

An Investigation Into English Learners' Speech Acts Using a Semi-Automatic Annotation Tool

Zhaoyi Pan

English Language Coordinating Section, School of Information Technology, King Mongkut's University of Technology Thonburi, Bangkok, Thailand

Abstract—This paper explores the speech acts most frequently performed by learners of English as a foreign language (EFL) through the Dialogue Annotation and Research Tool (DART), a semi-automatic annotation tool. The study used learner corpus research and an English as the first language (L1) speaker corpus as a reference to compare the speech acts performed by the EFL learners. The study involved 90 EFL learners of four nationalities. A spoken learner corpus was built with the dyadic interlanguage English conversation register. The most frequently performed speech acts were generally different for the English L1 speakers and the EFL learners. The speech acts labeled as *correctSelf*, *referProcess*, *expressNonAwareness*, and *stateReason* were in the top 10 for EFL learners, whereas they were not in the top 10 for the English L1 speakers. This difference was caused by three factors: the English proficiency of the EFL learners, the frequently used formulaic phrases, and the task requirements. Moreover, the corpus annotation has a problem with unrecognized speech acts, primarily due to the redundancies and fragments in the learner discourses. The findings of this study reveal the complementary relationship between grammatical and pragmatic performances.

Index Terms—speech acts, semi-automatic annotation tool, DART, interlanguage pragmatics, learner corpus research

I. INTRODUCTION

Interlanguage pragmatics is the study of “language use in social interaction” by learners of English as a foreign language (EFL) (Fernández & Staples, 2021, p. 240). One of the significant areas of interlanguage pragmatic studies is the speech acts performed by EFL learners in interlanguage communication (Staples & Fernández, 2019; Fernández & Staples, 2021). A plethora of studies in this field have examined specific types of speech acts, such as requests, refusals, complaints, and apologies (Akmal et al., 2022; Khamkhien, 2022) using discourse-completion tasks (DCTs) to elicit the required speech acts. Despite the fruitful results of these studies, critical concerns have been raised. The studies on specific types of speech acts are primarily based on “rather limited, and mostly abstract, taxonomies” developed by Austin (1962) and Searle (1979; Weisser, 2020, p. 400). Moreover, a lack of corpus methodology leads to the challenge that only “a small number of variant patterns” of the speech acts can be investigated, resulting in a “rather narrow focus of interlanguage pragmatics research on speech acts” (Staples & Fernández, 2019, pp. 241-242). Thus, the use of DCTs rather than naturally occurring data may not reflect the real situation of the speech acts performed by EFL learners (Pan, 2023a).

The present study investigates the speech acts performed by EFL learners in interlanguage communication holistically through a semi-automatic annotation tool: the Dialogue Annotation and Research Tool (DART) created and developed by Weisser (2018, 2019b, 2020) to reveal the speech acts in Weisser's (2020) extended speech act taxonomies that are most frequently performed by EFL learners. This study uses learner corpus research (LCR) methodology and contrastive interlanguage analysis (CIA) (Granger, 2015) to demonstrate the significance of using corpus linguistics to study interlanguage pragmatics and further dedicates the relative lacuna of LCR in spoken data (Fernández, 2023; McEnery et al., 2019; Yoon, 2020).

This paper is divided into five sections. After the introduction, Section 2 provides a literature review. Section 3 describes the research methodology. Section 4 presents the results and discussion based on each research question. A conclusion follows in Section 5.

II. LITERATURE REVIEW

This literature review is divided into four parts: speech act theory, interlanguage pragmatic studies, LCR and corpus pragmatics, and DART.

A. Speech Act Theory

To study the actual language use in different social and situational contexts, Austin (1962) elucidates three acts that can be performed by a speaker in any context: locutionary acts, illocutionary acts, and perlocutionary acts. Locutionary acts are the actions taken when a person makes an utterance in which literal meaning is conveyed. Illocutionary acts, as

they have been widely studied as speech acts, indicate the intention of the speaker embedded in an utterance. Perlocutionary acts refer to the result of an utterance. Harris (2019, p. 53) contends that a speech act is an action “with a communicative intention”; thus, different speech acts “involve intentions to affect hearers in different ways.” Speech acts have received more attention since “the central mechanism of human communication is intention recognition” (Harris, 2019, p. 53).

Searle (1979) further separates Austin’s (1962) illocutionary acts into five categories: assertives, directives, commissives, expressives, and declarations. Moreover, speech acts have been classified as direct and indirect (Staples & Fernández, 2019). These fundamental speech act theories have guided a strand of pragmatic studies, especially in the form-to-function approach. A form might not correspond with its apparent function owing to an indirect speech act. On the other hand, the correspondence between a form and a function might not be one-to-one, since contexts and interactants’ variables can differ, such as the social contexts and social powers of interactants in real-world communication (Akmal et al., 2022; Yuan & Lyu, 2022). These degrees of indeterminacy between an utterance and its speech act have motivated researchers to study different types of speech acts in various registers with a diversity of interactants.

A major criticism of the pragmatic studies following Austin (1962) and Searle (1979) is that they focus only on specific types of speech act, while little attempt has been made to study speech acts systematically (House & Kádár, 2023; Weisser, 2020). The speech act taxonomy proposed by Searle (1979) is limited, which restricts the scope of the examination of the types of speech acts (Weisser, 2020), such as assertion and promise (Yuan & Lyu, 2022). House and Kádár (2023, p. 3) point out that focusing on certain types of speech acts “unavoidably led to a top-down take on speech acts,” and this approach “cannot capture more complex speech act-related problems.” Their viewpoint indicates that the studies on individual types of speech act cannot indicate the speech act performance of an interactant comprehensively, and this top-down approach apparently cannot solve all the issues in this field. This criticism may be more perceptible in the interlanguage pragmatic studies discussed in the next section.

B. Interlanguage Pragmatic Studies

As an important concept of the EFL field, interlanguage pragmatic studies have increased rapidly and changed their focus in the past four decades (Taguchi, 2022). Early research tried to determine which factors account for EFL learners’ pragmatic competence, including “pragmalinguistics and sociopragmatics” or “functional knowledge and sociolinguistic knowledge” (Taguchi, 2022, p. 8). At the same time, studies compared the performance of speech acts of speakers of English as the first language (L1) and EFL speakers were based on Austin’s (1962) and Searle’s (1979) development of speech acts and speech act taxonomies. DCTs have been used since then and are still used in most research on specific types of speech act.

As mentioned earlier, most studies in the interlanguage pragmatic field have examined specific types of speech act, resulting in an excess of investigation on requests, refusals, apologies, compliments, and complaints (Akmal et al., 2022; Budeng & Merza, 2023; Khamkhien, 2022). Given the interplay between these types of speech act and the direct and indirect acts, these studies have attempted to examine different forms produced by interactants with one type of speech act, categorize the forms into different degrees of directness or indirectness, and compare their use by English L1 speakers and EFL learners. The results of these comparative studies of speech acts generally reveal significant differences in the performance of the same types of speech act between English L1 speakers and English interlanguage learners. For example, EFL learners tend to use a lower degree of directness in requests, while English L1 speakers flexibly choose different degrees of directness depending on the social context, situational environment, and the social status of the interactants (Akmal et al., 2022; Pan, 2023a). Furthermore, EFL learners tend to have limitations in choosing the lexical items for the same type of speech act due to their English proficiency levels (Pan, 2023a).

Limitations of interlanguage pragmatic studies have been discussed in recent years. First, similar to the criticism of the studies of speech acts in the previous section, interlanguage pragmatic studies have mainly restricted their interests to a small range of speech acts, and how interlanguage learners perform speech acts in interlanguage communication remains unknown (Fernández & Staples, 2021). Moreover, this small range of speech acts may not fully reveal EFL learners’ pragmatic competence, since speech acts vary, and the range can be extended with different interactional environments and the innovations of further studies (Fernández & Staples, 2021; Weisser, 2018). In addition, since many studies in this field have relied upon DCTs, the results from these studies may not be completely reliable, as participants were usually given time to answer each question in the DCT, and, more importantly, the data from the DCT is not the naturally occurring discourses that participants engage in real interlanguage communication (Pan, 2023a; Staples & Fernández, 2019). To overcome these limitations, researchers in this field have introduced corpus linguistics.

C. Learner Corpus Research and Corpus Pragmatics

LCR is a relatively new field in corpus linguistics, and “a relatively newer methodology within interlanguage pragmatics” (Fernández & Staples, 2021, p. 244). Unlike DCTs, LCR collects the naturally occurring data, both spoken and written discourses, used by language learners (McEnery et al., 2023). Although a new development in SLA, LCR has been used to investigate EFL learners’ pragmatic competence in various sub-fields, such as speech acts, pragmatic markers, formulaic language, and pragmatics in prosody (Fernández, 2022, 2023). Given the innovation of computational technology, LCR facilitates the analysis of EFL learners’ competence in different pragmatic sub-fields.

Hence, the collection of EFL learners’ naturally occurring spoken discourses and the analysis of the speech acts performed by them with the aid of modern computational technology will eventually allow researchers to comprehend how EFL learners perform various types of speech acts in real interlanguage communication. This approach avoids the elicited data that may not reflect EFL learners’ use of speech acts.

As the principles of corpus linguistics were introduced to the field of pragmatics, corpus pragmatics emerged, “combining both key methodologies of corpus linguistics and pragmatics” (Huang, 2021, p. 101). In this study, it is contended that an interlanguage corpus pragmatics field can be further developed as a sub-field of corpus pragmatics in which EFL learners’ performances of a diverse range of pragmatic competence are examined and analyzed via the LCR principles and methods. In interlanguage corpus pragmatics, the learner corpora built by researchers are usually in small, especially those of spoken registers (Huang, 2021; Taguchi, 2022). Many researchers have pointed out that learner corpora in different spoken registers are insufficient (Fernández, 2023; McEnery et al., 2019; Yoon, 2020), thus imposing a constraint on the analysis of EFL learners’ use of pragmatics in oral communication. More learner corpora are needed for studying interlanguage pragmatics so that a diversity of EFL learners with different L1 backgrounds and English proficiency levels can benefit from the results of this field (Yoon, 2020). Modern computational technology makes it possible to produce concordances in which the required forms can be retrieved and extracted from the long discourses produced by EFL learners for the analysis of individual pragmatic phenomena. For different spoken registers, such as dyadic conversations and group discussions, corpus pragmatics provides both horizontal concordances and vertical contexts in any given dialogic activity for researchers to manually annotate the forms and functions aligned with their goals and frameworks (Budeng & Merza, 2023; Weisser, 2020). There is a major issue with this annotation, however, as discussed in the next section.

D. Annotation in Pragmatic Studies

Previous studies of pragmatics and interlanguage pragmatics have mainly used manual annotation. Depending on the purpose of the research, researchers have annotated different types of pragmatic use in various corpora, including individual types of speech act, pragmatic markers, prosodic features, and non-linguistic features in multimodal pragmatic studies (Akmal et al., 2022; Budeng & Merza, 2023; Huang, 2021; Pan, 2023a). Manual annotation can be aligned with the requirements of the researchers. Moreover, certain types of annotations must be manual, such as learner errors in learner corpora (McEnery et al., 2019). However, manual annotation consumes a large amount of time and may incur more costs. Some studies used two or more raters to ensure the reliability of the manual annotation (Pan, 2023a). Despite these weaknesses, perhaps the most significant problem with manual annotation in interlanguage pragmatic studies is its sparsity in LCR (Fernández, 2022, 2023; McEnery et al., 2019). Not many learner corpora have been annotated. In addition, a learner corpus with comprehensive pragmatic annotation is difficult to find, especially for spoken registers (Fernández & Staples, 2021).

This situation in the interlanguage pragmatic studies is primarily caused by the limited annotation tools for pragmatics. DART is a relatively well-developed example. It uses extensible markup language (XML) to produce different notations for each utterance. As the main purpose of the tool is to retrieve and analyze the speech act performances in a given corpus, DART provides a more fine-grained extended speech act taxonomy (Weisser, 2019a, 2020) than Searle’s (1979). The extended speech act taxonomy is “applicable to the annotation and analysis of any type of naturally occurring dialogue” (Weisser, 2020, p. 405). DART’s extended speech act taxonomy contains nine super-categories with 162 detailed speech act categories (SACs) with a justification for each category. Table 1 illustrates the nine extended speech act super-categories and gives an example of a SAC in each super-category based on the DART scheme (Weisser, 2018, 2019a, 2020).

TABLE 1
SPEECH ACT SUPER-CATEGORIES IN DART SCHEME

Super-category	Explanation	Example	Function
1 information or option seeking acts	different categories in requests	reqInfo	requesting verbal information
2 (non)cohesive acts	managing the cohesion for texts and interaction	answer	answering a question
3 information providing and referring acts	providing various types of information	state	conveying information/awareness
4 suggesting or commitment indicating acts	giving various degrees of suggestion or commitment	suggest	proposing joint or interlocutor’s potential action
5 negotiative acts	helping to negotiate	accept	responding in an active positive way
6 evaluating or attitudinal acts	indicating personal opinions or stance	agree	signaling explicit agreement
7 reinforcing acts	emphasizing content	emphatic	repeating something for emphasis
8 social, conventionalized acts	expressing daily and social rituals	greet	greeting the interlocutor
9 residual acts	unintelligible or uninterpretable	uninterpretable	uninterpretable, due to missing or incoherent information

DART is accessible, applicable, convenient, and user-friendly regarding the classification of utterances, such as the single units of *yes* and *no*, the pragmatic markers, and the backchannels, which are more relevant to the linguistic components of the annotation of speech acts. Weisser’s (2020, 2021) series of studies have shed light on computational pragmatic annotation and point toward the future of corpus pragmatics.

Nevertheless, the DART program is generally acknowledged to be a semi-automatic form of annotation for two main reasons. First, problems involving misunderstandings or unidentified instances have been found in the annotation process (Weisser, 2018; Verdonik, 2022); thus, users are advised to examine the automatic results in the post-process. In addition, DART relies on the literal meaning of utterances in the annotations. However, the under-determinacy of the oral interactions may lead to an unstable basis for the identification of the speech act in an utterance, which suggests possible difficulties in the pragmatic annotation programs.

The present study has three main objectives. First, it aims to compare the performances of EFL learners and English L1 speakers in the most frequently performed speech acts based on the DART extended speech act taxonomy. Second, it seeks to determine the factors that cause the differences. Third, it reports the problems with using DART to annotate the speech acts in learner corpora to improve the development of future annotation tools in the interlanguage pragmatic field. The study seeks to answer three research questions:

- 1) What are the differences in the most frequently performed speech acts between EFL learners and English L1 speakers?
- 2) What are the factors that cause the differences in the most frequently performed speech acts between EFL learners and English L1 speakers?
- 3) What are the problems with using DART to annotate the speech acts in learner corpora?

III. METHODOLOGY

A. *Participants and Data Collection*

To collect the speech acts performed by EFL learners with different L1 backgrounds, this study engaged 90 participants with four nationalities: 30 Thai (TH) EFL learners, 22 Indonesian (IN) EFL learners, 20 Chinese (CH) EFL learners, and 18 Burmese (MM) EFL learners. All the participants attended the same government university in Bangkok, Thailand. There were 52 males (58%) and 38 females (42%). They were aged between 18 and 23 years and were in year 1 to year 4 at the undergraduate level. They belonged to the English programs or the international programs of four faculties. Each participant's L1 was their country's official language, and they had studied EFL for 11 to 14 years in their own countries. No participant had lived in an English-speaking country before the data collection. Their English proficiency levels were all above B1 based on the valid scores of an international English examination, such as TOEFL iBT, and the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2020). Thus, the participants in this study represent EFL learners with different L1 backgrounds at the undergraduate level. Gender was not considered a variable in this study.

The data were collected on the campus of the university between April and May 2023. Each participant was required to have a casual English conversation with another participant of the same nationality. The spoken register in this study is therefore the dyadic English interlanguage conversation. To ensure that each conversation was close to naturally occurring spoken discourse, each pair of participants could choose the topics to discuss. In case any pair did not know what to discuss, they were given the topics of an English course titled "General English" that is a required part of the English courses at the university. All the topics pertain to general daily or social life, such as happiness and technology. Each pair was requested to have this English conversation for approximately 15 minutes without any preparation. They were not allowed to use their L1, and they had all been informed that their conversation would be recorded by the iPhone Voice Memo application. The researcher did not appear while each pair was having the conversation to ensure that they would feel relaxed and comfortable and the conversation would be as natural as possible. Thus, 45 pairs of dyadic English interlanguage conversations were collected.

B. *Data Analysis*

This study used both quantitative and qualitative approaches in the LCR field. DART was used to examine the speech acts that the participants performed. DART uses XML to produce different notations for each utterance. All the spoken discourses were transcribed into written texts using the XML format. The interface of the DART program on the Windows computer system is illustrated in Figure 1. The input and output interfaces of several corpora and the interface for the analysis pertaining to the speech acts (the functions) and the linguistic components (the forms) are provided in the figure below.

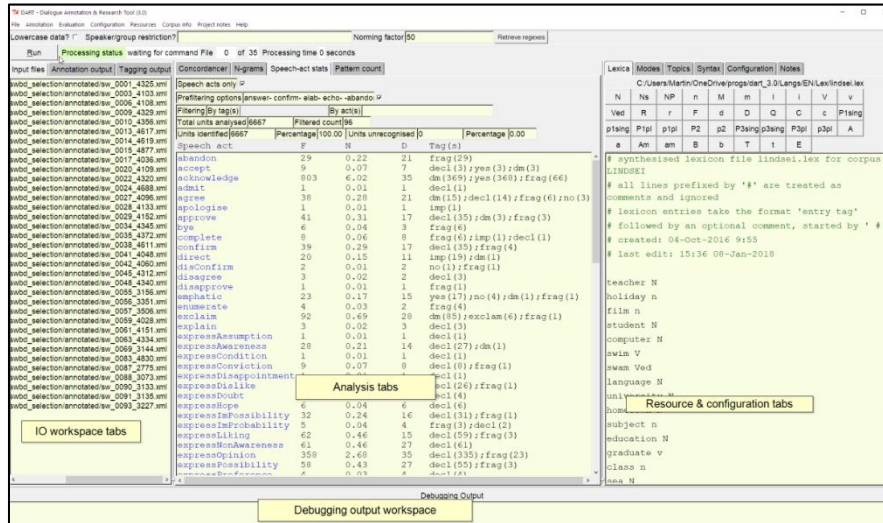


Figure 1. An Example of the Interface of the DART Program (Weisser, 2019a, p. 1)

Two main functions are provided by DART. The first function is to output the speech acts found in all the utterances in a corpus. This function is usually completed by three processes: the pre-process in which a revision of the original data is performed by the user; the computational process of outputting the results for the speech act; and the post-process in which modifications to the problematic annotations can be made manually by the users (Weisser, 2018). The second function enables the relevant pragmatic analysis, which DART achieves by integrating basic analytical tools to retrieve keywords, N-grams, and concordances, thus recording the frequencies of the keywords or the lexical patterns.

To compare the different uses of the SAC between EFL learners and English L1 speakers, the Switchboard Corpus (SC) annotated by Weisser (2018, 2020) was used as the reference of the performance of the speech acts by English L1 speakers based on the principle of CIA² (Granger, 2015). Granger (2015) contends that there should be a discrete consideration about the comparison of the language use between EFL learners and English L1 speakers since EFL learners with different L1 backgrounds have their own patterns of using English as a lingua franca (ELF) or an international language (EIL). Instead, the speech acts performed by English L1 speakers can be used as a for several different patterns of language use in the comparative studies for EFL learners' language improvement. SC was chosen because it contains "American interlocutors of equal status talk[ing] about a variety of different topics" (Weisser, 2020, p. 407). It is therefore comparable to the spoken register in this study. The differences of the most frequently performed speech acts were scrutinized by DART, and the factors that caused the differences were examined by the concordances provided by DART.

IV. RESULTS AND DISCUSSION

A. Different Uses of the Most Frequently Performed Speech Act Categories

This section presents the quantitative results of the different performances of the most frequently performed speech acts in the two corpora. The fine-grained qualitative analysis is presented in the next section. In this study, a learner corpus titled Speech Acts of Learners Corpus (SALC) was built. The corpus contains 45 samples of dyadic English interlanguage conversations, which were all given an identification number based on the input requirement of DART. For example, SALC01 refers to the first pair of conversations between participants 01 and 02. According to the descriptive data provided by DART, SALC contains 67,612 tokens. After DART's automatic annotation and a few manual changes suggested by Weisser (2018), this study found 83 SACs in the nine super-categories performed by all the participants.

Table 2 lists the 10 most frequently performed speech acts in SALC, along with the normalized frequency (NF per 100,000 tokens) of each SAC, the number (N) of samples and the corresponding proportion (P) of the entire sample in which each SAC was found, and the super-category (S) that each SAC belongs to. Since SC is regarded as the reference, the same information of each SAC in SC is listed for comparison. A significant difference was not considered necessary. The performance of the speech acts by English L1 speakers was mainly used to show the features of the performance of the speech acts by EFL learners by comparing the differences. Any significant difference in the overuse or underuse results may not have any substantial meaning towards the performance of the speech acts in English interlanguage communication.

TABLE 2
MOST FREQUENTLY PERFORMED SAC IN SALC AND SC AS THE REFERENCE

R	S	SAC	SALC		SC		R
			NF	N/P (%)	NF	N/P (%)	
1	information providing and referring acts	state	913	45/100	1,119	35/100	1
2	(non)cohesive acts	hesitate	890	45/100	405	35/100	4
3	information or option seeking acts	reqInfo	719	45/100	150	35/100	9
4	(non)cohesive acts	answer	697	45/100	88	34/97	11
5	evaluating or attitudinal acts	expressOpinion	528	45/100	235	35/100	6
6	information providing and referring acts	referProcess	466	40/89	11	8/23	40
7	information providing and referring acts	stateReason	376	45/100	78	30/86	12
8	information providing and referring acts	expressNonAwareness	280	42/93	40	27/77	25
9	negotiative acts	correctSelf	219	37/82	-	-	-
10	(non)cohesive acts	acknowledge	192	45/100	528	35/100	2

Based on the table above, the speech act labeled as state was performed the most frequently by both EFL learners and English L1 speakers. As Weisser (2018, 2020) points out, this speech act is labeled by DART when a declarative or fragment with a finite verb is uttered to provide general information. It is not surprising that this speech act was performed the most in both corpora, since providing information is a fundamental behavior in any type of communication (Weisser, 2020). Another two speech acts, namely hesitate and express opinion, are ranked similarly in the two corpora.

According to the ranking in Table 2, six speech acts are ranked higher in SALC than in SC. The biggest difference is the speech act labeled as correctSelf. It was not found in SC, whereas it is the 9th most frequently performed speech act by EFL learners. Another two speech acts, namely referProcess and expressNonAwareness, are ranked much higher in SALC: 40th and 25th in SC compared to 6th and 8th in SALC, respectively. Three other speech acts are also ranked higher in SALC than in SC: the speech act labeled answer is ranked 11th in SC and 4th in SALC; the speech act labeled as reqInfo is ranked 9th in SC and 3rd in SALC; and the speech act labeled as stateReason is ranked 12th in SC and 7th in SALC. Conversely, one speech act is ranked lower in SALC than in SC. The speech act labeled as acknowledge is 2nd in SC and 10th in SALC.

B. Factors That Cause the Differences

(a). Influence of English Proficiency

This study examined the concordances of each of the most frequently used speech acts to identify the factors that may influence EFL learners' use of the speech acts. The analysis begins with the biggest difference in performance between the two corpora: correctSelf. As Weisser (2018) states, this speech act is labeled when the speaker corrects their own utterance, as illustrated in examples (1) and (2):

- (1) <decl n="127" sp-act="answer" polarity="positive" mode="opinion" topic="opinion-option">
i may give money to the poor or to child <punc type="stop" />
<decl n="128" sp-act="correctSelf">
i will give my money to children with [with no family <punc type="stop" /> (SALC02)
- (2) <frag n="119" sp-act="answer" polarity="positive" topic="location" mode="">
we go to south and </frag>
</dm>
<no n="120" sp-act="negate">
no <punc type="stop" />
<decl id="67" sp-act="correctSelf" polarity="positive" topic="" mode="">
we went to the south </decl> (SALC15)

Both examples above exhibit three common self-correction behaviors of EFL learners. First, both participants change the tense of the verb ("give" in the first example and "go" in the second), revealing a correction of the grammatical structures of the utterances. Second, both participants correct certain lexical items. The word "child" is corrected to its plural form "children" in (1), and "south" is corrected to "the south" by adding a definite article in (2). Finally, the participant in (1) changes the structure of the original utterance by adding the expression "with no family" to describe the noun "children." Through these two examples, it can be seen that the EFL learners' English proficiency leads to the speech act labeled as correctSelf. Given that the conversation is spontaneous and filled with indeterminacy (Pan, 2023a), EFL learners correct the errors during the conversation. Weisser (2018) likewise mentions that this speech act is expected to be found in learner discourses.

Previous research on interlanguage pragmatics has contended that EFL learners with different levels of English proficiency use pragmatics to different degrees, including speech acts or pragmatic markers (Pan, 2023a; Staples & Fernández, 2019; Taguchi, 2022). Some studies have shown that the proficiency levels and the pragmatic performance

are proportional, while others have not found a significant relation between the two (Taguchi, 2022). Nevertheless, these studies have rarely concentrated on whether EFL learners attempt to correct the grammatical errors to express the correct speech acts based on Searle's (1979) taxonomy. This use of the speech act labeled as correctSelf according to the DART taxonomy shows that EFL learners try to repair different types of errors that they make by themselves. It indicates that EFL learners have an awareness of using the correct grammatical structure to produce the correct pragmatic sense. Thus, correcting the verb tense and adding an expression in (1) makes the intention of the participant more certain; and correcting the verb tense and adding a definite article in (2) allows the declarative sentence to be more easily comprehended.

The influence of the English proficiency level of the EFL learners is also revealed in the speech act labeled as hesitate. This should not be surprising, since DART "captures initial fillers that indicate a short-time planning process" in front of the utterances, including *uh* and *um* (Weisser, 2018, p. 224). In line with Weisser (2018), Tübben and Landert (2022) argue that instead of regarding *uh* and *um* as fillers without any pragmatic sense, they should be considered as the planners that give the interlocutors more time to think about the next utterance. The fact that the speech act labeled as hesitate was the second most frequent speech act in SALC shows that EFL learners frequently use *uh* or *um* to think about the next utterance. It is conceivable that, as learners, they have more opportunity to use *uh* or *um* due to their English proficiency.

(b). Influence of Frequently Used Expressions

Some speech acts are marked as such by DART because of key words pertaining to the core functions of the respective speech act (Weisser, 2018, 2021). The present study found three speech acts that are relevant to this situation: stateNonAwareness, stateReason, and expressOpinion. The speech act labeled as stateNonAwareness has the biggest difference in ranking between the two corpora. This is a speech act that reflects a lower degree of awareness of the content discussed in a prior utterance by the speaker. Some expressions commonly used in spoken discourse, such as "I don't know" and "no idea," generally indicate the speaker's unawareness of certain things, as in (3) below:

- (3) <q-yn n="35" sp-act="reqInfo" polarity="positive" topic="details" mode="closed-query">
do you have some example <punc type="query" /></q-yn>
</turn>
<turn n="42" speaker="19">
<del n="51" sp-act="expressNonAwareness" polarity="negative" mode="nonawareness-decl">
i don't know <punc type="stop" />
<decl n="52" sp-act="expressPossibility" polarity="positive" mode="poss-decl">
i can tell you a story <punc type="stop" /></decl> (SALC10)

A closer examination of the expression "I don't know" used by EFL learners reveals different levels of unawareness. A simple "I don't know" without more explanation or information indicates the speaker's complete lack of knowledge. The "I don't know" in example (3) may not be as simple, however, since it is followed by another declarative utterance, "I can tell you a story." This addition suggests that the speaker has some information relevant to the question, but at a low level. Previous studies have found that EFL learners use "I don't know" as a formulaic phrase with different pragmatic uses, including indicating complete unawareness and allowing themselves more time to think about how to explain something (Hosoda & Aline, 2021; Pan, 2023b). Pan (2023b) examined the use of "I don't know" by EFL learners with different English proficiency levels and found that this formulaic language was used at a high frequency with both functions mentioned above. The present study further confirms the common use of this formulaic language in learner discourse.

Concerning the speech acts labeled as stateReason and expressOpinion, this study found that EFL learners frequently used the word "because" to state reasons and the formulaic phrase "I think" to express personal opinions, as in (4) and (5):

- (4) <del n="12" sp-act="state" polarity="positive" topic="arrival" mode="information">
my brother got there <punc type="level" />
<del n="13" sp-act="stateReason" polarity="positive" mode="reason-decl">
because he want <pause /> **see the [the show** <punc type="stop" /> (SALC23)
- (5) <del n="73" sp-act="expressOpinion" polarity="positive" mode="opinion-decl">
i think we are just lazy <punc type="stop" /> (SALC05)

The use of the expression "I think" by EFL learners in spoken registers has been found in previous studies (Pan, 2023b). The frequent use of "because" and "I think" by EFL learners in spoken discourses is not surprising, since they are directly connected to their speech act functions and are usually learned by EFL learners during their early English education. This study found that the frequent use of certain words directly pertaining to their speech act functions led to the high ranking of some speech acts in SALC.

(c). *Influence of Task and Effects of Multiple Factors*

Since the spoken register in this study is the dyadic English interlanguage conversation, which all occurred between two participants of the same nationality, it is conceivable that the task itself had an impact on the performance of the speech acts. Both participants in each pair had to complete the task in around 15 minutes, thus keeping the interaction flowing. Asking for information (the speech act labeled as reqInfo) and answering questions (the speech act labeled as answer) are fundamental strategies to maintain the interaction, as in example (6):

- (6) <del n="59" sp-act="reqInfo" polarity="positive" topic="life" mode="query">
what is your school life like <punc type="query" />
 </turn>
 <turn n="28" speaker="21">
 <frag n="26" sp-act="answer" polarity="positive" mode="decl">
it's ok and fun <punc type="stop" /></decl> (SALC11)

A series of questions and responses, such as those in example (6), can help keep the conversation flowing. Theoretically, the numbers of the speech acts reqInfo and answer should be equal, since the speech act labeled as answer is the response to the speech act labeled as reqInfo based on the DART scheme (Weisser, 2019b). However, the interlocutor often simply utters “yeah,” “alright,” or “okay” (the speech act labeled as acknowledge) in response to a prior question (Weisser, 2018). Moreover, not every question is answered in conversations, as counter-questions, fragments, rejections, and other situations influence whether reqInfo and answer are equal in number (Weisser, 2018). This phenomenon also explains the decrease in rank of the speech act labeled as acknowledge in SALC. In SALC, questions more often took the form of special interrogatives that needed to be answered with substantial information than general questions. As a result, the speech act labeled as acknowledge was performed less often than the speech act labeled as answer. By contrast, the English L1 speakers in SC may not have felt the same pressure to complete a task as the EFL learners, since English is their L1. Hence, relatively more speech act labeled as acknowledge was performed in SC and was performed more casually.

Finally, another speech act that had a higher rank in SALC was referProcess. Following careful examination of the performance of this speech act, this study found that the task and the English proficiency level interact to give this speech act a high rank, as in example (7):

- (7) <frag n="63" sp-act="referProcess" polarity="positive" mode="frag">
He to study [study hard [hard <punc type="level" /></frag>
 <frag n="64" sp-act="referProcess" polarity="positive" mode="frag">
 <pause /> **going university to study** <punc type="level" /></frag>
 <dm n="81" sp-act="hesitate">
 um </dm>
 <frag n="65" sp-act="referProcess" polarity="positive" mode="frag">
 <pause /> **to also part time work** <punc type="level" /></frag> (SALC32)

This is a typical example of this speech act, which was performed by various participants in SALC. First, the speaker in (7) attempts to interact with the interlocutor by describing a series of events. This process is determined by the task itself. However, due to the limitation of the speaker’s English proficiency, DART cannot capture the three declaratives that were supposed to be uttered with the correct verb forms. In fragment 63, no verb is given after the pronoun “he.” In fragment 64, the non-finite verb “going” is uttered without a subject. In fragment 65, no subject or verb is provided. In these circumstances, DART can only mark the three fragments with the speech act labeled as referProcess, indicating that several actions are ongoing (Weisser, 2018). This phenomenon occurs because the English proficiency of the speaker does not allow them to use the correct grammatical structures to utter complete utterances. Both these factors (the task requirement and the English proficiency) cause the speech act labeled as referProcess to be highly ranked in SALC.

C. *Unrecognized Annotation*

DART uses the annotation label “unrecognized” to refer to the utterances for which DART cannot identify any SAC in the DART taxonomy (Weisser, 2019a, 2019b). Unrecognized annotation occurs because of the indeterminacy of spoken registers (Weisser, 2018). For SALC, DART marked 37 unrecognized instances, 25 of which belong to fragment units and 12 are fragments with overlapping. After manually annotating based on the DART taxonomy, four SACs were applied to the 37 instances. Table 3 shows the four SACs annotated manually with an example for each condition.

TABLE 3
SPEECH ACT CATEGORIES ANNOTATED MANUALLY

SAC annotated manually	syntactic structures	number of instances	one example
acknowledge	fragment	12	<frag n="20" sp-act="acknowledge" polarity="positive" mode=""> yeah it yes yes <punc type="stop" /> (SALC17)
	fragment with overlapping	8	<frag n="85" sp-act="acknowledge" polarity="positive" mode="confirm-frag"> you ok ok yes <overlap pos="start" /> one <punc type="stop" /> (SALC09)
refer	fragment	5	<frag n="102" sp-act="refer" polarity="positive" mode="frag"> post to [to post online <punc type="stop" /> (SALC23)
	fragment with overlapping	3	<frag n="11" sp-act="refer" polarity="positive" mode="frag"> going [going go there <overlap pos="end" /> one <punc type="stop" /> (SALC36)
answer	fragment with overlapping	6	<frag n="54" sp-act="answer"> it [it rain <overlap pos="start" /> one <punc type="stop" /> (SALC04)
echo	overlapping	3	<frag n="117" sp-act="echo" polarity="positive" mode="query"> him he brother <punc type="query" /> (SALC30)

As Table 4 illustrates, both fragments in the speech act labeled as acknowledge and refer annotated manually cannot be assigned to any SAC by DART, which may have been due to the redundancy in the utterances. The double use of “yes” appears after an “it” that seems to have no literal sense in the example of the speech act labeled as acknowledge retrieved from SALC17; and an original form of the verb “post” is used, followed by the non-finite verb form “to post” with a double use of “to” in the example of the speech act labeled as refer retrieved from SALC23. The same phenomenon is also captured in the fragment with overlapping, which is when the fragment uttered by the current speaker overlaps with a part of the utterance uttered by the previous speaker, such as the double use of “OK” in the middle of a fragment in the speech act labeled as acknowledge retrieved from SALC09. The dyadic interlanguage English conversation register used in this study is an informal spoken register that often has “incomplete or redundant, inconsistent, syntactic structure, and non-standard dictions” (Akmal et al., 2022, p. 505). Redundancy in the learners’ discourses was found to be due to their English language proficiencies (Staple & Fernández, 2019). The frequent redundancies in utterances result in an unclear semantic sense and lead to the non-recognition of the speech acts (Verdonik, 2022). Weisser (2018) mentions this potential problem with DART. Thus, DART is a semi-automatic annotation tool, and manual annotation is necessary after using the tool. The findings of this study show that DART has an annotation issue when annotating learner corpora and the semantic sense is unclear, primarily owing to the frequent redundancies in the learner discourses.

The findings and discussion above show that the differences in the way the most frequent speech acts are performed by EFL learners and English L1 speakers reflects the interlanguage pragmatic features of EFL learners. These features can be explained by the factors found in this study. Through the macro-perspective of studying the speech acts performed by EFL learners, this study found a complementary relationship between the grammatical performance and the pragmatic performance of the EFL learners. This finding is aligned with those of Mao (2021) and Mao and He (2021). The relationship between grammatical competence and pragmatic competence in interlanguage pragmatic studies has been studied for a long time, but inconsistent results—either a proportional relationship or no significant correlation—bring uncertainty to this issue (Mao, 2021; Pan, 2023a). As the analysis above shows, the complementary relationship of the two performances can be found in each speech act. For example, the correction of grammatical errors made by the speaker led to the correctSelf speech act, whereas the intention of correcting their own errors naturally led participants to utter a new repaired utterance; the tendency to use formulaic phrases, such as “I don’t know”, caused the high rank of the expressNonAwareness speech act, whereas the intention of indicating the participants’ lower degrees of certainty causes them to use formulaic language; the series of utterances with insufficient grammatical correctness result in the speech act labeled as referProcess, whereas the intention of completing the task and explaining the content more clearly causes the participants to process more utterances. Mao and He (2021, p. 10) used a new integrated model and found a “complementary relationship between grammatical competence and pragmatic competence, rather than the latter containing the former.” Through the results of DART and the researcher’s own analysis, this study found further evidence of the complementary relationship of the grammatical and pragmatic performances in the actual use of English by EFL learners in dyadic interlanguage English conversations.

Moreover, the macro-investigation of the speech acts performed by EFL learners revealed that the individual speech acts studied by previous researchers, such as request, refusal, complaint, and apology, were not used at high frequencies in the dyadic interlanguage English conversations. This research is consistent with the findings in those studies, but whether the performance of these individual speech acts by EFL learners can fully exhibit their pragmatic competence may need to be reconsidered. After all, previous research has indicated the issue whether speech acts can represent pragmatic competence, let alone the narrow focus on certain individual speech acts that are not used frequently in real communication (Fernández & Staple, 2022).

V. CONCLUSION

The results of this study reveal the pragmatic features of EFL learners' performance of speech acts by comparing their performance to that of English L1 speakers. This finding indicates the different performance of the speech acts and the different use of EFL in the interlanguage pragmatic field from a computational pragmatic perspective. Through the DART taxonomy and its semi-automatic annotation of the speech acts, the complementary relationship between the grammatical and pragmatic performances of EFL learners is revealed. This finding provides a new viewpoint based on the semi-automatic annotation tool with the extended speech act taxonomy.

In addition, this study used the semi-automatic annotation tool to annotate the speech acts in a learner corpus. The intention was to expand the annotation to interlanguage pragmatic studies, which has rarely been done (Staple & Fernández, 2019; Fernández, 2022). Problems in the annotation of the learner corpus were discussed. Given that the learner discourses may be more likely to contain redundancies and fragments, the annotation problems discussed in this study may provide insights for the development of the annotation tool.

This study involved participants from Asia. Future research could investigate speech acts performed by speakers with other L1 backgrounds. Furthermore, the semi-automatic annotation tool should be used in different spoken registers with different learner corpora.

APPENDIX XML CONVENTIONS

All participants' names are pseudonyms.

<>	XML format for decoding information
	repetition of the same word
<pause />	longer pause
del	declarative
sp-act	speech act
"level"	uncompleted utterance
"stop"	utterance completed
"query"	a question
<turn />	the turn-taking to a new interlocutor

REFERENCES

- [1] Akmal, H., Syahriyani, A., & Handayani, T. (2022). Request speech act of Indonesian English learners and Australian English speakers through cross-cultural pragmatic perspectives. *LEARN Journal: Language Education and Acquisition Research Network*, 15(2), 498-520.
- [2] Austin, J. L. (1962). *How to do things with words*. Oxford University Press.
- [3] Budeng, R. B., & Merza, H. N. M. (2023). Assessing interlanguage pragmatic competence on speech acts in a Filipino EFL context. *Corpus Pragmatics*, 7(2), 85-102. <http://dx.doi.org/10.1007/s41701-023-00137-y>
- [4] Council of Europe. (2020). *Common European Framework of Reference for languages: Learning, teaching, assessment-companion volume*. Retrieved May 10, 2023, from *Council of Europe Publishing*. www.coe.int/lang-cefr
- [5] Fernández, J. (2022). Corpus linguistics in L2 pragmatics research. *Applied Pragmatics*, 4(2), 178-198. <https://doi.org/10.1075/ap.00008.fer>
- [6] Fernández, J. (2023). Pragmatics in learner corpora. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*, (pp. 1-6). Wiley.
- [7] Fernández, J., & Staples, S. (2021). Pragmatic approach. In N. Tracy-Ventura, & M. Paquot (Eds.), *Routledge handbook of second language acquisition and corpora* (pp. 240-251). Routledge.
- [8] Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, 1(1), 7-24. <http://dx.doi.org/10.1075/ijlcr.1.1.01gra>
- [9] Harris, D. W. (2019). Intention and commitment in speech acts. *Theoretical Linguistics*, 45(1-2), 53-67. <https://doi.org/10.1515/tl-2019-0004>
- [10] Hosoda, Y., & Aline, D. (2021). Deployment of I don't know and wakannai in second language classroom peer discussions. *Text & Talk*, 42(1), 27-49. <https://doi.org/10.1515/text-2019-0275>
- [11] House, J., & Kádár, D. Z. (2023). An interactional approach to speech acts for applied linguistics. *Applied Linguistics Review* [Online first], 1-21. <http://dx.doi.org/10.1515/applirev-2022-0116>
- [12] Huang, L. (2021). Toward multimodal corpus pragmatics: Rationale, case, and agenda. *Digital Scholarship in the Humanities*, 36(1), 101-114. <http://dx.doi.org/10.1093/llc/fqz080>
- [13] Khamkhien, A. (2022). Speech acts or speech act sets of refusals: Some evidence from Thai L2 learners. *Novitas-Royal*, 16(1), 97-121.
- [14] Mao, T. (2021). Investigation of L2/Ln pragmatic competence: Its core and route map. *Frontier in Psychology*, 12, 1-10. <https://doi.org/10.3389/fpsyg.2021.690550>
- [15] Mao, T., & He, S. (2021). An integrated approach to pragmatic competence: Its framework and properties. *SAGE Open*, 11(2), 1-13. <http://dx.doi.org/10.1177/21582440211011472>
- [16] McEnery, T., Brezina, V., Gablasova, D., & Banejee, J. (2019). Corpus linguistics, learner corpora, and SLA: Employing technology to analyze language use. *Annual Review of Applied Linguistics*, 39, 74-92. <http://dx.doi.org/10.1017/S0267190519000096>
- [17] Pan, Z. (2023a). Examination of data collection methods for pragmatic competence assessment of EFL learners. *International Journal of English Language and Literature Studies*, 12(2), 159-171. <http://dx.doi.org/10.55493/5019.v12i2.4798>

- [18] Pan, Z. (2023b). Formulaic language use by learners of English in interlanguage communication. *World Journal of English Language*, 13(7), 128-139. <https://doi.org/10.5430/wjel.v13n7p128>
- [19] Searle, J. (1979). *Expression and meaning studies in the theory of speech acts*. Cambridge University Press.
- [20] Staples, S., & Fernández, J. (2019). Corpus linguistics approaches to L2 pragmatics research. In N. Taguchi (Ed.), *The Routledge handbook of second language acquisition and pragmatics* (pp. 241-254). Routledge.
- [21] Taguchi, N. (2022). Second language pragmatics: A historical overview and future directions. In W. Ren (Ed.), *Second language pragmatics* (pp. 7-25). Cambridge University Press.
- [22] Tübben, I. T., & Landert, D. (2022). *Uh* and *um* as pragmatic markers in dialogues: A contrastive perspective on the functions of planners in fiction and conversation. *Contrastive Pragmatics*, 4(2), 350-381. <https://doi.org/10.1163/26660393-bja10049>
- [23] Verdonik, D. (2022). Annotating dialogue acts in speech data: Problematic issues and basic dialogue act categories. *International Journal of Corpus Linguistics*, 28(2), 144-171. <https://doi.org/10.1075/ijcl.20165.ver>
- [24] Weisser, M. (2018). *How to do corpus pragmatics on pragmatically annotated data: Speech acts and beyond*. John Benjamins.
- [25] Weisser, M. (2019a). *Manual for the Dialogue Annotation & Research Tool (DART)*. Retrieved March 1, 2023, from http://martinweisser.org/publications/DART_manual_v3.0.pdf
- [26] Weisser, M. (2019b). The DART annotation scheme: form, applicability & application. *Studia Neophilologica*, 91(2), 131-153. <https://doi.org/10.1080/00393274.2019.1616218>
- [27] Weisser, M. (2020). Speech acts in corpus pragmatics: Making the case for an extended taxonomy. *International Journal of Corpus Linguistics*, 25(4), 400-425. <https://doi.org/10.1075/ijcl.19023.wei>
- [28] Weisser, M. (2021). Profiling learners through pragmatically and error annotated corpora. In P. Pérez-Paredes, & G. Mark (Eds.), *Beyond the concordance: corpora in language education* (pp. 121-148). John Benjamins.
- [29] Yoon, S. (2020). The learner corpora of spoken English: What has been done and what should be done. *Language Research*, 56(1), 29-51. <https://doi.org/10.30961/lr.2020.56.1.29>
- [30] Yuan, W., & Lyu, S. (2022). Speech act matters: Commitment to what's said or what's implicated differs in the case of assertion and promise. *Journal of Pragmatics*, 191, 128-142. DOI: 10.1016/j.pragma.2022.01.012



Zhaoyi Pan holds a Ph.D. in applied linguistics from the English as an International Language Program, the Graduate School, Chulalongkorn University, Bangkok, Thailand. He is now working as an English specialist in the English Language Coordinating Section, the School of Information Technology, King Mongkut's University of Technology Thonburi. His research interests are interlanguage pragmatics, intercultural pragmatics, corpus linguistics, and learner corpus research.