, offers several key advantages in the context of EFL research. Given that EFL learners often have limited opportunities for spontaneous English interaction, structured monologic speech can still yield valuable insights into their pragmatic competence. In addition to being free and publicly accessible (<http://language.sakura.ne.jp/icnale/>), ICNALE offers three features that are especially critical for the goals of this study: (1) standardized spoken tasks across all participants, allowing for consistent functional level comparison; (2) a large and diverse sample of speakers from multiple L1 backgrounds; and (3) CEFR-based proficiency classifications, which support developmental analysis of DM usage. These features make ICNALE one of the most suitable resources for large-scale, cross-sectional analysis of discourse marker use in L2 spoken English. The total word count used in this study is 203,243 words, with distribution shown in Table 1.

Table 1. Corpus overview

	# of files	Word tokens
NNS-high (B2)	637	70,669
NNS-low (A2)	398	36,415
NS (ENS)	614	96,159
Total	1,649	203,243

3.2 Data analysis

This study investigates the distribution and functional use of spoken DMs by native and non-native speakers. Building on the functional pragmatic approach outlined in Section 2.1, the classification of DMs in this study was based on the framework proposed by Šimčikaitė (2012), who adapted the functional classification developed by Carter and McCarthy (2006) for spoken English. In this model, DMs are identified by their discourse organizing roles such as signaling sequencing, reformulation, and epistemic stance rather than on fixed lexical forms. This analytical perspective allowed the study to capture the nuanced ways in which DMs contribute to coherence and

speaker stance in spoken monologic discourse.

Building on this framework, the study also incorporated Fraser's (2010) distinction between different categories of DMs, with particular attention to those expressing epistemic stance (e.g., *I think, I suppose*) and vagueness (e.g., *kind of, things like that*). These types of markers are especially relevant in spoken monologic tasks such as those found in the ICNALE corpus, as they help speakers convey degrees of certainty, belief, or approximation, key aspects of pragmatic competence.

The final list of 32 DMs used in this study was compiled through a combination of theory-driven selection and corpus-based verification. Drawing on the functional frameworks outlined above, an initial list of 42 spoken DMs, grouped into 16 functional categories, was first assembled. This preliminary pool was then refined by examining the frequency and discourse functions of each marker within the ICNALE corpus. Only those that appeared at least once in the data and fulfilled a discourse-pragmatic function were retained. As a result, a total of 32 DMs across 14 functional categories were selected for analysis. This dual approach, combining theoretical grounding with empirical observation, was necessary due to the lack of a universally agreed-upon taxonomy of spoken DMs. Where functional boundaries were ambiguous, operational criteria (Table 2) were applied consistently to ensure interpretive reliability.

This broader approach allows for a more comprehensive understanding of how DMs function in managing spoken interaction. Since this study focuses only on identifying instances where expressions function as discourse markers, it was necessary to provide ChatGPT with detailed operational definitions to help exclude non-DM uses. However, such clarification was necessary only for markers where functional classification could be ambiguous. Table 2 outlines functional criteria focusing on boundary cases where DM interpretation varies depending on context, rather than exhaustively listing all markers included in the analysis.

Table 2. Functional criteria for identifying DMs

Discourse marker	Included use (as DM)	Excluded use
<i>And</i>	Used at the beginning of a sentence to add a point or continue a narrative	As a coordinating conjunction between two clauses
<i>So</i>	Used for sequencing, indicating results, or resuming a topic	As an intensifier (e.g., "so good"); uses indicating summarizing, topic shift, or

		hedging without clear discourse-organizing function
<i>Of course</i>	Used as a resultative (to show logical consequence) or concessive (to acknowledge a point before contrast)	As a general emphatic phrase
<i>Well</i>	Used to reformulate, hesitate, or shift the topic	As an adverb meaning “in a good way”
<i>I mean</i>	Used to clarify or modify previous speech	Literal use of the verb “mean”
<i>Actually</i>	Used to adjust or contrast a prior statement	Used purely as an adverb meaning “in fact”
<i>Then</i>	Used to imply inference (“in that case,” “as a result”) inferring a logical consequence or result	Temporal adverb referring to past time
<i>Now</i>	Used to initiate or shift the topic	Temporal adverb meaning “at the present moment”
<i>Talking about</i>	Used to introduce or shift to a new topic	Literal progressive verb phrase (e.g., “was talking about”)
<i>Look</i>	Used to draw attention to what follows	Literal verb usage (“I look at...”)
<i>You know</i>	Used to assume shared knowledge or to seek listener agreement	Used habitually without functional intent, or in contexts with no clear listener engagement (especially in monologic data)
<i>Like</i>	Used to manage hesitation, approximation, or reformulation	Literal verb meaning “to enjoy or prefer”; filler uses that merely sustain speech flow without contributing to discourse structure or speaker stance (e.g., the smoke of the cigarette will be – um – <i>like</i> in – <i>like</i> especially the food); literal use as a preposition
<i>Right</i>	Used as a closing marker or transition	Adjective meaning “correct”
<i>Kind of/ Sort of</i>	Used to soften a claim or imply uncertainty	Used in its literal sense to refer to category or type (e.g., “a kind of fruit”)
<i>Looks like</i>	Used to soften a claim or imply uncertainty	Literal verb usage (“It appears to be...”)

3.3 ChatGPT-assisted analysis

To identify and functionally classify DMs within the ICNALE corpus, this study employed a ChatGPT-assisted procedure combining automated detection, selective manual review, and validation. First, regular expression (regex) patterns were designed based on predefined inclusion/exclusion criteria outlined in Table 2 to flag candidate DMs reflecting core functional categories (e.g., *so* + result clause). Selected examples involving *so* (Table 3) demonstrate how ChatGPT's initial classifications were refined through contextual evaluation. Similar procedures were applied to other discourse markers in the dataset.

Table 3. Filtering and functional classification of *so* in learner speech

No.	Sentence	Functional use	Analysis
1	So, this part-time job can make us make full use of our spare time.	Functional (Resultative)	The previous sentence mentions having a lot of free time, and this sentence presents a logical consequence, showing how that time can be used effectively.
2	Although the job is useful than to gain some <i>so</i> – society ex – uh – ex – experience, it also could, uh, be...	Non-functional (Partial word/error)	<i>So</i> is part of the interrupted word “society” and does not function as a discourse marker.
3	So, say time could, uh, bring – bring some money.	Non-functional (Unclear)	There is no clear logical connection to the previous sentence, and <i>so</i> seems to function as a filler rather than signaling a result or sequence.
4	So, um, we don't – we don't know how to use it creatively.	Functional (Resultative or Weak Sequencing)	A vague causal or sequencing relationship is implied, though the connection is weak. Retained as a minimal-function DM.
5	Smoking makes other people bad. So, I think – and people smoking in restaurant is especially banned.	Non-functional (Filler)	This instance of “ <i>So</i> , I think” is formulaic and appears without a clear functional link to the preceding clause. It was excluded from analysis.

Given the potential limitations of large language models such as ChatGPT in consistently interpreting pragmatic nuance, this study adopted a hybrid verification process to enhance reliability. While ChatGPT's role was primarily as a first-pass classifier, its outputs were cross validated through manual review and AntConc

concordance checks: Ambiguous or borderline cases like sentence (4) in Table 3 were reviewed manually by the researcher to ensure functional clarity. Also, to validate the results, representative examples were cross checked with actual concordance lines using AntConc, and repeated frequency checks were conducted on separate days to confirm consistency. Only instances with clear discourse functions, confirmed through triangulated review, were retained for final analysis. This triangulated approach strengthens the methodological rigor by combining the scalability of AI tools with human pragmatic intuition, ensuring that the findings are both replicable and contextually accurate.

4. Results

A standardized Type-Token Ratio (TTR) of spoken DMs occurring in the three corpora is presented in the following table. Relying solely on raw frequency or simple type counts can lead to misleading interpretations, especially when comparing corpora of different lengths. To ensure a more meaningful comparison, this study adopts the TTR, which normalizes the number of unique DM types relative to the total number of tokens.

Table 4. Standardized TTR of spoken DMs across three corpora

Functional category	Spoken DMs	NNS high	NNS low	NS
		TTR	TTR	TTR
Sequencing	<i>And</i>	0.00195	0.00126	0.00106
	<i>So</i>	0.00072	0.00178	0.00048
	<i>In general</i>	0.00003	0	0.00005
	<i>Going back to</i>	0.00001	0	0.00002
Resultative	<i>So</i>	0.00096	0.00258	0.00092
	<i>Of course</i>	0.00003	0.00005	0.00003
Reformulation	<i>In other words</i>	0	0	0.00001
	<i>Well</i>	0.00004	0	0.00010
	<i>I mean</i>	0.00040	0.00027	0.00047

	<i>actually</i>	0.00055	0.00019	0.00044
	<i>By the way</i>	0	0	0.00001
	<i>So to speak</i>	0	0	0.00002
Concessive	<i>Of course</i>	0.00008	0.00016	0.00011
Resuming	<i>So</i>	0.00010	0.00008	0.00019
Inference	<i>Then</i>	0.00008	0.00019	0.00064
Opening	<i>Now</i>	0.00004	0	0.00012
	<i>Talking about</i>	0.00003	0	0.00001
Focusing attention	<i>Remember</i>	0	0	0.00006
	<i>Look</i>	0	0	0.00002
	<i>Just think</i>	0	0	0.00002
	<i>Yeah/ Yes</i>	0.00079	0.00041	0.00088
Shared knowledge	<i>You know</i>	0.00051	0.00019	0.00266
Difficulty to formulate	<i>Like</i>	0.00085	0.00033	0.00149
Diverting	<i>Oh</i>	0	0	0.00009
Closing	<i>Right</i>	0	0	0.00007
Epistemic stance	<i>I think</i>	0.00930	0.00349	0.00722
	<i>I suppose</i>	0	0	0.00006
Vagueness	<i>kind of</i>	0.00069	0.00016	0.00099
	<i>sort of</i>	0.00007	0.00003	0.00018
	<i>something like</i>	0.00010	0.00014	0.00015
	<i>things like (that)</i>	0.00010	0.00005	0.00021
	<i>looks like</i>	0	0	0.00001

Table 4 shows that the most frequently used functional categories in all three corpora are epistemic stance (*I think*), resultative (*so*), and sequencing (*and*) discourse markers. Interestingly, the overall frequencies of *I think*, *so*, and *and* are relatively lower in the NS corpus (0.00722, 0.00092, 0.00106) compared to those in the NNS-high (0.00930, 0.00096, 0.00195) and NNS-low (0.00349, 0.00258, 0.00126) groups. The following examples were selected to represent the most frequent usage patterns

observed in each group, based on concordance lines and functional coding.

4.1 Sequencing markers

Sequencing markers help structure spoken discourse by indicating the order of ideas. Among these, *and* was used more often by NNS-high speakers (0.00195) than by NNS-low (0.00126) and NSs (0.00106) speakers, suggesting NNSs rely more on this basic connector, whereas NSs may use a broader range of sequencing markers (e.g., *in general*, *going back to*), indicating greater functional variety in discourse structuring.

- (1) NS Example: “But *going back to* the topic, I believe that people have the rights to smoke...”
- (2) NNS-high Example: “it hurts your health seriously, so I don’t think it’s a good idea to smoking and... *And* I also think smoking people, they don’t have to smoke in the restaurant necessarily.”

Here, *And* links two related opinions, continuing the speaker’s stance against public smoking. Although the sentence is grammatically flawed, *And* is functioning as a sequencing marker by adding a secondary argument.

In the case of *so* used for sequencing, NNS-low showed the highest TTR (0.00178), followed by NNS-high (0.00072) and NS (0.00048) groups. This may indicate NNS-low speakers use *so* a lot because it’s a familiar and easy word that helps them organize their speech either as a smart choice or because they don’t have many alternatives yet. In contrast, NNS-high and NSs appear to diversify their sequencing strategies more effectively, using *so* more selectively and in conjunction with other markers.

- (3) NNS-low Example: “First, I want to get some experiences from the part-time job. *So*, I started working at the convenience store during the semester.”

So is used here to link the intention (getting experience) with the resulting action (starting the job). The sentence shows a logical, temporal progression of ideas, a key trait of sequencing discourse markers. While the TTR values for sequencing *so* were filtered to include only functionally accurate discourse marker uses, the example below

highlight typical misuses that were excluded from the count due to mechanical, ambiguous, or unclear usage.

- (4) NNS-low Example: “ I can take care of myself good and – uh – uh – I –I also could – uh – um – in – increase my – my – my... So, in my opinion of my college life, I think the most important thing is cultivating all individuals”

In this example, *so* appears to function more as a filler or vague transition than as a clearly sequenced discourse marker. The phrase “in my opinion of my college life” is awkward, and the sentence lacks a smooth progression of ideas.

4.2 Resultative markers

So again appeared frequently as a resultative marker, with higher TTRs among NNS-low speakers (0.00258) compared to NNS-high (0.00096) and NSs (0.00092) groups. However, not all instances of *so* functioned appropriately as resultative markers. In particular, NNSs often used *so* without a clear cause-effect relationship, resulting in exclusions from the DM count. The following example shows qualitative differences in usage between NSs and NNSs.

- (5) NS Example: “Many students struggle financially. *So*, taking a part-time job helps them...”
 (6) NNS-high Example: “... it is the essential problem about the money. *So*, while students are studying, they are already 18 years old.”

The NS example appropriately aligned the marker with logical conclusions. In contrast, the sentence before *so* in the learner example talks about financial problems as a reason to get a part-time job. The sentence after *so* simply states a fact (students are 18 years old) without a clear causal link to the financial issue.

Of course, used to mark expected results, showed limited use across all groups even though the TTR for resultative *of course* is little bit higher in NNS-low data (0.00005) compared to NNS-high (0.00003) and NS (0.00003) groups.

- (7) NNS-low Example: “it is – um – opp – opportunity to let us know more about the society and we can – uh – know more people according to this part-time job. *Of course*, we can earn some money...”

The marker *Of course* here was used to signal that the third outcome, earning money, is a natural or expected result of the previously mentioned activity (having a part-time job). It's functioning to reinforce a conclusion that follows logically and predictably from the prior information. The example below is a typical NSS utterance which applied the marker more as a general emphasis marker rather than to indicate a clear result.

- (8) NNS-low Example: “I think it is not important to students to have part-time jobs because *of course* they have a lot of assignments to do.”

The sentence after *of course* does not directly result from the previous statement. NNS-low learners might be using it mechanically rather than purposefully to indicate a result, which was excluded from the DM count. A true resultative example would look like: “Students have a lot of assignments. *Of course*, this means they don't have time for part-time jobs.”

4.3 Reformulation markers

Markers like *I mean*, *well*, and *actually* serve to clarify or adjust previous utterances. In particular, *I mean* appeared much more frequently among NS speech (0.00047) than among NNS-high (0.00040) or NNS-low (0.00027) groups. However, beyond frequency, qualitative differences in usage were also observed. In NS speech, *I mean* typically functions to refine or specify a prior idea, contributing to discourse coherence.

- (9) NS Example: “...to completely ban something like smoking is a bit ridiculous, especially considering – *I mean* while it's a bit lethal, it's not ...”

This use shows clear rephrasing that modifies and softens the speaker's earlier

evaluation. In contrast, NNS-low speakers often used *I mean* in a more mechanical or less functional way.

- (10) NNS-low Example: “I think I not really agree that a smoking should be completely banned in our country because – a – smoke – *I mean* cigarettes company have a big influence in our – in our investment, *I mean* – a – *I mean* investment in our country and then it is –”

Here, *I mean* appears three times in rapid succession but fails to introduce meaningful clarification. Instead, it is followed by vague or repetitive content, indicating that the marker is being used as a filler or hesitation device rather than a true reformulation cue, resulting in exclusion from the DM count.

In particular, *actually* appeared much more frequently among NNS-high speech (0.00055) than among NSs (0.00044) or NNS-low (0.00019) groups.

- (11) NNS-high Example: “Moreover – the reason for this is moreover that student may know the society life after the graduation. It will give some kind of practice for having jobs, especially – *actually*, sometimes you may be able to continue the same job as a full-time worker – working as a journalist for example.”

In this sentence, *actually* reformulates the previous statement (“some kind of practice for having jobs”) by introducing a more concrete and realistic example. Specifically, within the flow from a general statement to a specific instance, *actually* does not merely serve as an intensifier or habitual filler. Instead, it functions to adjust or supplement the discourse meaningfully.

The notably higher frequency of *actually* among NNS-high learners compared to NNS-low learners may reflect their increased exposure to academic discourse, where *actually* is often used to introduce clarifications or corrections. This suggests that NNS-high speakers may be more used to expressing their thoughts clearly and carefully, skills they likely developed through formal language education and practice in structured speaking settings. In contrast, NSs distributed reformulative functions across a wider range of discourse markers, not only *actually*, but also expressions like *I mean*, *well*, *in other words*, *by the way*, and *so to speak*, which may explain

their lower reliance on *actually* alone.

4.4 Concessive and resuming markers

Of course also functioned as a concessive marker. NSs used it to acknowledge and then contrast a point. *Of course* as concessive markers was used more often by NNS-low speakers (0.00016) than by NSs (0.00011) or NNS-high (0.00008) speakers. The slightly higher use of *of course* as a concessive marker among NNS-low speakers may reflect a tendency to rely on formulaic or overgeneralized expressions that seem polite or assertive in tone. Due to limited discourse awareness or pragmatic sensitivity, lower-proficiency speakers might overuse it to sound more convincing or cooperative, even when the concessive function is only loosely appropriate or adds little to the coherence of the argument. In contrast, NSs and more proficient NNSs may choose from a wider range of concessive markers (e.g., *although*, *but*, *even so* which aren't necessarily the DMs under investigation in this study) or contrastive use of *well* or *actually* depending on the context.

- (12) NNS-low Example: "It will be better for them to find their jobs after they graduate because their resume will – we recently have an experience that you have pick have a job while you are still college student and you might a chance easily if you have an experience and I think another reason is *of course* while you're having a part-time job in your college when you are still a college student is good, is because..."

The speaker acknowledges the common belief that getting a job after graduation is better. Then contrasts it with a different view, emphasizing that having a job during college can also be beneficial. *Of course* is used to acknowledge the opposing idea before asserting their own. This is a classic concessive structure, where *of course* serves to concede a generally accepted point while preparing to contrast it.

Besides a sequencing or resultative marker, *so* also functioned as topic resuming markers. It's used to bring the conversation back to a previous topic or to resume after an interruption. The marker showed the highest frequency in the NS group (0.00019), followed by NNS-high (0.00010) and NSS-low groups (0.00008). The

NNS-high example below demonstrates a return to a previous point to reinforce the main idea, functioning as resuming:

- (13) NNS-high Example: “So, I think smoking should be banned ultimately in order to preserve our health – and another argument that I support smoking should be banned is because... So, as I have mentioned before I think smoking should be banned in all restaurants, ”

In contrast, the NNS-low example below shows that NNS-low speakers tend to use *so* more often as a sequencing or resultative marker rather than for resuming functions:

- (14) NNS-low Example: “I agree this statement. I am now live alone *so* my parents give the money for me, *so* I want to earn myself. And – and a part-time job is the important experience for me, and – I can study about society. *So*, I think it is important for college statement to have a part-time job.”

The first two *so* instances are used to indicate cause-effect relationships while the third “*So*, I think...” functions as a sequencing marker. It does not return to an earlier topic but instead moves the discourse forward with a concluding opinion.

4.5 Inference markers

Then as an inferential marker indicating a logical consequence or implied result from the prior context was most frequently used by NSs (0.00064), followed by NNS-low (0.00019) and NNS-high (0.0008) speakers.

- (15) NS Example: “But I don’t think it’s mandatory that everybody should have a part-time job. Uh, I mean if – if that’s what you want to do *then* that’s great and you should go and do it...”

Here, *then* shows inference where the speaker draws a conclusion based on a

conditional clause (“if that’s what you want to do”). It’s not about time or sequence, but rather a result that logically follows from the previous idea. In contrast, the NNS-low example below shows that NNSs often used *then* to simply mean “next,” thus missing its inferential function, which was excluded from analysis as an inferential marker in this study.

- (16) NNS-low Example: “Smoking is bad to ... So this is one point and *then* smoking will - smoking is bad to the air environment...”

In this case, *then* is used to mean “next” in a list of points. It does not convey any logical consequence, but simply indicates the next idea in a series.

4.6 Opening markers

Markers like *now* and *talking about* served to open or transition into topics. *Now* showed the highest frequency in NS speech (0.00012), followed by NNS-high speakers (0.00004). In spoken discourse, *now* is often used to frame a new point, soften a transition, or refocus the listener’s attention. In short, *now* often signals a shift in the speaker’s talk where they are about to explain something, express a belief, or make a broad statement.

- (17) NS Example: “I’ve also been through several nice little cafes that have been ruined by the smoking and it’s unfortunate that they can’t make the restaurant or café non-smoking. *Now*, I always believe that it would help people, uh, reduce smoking in public restaurants if the government would assist in passing a – a law...”

The use of *now* introduces the speaker’s general opinion or belief related to the previous topic (smoking in restaurants), but moves from a description of experiences to a more abstract or evaluative statement.

Interestingly, *talking about* occurred at a higher rate in NNS-high data (0.00003) compared to NS data (0.00001). NNS-low group never used both markers.

- (18) NNS-high Example: “Firstly, *talking about* the benefits, there are economic benefits...”

This suggests that NSs might use alternatives like *and then*, *so*, *well*, or just shift topics without needing an explicit phrase as in the example below.

- (19) NS Example: “*And then* the second one I would say is kind of life skills, um, so you learn to be on time, you learn to communicate with other people...”

This sentence shows a shift to a new point (life skills), smoothly introduced with *and then*, a natural implicit topic shifting marker often used by NSs. The speaker introduces a new thematic point without using a formulaic phrase like “talking about life skills.”

4.7 Focusing-attention markers

Focusing-attention markers such as *remember* (0.00006), *look* (0.00002), and *just think* (0.00002) appeared only in the NS data, indicating that NSs are more likely to use a wider variety of interactive strategies to guide the listener’s attention or emphasize a point. These markers function not only as continuers or agreement signals but also serve to engage the listener more actively, which is a typical feature of spoken discourse. In contrast, NNSs never used these particular expressions, possibly due to limited exposure to such native-like conversational devices or a lack of pragmatic awareness in spontaneous speech.

However, *yeah/yes* which can also function as focusing-attention or agreement markers was used by both NSs (0.00088) and NNS-high (0.00079) speakers at a comparable rate, and also used by NNS-low speakers (0.00041) to a certain extent. This suggests that *yeah/yes* may be more universally accessible or easily acquired, possibly because of its high frequency in everyday conversation and classroom interactions, making it more familiar to learners regardless of proficiency level.

- (20) NNS-high Example: “A key I feel is the power of contributing out to

the whole economy as well. *Yeah*, I also think actually part-time – doing part-time work during our university is once in a lifetime experience...”

The *yeah* here is not a simple agreement with another speaker and nor is it just a backchannel like supportive feedback. It guides the listener’s attention to a personal reflection or elaboration.

4.8 Shared knowledge markers

You know was strikingly more frequent among NS speakers (0.00266) compared to NNS-high (0.00051) or NNS-low (0.00019) groups.

- (21) NS Example: “... because, uh, they’re just having a little bit of extra pocket cash to *you know* do things, socialize, and eat out ...”

In this example, *you know* functions as a shared knowledge marker, inviting the listener to empathize or agree with what’s being said. This is a typical pragmatic use among native speakers, helping to maintain rapport and conversational flow. In contrast, NNSs, especially NNS-low speakers, used *you know* far less frequently with this function. This may be because, as shown in the example below, NNSs often used it to fill pauses or while planning their speech, rather than to build a shared understanding with the listener. Therefore, such uses were excluded from this study’s analysis of *you know* as a shared knowledge marker.

- (22) NNS-low Example: “And I told that smokers pay a lot of money to buy the cigarette, I think it’s more – it’s better for them to – to allocate their money itself to more – to a better – better – better area like, *you know*, like spending money to doing some sport or...”

Here, *you know* appears during a disfluent moment in speech production, serving more as a planning or hesitation marker than as a shared knowledge cue. It doesn’t seem to function as an intentional appeal to listener understanding, which aligns with the finding that NNS-low speakers rarely used it pragmatically (Taguchi and Kim 2018).

4.9 Difficulty to formulate markers

Like was categorized under difficulty to formulate functions. NSs used *like* to fill pauses, approximate, or clarify, thus helping to maintain fluency. *Like* was used more often by NS speakers (0.00149) than by NNS-high (0.00085) or NNS-low (0.00033) speakers. The examples below show that NSs and learners differ in their use of *like* qualitatively as well as quantitatively.

- (23) NS Example: “I don’t like it either sometimes even I’m a smoker, but you know sometimes, it’s better that people *like* – they go *like* to – *like* a smoking area...”
- (24) NNS-high Example: “So maybe we can *like* have an isolated area for them to smoke but not – not *like* anywhere in a restaurant where they can affect everyone else who is dining in the restaurant.”

The examples show that NSs use *like* naturally to maintain fluency during reformulation or planning what to say next, not interrupting communication. In the NNS-high example, the first *like* was used to fill a pause, buy time, or ease formulation, which is precisely what the “difficulty to formulate” function refers to. However, the second *like* was used in its literal, content word sense (a preposition in this case), not as a DM. This suggests that while learners may know and use *like* in ways similar to native speakers, they may not yet fully control its pragmatic/discoursal functions, especially when formulating spontaneous speech (cf. Müller 2005).

4.10 Diverting and closing markers

Oh (diverting) and *right* (closing) were more typical in NS speech. NSs used *oh* to introduce a shift in topic or tone while *right* helped wrap up a discussion and move on. The markers *oh* and *right* appeared predominantly in NS speech data (0.00009) and (0.00007), respectively. NNS speakers did not use them at all.

- (25) NS Example: “... for tests like the TOEIC ... a student having a part-time job could be useful. On the other hand – *oh*, and by the way, a student

also needs money...”

- (26) NNS-low Example: “First – *oh* this is very difficult example [***]. I think I agree because smoking – yeah, I smoking too. Smoking is everyone–everybody knows as time goes by. Very dangerous in our health and smoking...”

The native speaker was discussing one aspect (for tests), but then suddenly shifts focus to a different but related point (students need money to enjoy themselves). It helps fluidly redirect the topic while maintaining coherence and interactional flow. In contrast, in the learner’s example, *oh* appears to be an emotional reaction (realizing the difficulty of the example), not a strategic shift in topic or a cue to divert attention from one point to another, and thus not included in the diverting DM count.

- (27) NS Example (closing): “They should be worrying about us putting up with them, not the other way around or they should – yeah, yeah. We don’t have to worry about putting up. *Right*. So the point of all this is, is that, you know, even if you can separate them, smokers are still imposing”

In the example, *Right* functions as a closing discourse marker, signaling agreement and the end of a thought unit. NSs use *oh* and *right* as a discourse tool to help structure spoken arguments by signaling shifts, elaborations, or contrasts in the conversation. On the other hand, the complete absence of these two functions in NNS speech may reflect a lack of pragmatic awareness or limited exposure to authentic conversational input, especially in spontaneous, informal spoken English.

4.11 Epistemic stance markers

I think was the most frequently used DM across all groups, with the highest TTR in NNS-high speech (0.00930), followed by NSs (0.00722) and NNS-low (0.00349) groups. As a DM, *I think* is commonly used as a hedge to soften the impact of a statement or as a hesitation device (Fraser 2010).

- (28) NNS-high Example: “*I think* part-time job is important for university students, because, uh, university students are in a situation before they enter the society.”

NNS-high group, having more advanced proficiency, used *I think* most frequently. This could indicate a preference for making their opinions explicit, possibly due to awareness of academic or formal English conventions. NNS-low group also used it to a certain extent, likely relying on it as a default strategy to express ideas cautiously, in the absence of a broader range of DMs.

- (29) NS Example: “Uhm, some restaurants a bit of smoke adds to the atmosphere *I suppose*. Uhm, you know if you got people smoking, people having a good time, ... I think it sometimes makes a bit of a better atmosphere.”

NSs used *I think* at a comparable rate to the NNS-high group because they may have a more diverse set of epistemic stance markers (e.g., *I suppose*), as illustrated in the example above, a pattern not observed in the learner data.

4.12 Vagueness markers

Vagueness markers are used as a hedge to avoid making strong claims or to give approximate examples/explanations and the hesitation/uncertainty. Vague expressions like *kind of* (0.00099), *sort of* (0.00018), *something like* (0.00015), and *things like* (that) (0.00021) were more commonly used by NSs. NNS-high speakers used *kind of* relatively often (0.00069), but *sort of* and *things like (that)* remained rare in learner speech. Vagueness markers serve important pragmatic roles in softening claims, expressing approximation and buying time to search for the right words. NSs naturally use a wider variety of vagueness markers. NNS-high speakers use some of these markers but at a lower frequency. NNS-low speakers rarely use them and rely more on simpler expressions like *something like* (0.00014) as in the example below.

- (30) NNS-low Example: “our knowledges that we got from the colleges, we

can just *something like* state - state our knowledges on our mind and our brain”

The instance of *something like* here functions as vagueness discourse markers, reflecting the speaker’s uncertainty or lack of precise lexical access. The frequency patterns across NS, NNS-high, and NNS-low groups indicate that the use of vague expressions develops with language proficiency and exposure. This suggests that explicit instruction in the use of vagueness markers could benefit learners, particularly those at lower proficiency levels, who tended to underuse or apply them awkwardly.

5. General discussion

This study explored the functional use of spoken DMs by native and non-native English speakers using a large-scale learner corpus (ICNALE) and a novel AI-assisted analysis method. In general, NSs exhibit a wider range, more nuanced use, and greater functional flexibility across DM types. In contrast, NNSs tend to rely on a narrower set of familiar markers (e.g., *I think*, *and*, *so*), primarily to signal the sequence of events or propositions, which supports Carter and McCarthy’s (2006) claims. This developmental trajectory aligns with prior research suggesting that markers such as *well*, *you know*, *I mean*, *then*, *in general*, and *kind of* serve important discourse functions, including resuming topics, signaling reformulations, or expressing vagueness, that are often underdeveloped in learner speech (Aijmer and Simon-Vandenberg 2006; Huang 2019). The fact that these markers appear more frequently in NNS-high speech than in NNS-low suggests that learners gradually acquire the pragmatic and functional use of these markers as their proficiency increases, reflecting a deeper command of interactional norms in English.

The results revealed clear developmental trends in DM use across proficiency levels, confirming previous research on the limited functional range among lower-proficiency learners and gradual expansion in more advanced learner speech (e.g., Müller 2005; Neary-Sundquist 2014; Diskin 2017; Huang et al. 2023). However, rather than focusing solely on frequency counts or specific markers, the broader pedagogical and methodological implications of these findings deserve further discussion. Given their pragmatic richness and frequent use in NS speech, these markers (e.g., *well*, *you know*,

I mean) represent valuable instructional targets for helping learners achieve more natural and context sensitive communication.

From a pedagogical perspective, the results highlight the importance of incorporating explicit instruction on DMs into EFL curricula. Instruction should go beyond the basic meanings of high-frequency markers like *so* (function as resultative or sequencing) or *actually* (often used for clarification), and address their multiple functions such as topic resumption (*so*), contrast (*actually*), reformulation (*well, I mean, like*), signaling shared knowledge (*you know*), and vagueness (*kind of, sort of*). It is also important to draw learners' attention to markers that help focus listener attention, such as *yeah/yes*, which were used less frequently by non-native speakers compared to native speakers. By increasing learners' awareness of the diverse pragmatic roles DMs play, including enhancing coherence, naturalness, and managing interaction, educators can help learners move beyond formulaic usage. Furthermore, exposing learners to a wider repertoire of expressions (e.g., *right, I suppose, things like (that)*) may help diversify their discourse strategies and reduce reliance on a narrow set of markers.

To apply these findings in classroom practice, pedagogical interventions can be designed around DM awareness and usage through three stages: (1) input enhancement, (2) guided practice, and (3) contextualized production. First, input enhancement can involve presenting learners with authentic examples of native speaker speech that highlight various functions of DMs (e.g., excerpts from interviews, podcasts, or classroom discourse). Second, guided practice should include activities such as matching DMs to their functions, sentence completion tasks using appropriate markers, or noticing tasks that contrast NS and NNS usage.

Finally, learners can engage in contextualized speaking or writing tasks, such as giving short guided talks, joining structured pair or group discussions, or narrating personal experiences using sentence frames, where they are encouraged to use selected discourse markers for specific functions like sequencing, reformulation, or expressing stance. For beginner to intermediate learners, these tasks can be supported through scaffolding techniques such as modeling with short video or audio clips featuring native speaker use of DMs, shadowing practice, focusing on one DM function at a time, and using familiar, personally meaningful topics to build fluency and confidence. Teachers and material developers can incorporate DM focused sections into textbooks, including reflection prompts (e.g., "What does this DM do in this sentence?") and

checklists (e.g., “Did I use at least one reformulation marker?”). In assessment, rubrics can include a category for pragmatic resourcefulness, helping students internalize the communicative value of appropriate DM use.

Methodologically, this study demonstrates how generative AI tools such as ChatGPT can enhance corpus-based pragmatics research. By supporting contextual accuracy classification and efficient filtering of DM usage, AI-assisted analysis reduces the labor-intensive demands of traditional annotation while retaining interpretive depth. This approach opens new possibilities for large-scale learner corpus analysis, particularly in areas like spoken pragmatics where function is often inseparable from context. In future applications, tools like ChatGPT could also be integrated into classroom practice to provide learners with real-time feedback on their spoken output, model DM use in different contexts, or generate targeted instructional materials.

Despite its contributions, this study has some limitations. First, the ICNALE corpus is composed of structured, monologic speech samples rather than spontaneous conversation. While this allows for standardized comparisons, it may not fully reflect the interactive dynamics of natural discourse. However, EFL learners’ use of DMs in ICNALE shows clear developmental trajectories that parallel findings from interactive settings. While their pragmatic choices are sometimes formulaic, they nonetheless reflect an emerging awareness and gradual acquisition of discourse functions that are pedagogically meaningful. Second, although ChatGPT-assisted annotation enhanced efficiency, some nuanced or borderline cases required human interpretation, meaning absolute consistency could not be guaranteed. Finally, the study focused on a predetermined set of 32 markers, which, while theoretically and empirically grounded, may exclude emerging or less frequent DMs relevant to certain learner groups. Future research could expand to include dialogic data, longitudinal tracking, or learner oriented intervention studies to further validate and extend these findings.

Future research could address these limitations by incorporating spontaneous dialogic data, exploring longitudinal changes in learner DM use, or developing classroom-based interventions focused on pragmatic instruction. Further integration of AI tools, for example, in learner facing feedback or task design, may also enhance both research scalability and pedagogical effectiveness. In sum, this study contributes to a growing body of work on L2 pragmatics by showing how DM development reflects broader patterns of communicative competence. By combining robust corpus data

with AI-enhanced analysis, it offers both empirical insight and practical tools for advancing the teaching and learning of spoken discourse in EFL contexts.

References

- Aijmer, Karin. 2002. *English discourse particles: Evidence from a corpus*. Amsterdam: John Benjamins.
- Aijmer, Karin and Anne-Marie Simon-Vandenberg. 2006. *Pragmatic markers in contrast*. Boston: Brill.
- Algamal, Ahmed M. A. 2020. Syntactic and prosodic features of discourse markers used by native and non-native speakers of English. *Journal of the Faculty of Arts and Humanities* 3(35): 96-133.
- Bachman, Lyle F. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bardovi-Harlig, Kathleen and Beverly S. Hartford. 2005. *Interlanguage pragmatics: Exploring institutional talk*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Blakemore, Diane. 2002. *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.
- Brezina, Voclav, Robbie Love, and Karin Aijmer. 2018. *Corpus approaches to contemporary British speech: Sociolinguistic studies of the spoken BNC2014*. London: Routledge.
- Carter, Ronald and Michael J. McCarthy. 2006. *Cambridge grammar of English*. Cambridge: Cambridge University Press.
- Cuenca, Maria Josep and Maria Josep Marín. 2012. Discourse markers and modality in spoken Catalan: The case of (és) clar. *Journal of Pragmatics* 44(15): 2211-2225.
- Diskin, Catherine. 2017. The use of the discourse-pragmatic marker 'like' by native and non-native speakers of English in Ireland. *Journal of Pragmatics* 120: 144-157.
- Ebrahimi, Hourieh and Ismail Xodabande. 2023. The functions of discourse markers in L2 teachers' and learners' speech across gender and proficiency: A corpus study. *MEXTESOL Journal* 47(1): 1-15.
- Fraser, Bruce. 2010. Pragmatic competence: The case of hedging. In Gunther Kaltenböck, Wiltrud Mihatsch, and Stefan Schneider (eds.), *New approaches to hedging*, 15-34. Bingley: Emerald.
- Fraser, Bruce. 2015. The combining of discourse markers: A beginning. *Journal of Pragmatics* 86: 48-53.
- Huang, Lan-fen 2019. A corpus-based exploration of the discourse marker well in spoken interlanguage. *Language and Speech* 62(3): 570-593.
- Huang, Lan-Fen, Eric Yen-Liang Lin, and Tomáš Gráf. 2023. Development of the use of discourse

- markers across different fluency levels of CEFR: A learner corpus analysis. *Pragmatics* 33(1): 49-77.
- Ishihara, Noriko and Andrew D. Cohen. 2022. *Teaching and learning pragmatics: Where language and culture meet*. London: Routledge.
- Kim, Ihnhee. 2012. Phenomena of discourse marker use of bilingual children and implications for heritage language education. *The Korean Language in America* 17(1): 24-54.
- Liu, Binmei. 2017. A contrastive study of discourse markers used by native and Chinese L2 English speakers across speech context. *East Asian Pragmatics* 2(1): 101-126.
- Müller, Simone. 2005. *Discourse markers in native and non-native English*. Amsterdam: John Benjamins Publishing Company.
- Neary-Sundquist, Collen. 2014. The use of pragmatic markers across proficiency levels in second language speech. *Studies in Second Language Learning and Teaching* 4: 637-663.
- Römer, Ute. 2022. Applied corpus linguistics for language acquisition, pedagogy, and beyond. *Language Teaching* 55(2): 233-244.
- Romero-Trillo, Jesús. 2002. The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics* 34(6): 769-784.
- Sankoff, Gillian, Pierrette Thibault, Naomi Nagy, Hélène Blondeau, Marie-Odile Fonollosa, and Lucie Gagnon. 2009. Variation in the use of discourse markers in a language contact situation. *Language Variation and Change* 9(2): 191-217.
- Shimada, Kazunari. 2011. Discourse marker use in learners and native speakers: A corpus-based analysis of spoken English. *ARELE: Annual Review of English Language Education in Japan* 22: 377-392.
- Šimčikaitė, Ale. 2012. Spoken discourse markers in learner academic writing. *Studies about Languages* 20: 27-34.
- Taguchi, Naoko and Youjin Kim. 2018. *Task-based approaches to teaching and assessing pragmatics*. Amsterdam: John Benjamins Publishing Company.

Soojin Ahn

Visiting Professor

Department of Liberal Education

Incheon National University

119 Academy-ro Yeonsu-gu,

Incheon, 22012, Republic of Korea

E-mail: sja@inu.ac.kr

Received: 2025. 05. 08.

Revised: 2025. 08. 01.

Accepted: 2025. 08. 01.