

A corpus-based analysis of *so* in written discourse: A comparison between L1 English speakers and Japanese EFL learners

Shie Sato

Ritsumeikan University

This study examines the functional variability of *so* in essays written by 200 L1 English speakers (ENSs) and 400 Japanese EFL learners (EFLs). Using quantitative and qualitative approaches, this study elucidates discourse marker usage of *so* in each group, thereby establishing the normative patterns of use among ENSs and the features specific to L2 English writers. The findings suggest that ENSs use *so* strategically as a preface to stance-taking by carefully selecting and adjusting the information to be established as common ground with the reader. EFLs use *so* in a manner distinctly different from ENSs, displaying varying degrees of understanding and difficulty in utilising the word's discursual properties. The study concludes that it is important for L2 English learners to learn the uses of *so* not only as a connective marker with resultative meaning but also as a resource for projecting stance and assertion.

Keywords: *so*, discourse marker, L2 English writing, corpus-linguistics

1. Introduction

Research on the meanings and functions of *so*, particularly those focusing on its discourse marker usage, indicates that despite its general acknowledgement as a marker that connects causes to results, *so* is a fairly complex marker. The *Cobuild English Dictionary* (Sinclair, 1995), for instance, lists more than 10 usages of *so* beyond its fixed usages seen in phrases such as *and so on/forth* and *or so* (pp. 1581–1582). Quirk, Greenbaum, Leech, and Svartvik (1985) explain the uses of *so* via a continuum, along which *so* is placed between “coordinators” and “subordinators” as a “conjunct” (pp. 927–928). Since *so* is commonly used in both spoken and written English, the complexities involved in using *so* inevitably impose a challenge for learners of English as a second/foreign language.

This study examines the functional variability of *so* in written essays produced by 200 L1 English speakers (ENNs) and 400 Japanese EFL learners (EFLs). This study has three main purposes: (1) to describe the key discoursal factors motivating the use of *so* in L1 writing; (2) to elucidate the characteristics of L2 usage of *so*; and (3) to reveal the differences in usage patterns of *so* between the two groups. To augment methods used in previous research, the present study analyses L2 learners' written discourse (rather than their spoken discourse), incorporating both qualitative and quantitative analyses.

The remainder of this paper is organised in the following manner. Section 2 presents an overview of the major functions of *so* as identified in the literature. Section 3 outlines the methodological aspects of the present study, describing the data and the analytical framework employed. Section 4 provides the results of the study regarding the overall distribution of usage types of *so* in each of the two essay sub-corpora and the flow of discourse information achieved by the use of *so*. Section 5 presents the implications of the study, extending the discussion to some spoken language phenomena. Section 6 presents concluding remarks, key aspects of the findings that have implications for L2 pedagogy, and suggestions for areas of future explorations.

2. Literature review

2.1 Discourse marker functions of *so*

The major stream of research concerning *so* in modern English can be traced back to Halliday and Hasan (1976) on the types of cohesive relations in discourse and van Dijk (1979) on pragmatic connectives. In research from the 1980s and 1990s, much discussion on *so* is found in studies delineating theoretical frameworks for examining discourse markers in general (Fraser, 1988, 1990, 1999; Schiffrin, 1987; Schourup, 1999).¹ During the past two decades, the scope of research has been extended to the semantic and pragmatic functions of *so* in spoken discourse, and then to the interactional role of *so* in the sequential development of talk (e.g., Bolden, 2006, 2008, 2009; Johnson, 2002; Raymond, 2004). The latitude of scholastic attention to the usage of *so* has also expanded to L2 learners' language use and to the field of L2 pragmatics (Babanoğlu, 2014; Buysse, 2007, 2012; Fung & Carter, 2007; Lim, 2016; Müller, 2005; Romero Trillo, 2002; Vickov & Jakupčević, 2017).

1. One notable exception is Blakemore (1988), whose work is grounded in relevance theory (cf. Sperber & Wilson, 1986).

So has been assigned various functional labels. In addition to general labels such as a discourse marker or pragmatic marker, *so* is specified in the literature as a “resultative marker” (Quirk et al., 1985), a marker of “main idea units” (Schiffrin, 1987; see Müller, 2005, for a slightly modified interpretation), an “inferential discourse marker” (Fraser, 1996), a “topic developer” (Johnson, 2002), and a “speech act marker” (Müller, 2005).

The differences in the approaches and analytical frameworks concerning the usage of *so* have yielded various and often contrasting views on its core semantic meaning. Schiffrin (1987), for instance, claims that *so* conveys several meanings, including: (1) “result” at the level of ideational structure (fact-based), (2) “inference” concerning the information state (knowledge-based), and (3) “action” in view of the action structure (action-based) (p. 202). Other researchers employing the relevance-theoretic approach (e.g., Blakemore, 1988) assert that *so* contains only a procedural meaning with little conceptual value, thus serving as a “constraint on relevance” (p. 183).

Table 1 presents the major discourse marker functions of *so* identified in the literature (see Appendix for examples).

Due to the different approaches employed, the ways of classifying these functions into distinct types (textual, interactional, and/or procedural) vary depending on the researcher(s). For instance, Müller (2005) considers (1)–(5) to be textual functions, categorising “prefacing question or request” (9) and “indicating TRP” (10) as interactional types of phenomena.² Buisse (2012), on the other hand, presents a three-way distinction. The function of indicating a result or consequence belongs to the category of ideational function, whereas the functions of introducing a summary, returning to the main unit, and marking a boundary are classified as part of the textual function category. According to Buisse, the interpersonal function includes drawing a conclusion, prompting action by the recipient, and holding the floor.

Bolden (2006, 2008, 2009) offers a detailed analysis of *so* in interactions based on the conversation-analytic framework. She claims that *so* marks the ensuing course of action as “having been ‘on the speaker’s mind’ or ‘on agenda’ for some time” (Bolden, 2009, p. 976), thus often prefacing “new and resumed pending interactional agendas” (Bolden, 2009, p. 977). Bolden (2006) also claims that *so*

2. Müller (2005) offers an in-depth analysis of L2 English learners’ use of *so* in the Giessen-Long Beach Chaplin Corpus, generating 14 functions of *so* found in the speech of (non)native speakers: five non-discourse marker (grammatical) functions, five textual functions, and four interactional functions.

Table 1. Discourse marker functions of *so*

(1)	Indicate result or consequence Buyse, 2012; Fung & Carter, 2007; Lam, 2009, 2010; Lim, 2016; Liu, 2017; Müller, 2005; Schiffrin, 1987; van Dijk, 1979; Vickov & Jakupčević, 2017
(2)	Introduce summary Buyse, 2012; Fung & Carter, 2007; Johnson, 2002; Lam, 2010; Lim, 2016; Liu, 2017; Müller, 2005; Vickov & Jakupčević, 2017
(3)	Preface rewording Lim, 2016; Müller, 2005; Vickov & Jakupčević, 2017
(4)	Give an example Lam, 2009, 2010; Müller, 2005; Vickov & Jakupčević, 2017
(5)	Return to a main course of talk (Main unit marker) Buyse, 2007, 2012; House, 2013; Lam, 2009, 2010; Lim, 2016; Müller, 2005; Redeker, 2006; Schiffrin, 1987; Vickov & Jakupčević, 2017
(6)	Introduce a new topic Bolden, 2009; Buyse, 2012; Fung & Carter, 2007; Lam, 2009, 2010; Lim, 2016; Liu, 2017; Vickov & Jakupčević, 2017
(7)	Mark textual/sequential transition (Boundary marker) Chapetón Castro, 2009; Johnson, 2002; Lam, 2010; Redeker, 2006; Schiffrin, 1987
(8)	Draw conclusion Buyse, 2007, 2012; van Dijk, 1979
(9)	Preface question or request Johnson, 2002; Lam, 2010; Lim, 2016; Müller, 2005; Schiffrin, 1987; Vickov & Jakupčević, 2017
(10)	Indicate Transition Relevance Place (TRP) Buyse, 2007; Lam, 2009; Liu, 2017; Müller, 2005; Schiffrin, 1987
(11)	Prompt action by recipient Bolden, 2009; Buyse, 2012; Raymond, 2004
(12)	Hold the floor Buyse, 2012; House, 2013; Lam, 2009, 2010; Liu, 2017

often precedes a course of action that is “other-attentive,”³ that is, an action displaying the speaker’s awareness of or concern for the recipient’s interests. Thus, *so* not only helps to establish discourse coherence but also contributes to achieving mutual understanding (Bolden, 2009).

In summary, although previous studies on *so* indicate considerable variations in their theoretical standpoints and proposed classifications, the findings are consistent regarding the primary function of *so* in marking discursal and interactional transitions.

3. *So* is found to be used by English as a lingua franca (ELF) speakers “as a deictic element” serving to accomplish “self-attentive matters”. ELF speakers use *so* when they prepare for upcoming attempts and “as a means of ‘looking backwards’ summing up previous discourse stretches” (House, 2013, p. 62).

2.2 L2 learners' use of *so*

L2 learners' use of *so* can be discerned through research investigating discourse/pragmatic markers in general, as in the case of L1 usage. Only a small number of studies have been conducted solely on the use of *so* among L2 learners (e.g., Buysse, 2007, 2012; Lam, 2009; Lim, 2016). The overall use of discourse/pragmatic markers by L2 English learners is generally considered to be limited in terms of both frequency and variation in comparison to L1 English speakers,⁴ as the markers with less propositional content tend to receive little pedagogical emphasis in formal education (Fung & Carter, 2007; Müller, 2005; Romero Trillo, 2002). Claiming *so* to be one of those markers, some studies report that non-native English speakers employ *so* less frequently than native speakers (Fung & Carter, 2007; Müller, 2005; Ding & Wang, 2015). However, the opposite is also reported by Buysse (2012) and Hays (1992), both of whom show that, in L2 data, *so* behaves differently from other discourse markers such as *well* and *you know*. Buysse (2012), for instance, suggests that non-native English speakers tend to use *so* for interpersonal purposes (i.e., drawing a conclusion, prompting, holding the floor) markedly more often than native English speakers. According to Buysse, a similar tendency is observed for the textual functions of *so*; a group of non-native English speakers majoring in English Linguistics is found to use *so* in initiating elaboration 2.3 times more frequently than native speakers. Chapetón Castro's (2009) study examining classroom interaction between a non-native English teacher and EFL students shows that assigned roles in interaction (i.e., teacher and student) may also affect the frequency of *so*; a non-native teacher is found to use *so* significantly more often than EFL students.

In sum, the existing findings in the literature indicate manifold interactions between the extensive discourse/pragmatic functions fulfilled by *so*, ranging from referential, structural, and interpersonal functions and the types of data analysed (including participants' backgrounds and assigned roles in an interaction). Crucially, although the primary feature of *so* as marking a textual and/or interactional boundary is confirmed by previous studies, in-depth qualitative analyses of the types of information preceding and following *so*, both in L1 and L2 English speakers, have not yet been conducted, and the question as to when and how *so* should be used remains unanswered. The present study, therefore, investigates the types of discoursal information segmented by *so* and offers a framework for its usage, which will serve as a guideline for teaching L2 learners how to use *so* in writing. The present study addresses the following research question: what are the characteristics of L1 and L2 English speakers' use of *so* as found in the corpus of written essays?

4. See Hellermann & Vergun (2007) for details.

3. Methods

3.1 Data

The data were obtained from a section of the “Written Essays” component of the International Corpus Network of Asian Learners of English (ICNALE) (Ishikawa, 2013). A description of the sub-corpora is given in Table 2.

Table 2. Description of ENS and EFL sub-corpora from ICNALE

Corpus	Participants	Essays	Total number of words
ENS	200	400	90,613
EFL	400	800	179,042

This study analysed a total of 1,200 academic essays produced by 200 native speakers of English (ENSs; 400 essays) and 400 Japanese college students studying English (EFLs; 800 essays). For the ICNALE project, each participant submitted two essays advocating positions either for or against the following statements by providing specific reasons and examples:

Statement A: It is important for college students to have a part-time job.

Statement B: Smoking should be completely banned at all the restaurants in the country.

Each essay ranged from 200–300 words, and the time allotted for writing each essay was 20–40 minutes. Participants were not allowed to use dictionaries.

The EFL participants were divided into four proficiency levels based on their estimated Common European Framework of Reference for Languages (CEFR) level.⁵ As shown in Table 3, the majority of the participants (83.3%) fall in the A2 and B1_1 levels (elementary to lower-intermediate).

Table 3. Description of ICNALE participants⁶

Corpus	Country/Area	CEFR-based L2 proficiency levels			
		A2	B1_1	B1_2	B2+
ENS	USA, UK, Canada, Australia, New Zealand	N/A			
EFL	Japan	154 (38.5%)	179 (44.8%)	49 (12.3%)	18 (4.5%)

5. The CEFR levels were obtained based on the participants' scores on the TOEFL PBT/iBT, TOEIC, IELTS, STEP, TEPS, CET, or L2 vocabulary size test. (See <<http://language.sakura.ne.jp/icnale/index.html#6>> for more information.)

6. See Ishikawa (2013) for details.

For the ENS participants, no information was given concerning their L1 proficiency. Considerable variation was assumed to be present in the ENS group, however, as their occupational backgrounds were rather varied (e.g., students, teachers of English, and other occupations, such as translators, company employees, medical doctors, and financial/IT consultants). Based on the classifications given in the ICNALE, the ENS examples presented in the results section were specified as “Student” or “Teacher”.

3.2 Data analysis

First, using AntConc (Anthony, 2014), all instances of *so* in ICNALE were searched (ignoring case), which yielded 389 and 1,582 tokens in the ENS and EFL corpora, respectively. The tokens in each corpus were then manually classified into discourse marker and non-discourse marker types. The non-discourse marker usage, which was excluded from the present analysis, includes the following grammatical uses of *so*: (1) a conjunction phrase specifying purpose (e.g., *so that*), (2) an adverb (e.g., *Everybody was so nice*), (3) a pro-form (e.g., *I presume so*), and (4) a part of a fixed phrase (e.g., *so-called*). Of the 1,582 tokens of *so* in the EFL corpus, 10 tokens were excluded from the analysis for being both ungrammatical and unintelligible. This part of the analysis had to be performed manually (cf. Vickov & Jakupčević, 2017), as the process involved making grammaticality/intelligibility judgments and discriminating certain types of uses (e.g., *so* describing purpose often appeared without *that*, which could be misidentified as a discourse marker use).

After analysing the data quantitatively, qualitative analyses were conducted on all tokens of discourse marker uses of *so* in the ENS corpus (186 tokens), as well as on 300 tokens of *so* randomly selected from the EFL corpus. Each token was examined with reference to its adjacent and overall contexts, with a particular focus on the flow of information presented using *so*. The cases of *and so* in ICNALE (41 and 12 tokens from the ENS and EFL corpora, respectively) were analysed separately due to this phrase’s feature of containing two discourse markers.

7. According to the figures generated by the ICNALE Online, the frequency of use of *so* by the Japanese EFLs (B1_2 and B2+ only) in written essays is markedly higher than that by other EFL learners of the same levels in China, Indonesia, Korea, Thailand, and Taiwan. In fact, the learners in Taiwan seem to use *so* much less frequently than ENSs, which suggests that different approaches are required depending on the sub-group of L2 learners.

various quantifiers (e.g., *some, quite, many, none*) and modal verbs (e.g., *can, could, may*). Other means to enhance the acceptability of the premise by the reader are also sought, such as the use of the collective first pronoun *we*, which contributes to enhancing the sense of comradery with the reader. Hence, *so* has two primary functions. One is to retroactively confirm that the information presented in (A) is to be interpreted as common ground on which to present the writer's stance. The other function is to serve as a preface for the writer's act of stance-taking. Consider Examples (1)–(2):

- (1) ICNALE ENS 099 (Student)
 I disagree with this statement because I think that college students should spend all of their time studying and not worrying about other things. When we graduate from College we will be able to work for around 40 years *so* I think we should concentrate 100% on school.
- (2) ICNALE ENS 159 (Teacher)
 Science and statistics all around the globe have already unequivocally proven, beyond any reasonable doubt that smoking causes death, disease, heart failure, stroke and the list goes on. *So*, why do people still feel the need to cater for what was once a majority group but is becoming more and more the minority group that call themselves smokers.

Here, both the general fact in Example (1) and the scientific/statistical evidence in Example (2) serve as shared ground upon which the writers' stance is introduced. The analysis suggests that the property of *so* in which it segments the information indicating the writer's dissimilar interests, which first involves the reader and then displays stance, is essential.

A notable usage characteristic of *so* observed in the ENS corpus is that, when displaying stance, the ENSs often indicate a change in disposition (i.e., a shift in their focus of attention). As illustrated in Examples (3) and (4), this type of transition is often seen at the end of an essay when a concluding remark is given.

- (3) ICNALE ENS 032 (Student)
 It is important to spend as much time gaining as much knowledge as we can in our own fields so that we can be successful when we have to enter the real world and start working at companies. *So* no, I don't think college student should have a part-time job, and I think that it will be a while before I start working.
- (4) ICNALE ENS 060 (Student)
 Many people have some strange psychological fear of success, and they allow this to impact their career and life choices as well as their level of happiness. *So*, buck up, find a part-time job, and start working your way to happiness today.

In both examples, the writers' focus of attention shifts from explaining the premise to displaying stance with *so*.

The cases in which *so* prefates a rewording of or elaboration on the first portion of the sentence, withholding a move toward stance-taking, amount to 24.2% of the tokens in the ENS data (45/186). In such cases, *so* can be paraphrased as *in other words* or *that means*, with the function of introducing information that is logically plausible in light of what is stated in the first part ([A] in Figure 1). Example (5) illustrates this case.

- (5) ICNALE ENS 111: Rewording (Teacher)
 Finally, part-time work generally pays very little, *so* the actual financial gain from such work is almost negligible. His time is much better spent in academic and social activities in order to prepare him for his future social responsibilities.

The discourse structure depicted in Figure 1 also applies to the cases of *and so*, with the only difference being the type of information given in (B) (i.e., the writer's main assertion). It is found that *and so* is used when the writer's decisive stance is presented, as shown in Example (6).⁸ Here, the writer's conclusive stance ("family restaurants should ban smoking") is introduced by *and so*.

- (6) ICNALE ENS 008 (Student)
 Everyone already knows that smoking gives you cancer and everyone also knows that secondhand smoke is much more dangerous than first hand smoke. The toxins contained in second hand smoke are especially dangerous for young children *and so* at the very least family restaurants should ban smoking.

In sum, the discourse marker *so* serves two primary functions in ENS writing: *so* retrospectively defines the statement just presented as a shared premise, which appeals to the reader's understanding, while simultaneously prefating the writer's display of stance. The function of signalling a transition in the type of information and goals pursued is considered the primary feature of *so* observed in the ENS corpus. The types of transition include shifting viewpoints, such as changing from objective to subjective standpoints (Examples (1) and (2)) and robust stance shifting (Examples (3) and (4)). There are also instances where *so* serves as a simple turnover in the construction: *A is the case; in other words, B* (Example (5)). The cases in which *so* intercedes a *reason-result* sequence are not as predominant in the present data as generally expected.

8. In the majority of cases (32/41 tokens), *and so* appears in the middle part of the essays; only two and seven instances were found in the first and final statements, respectively.

4.3 Qualitative analysis of discourse marker *so* in EFL writing

As mentioned above, the use of *so* as a discourse marker in the EFL corpus accounts for nearly 72% of the total tokens, suggesting that it is the primary usage among EFLs. Although the same tendency is observed in the ENS corpus, with discourse marker use accounting for 48% of occurrences, this analysis has revealed two distinct features of *so* in the EFL data: (1) repeated (redundant) use of *so*, and (2) absence of information serving as a premise.

4.3.1 *Repeated (redundant) use of so*

One prominent feature of *so* in the EFL writing is that *so* appears repeatedly in close proximity, often with other conjunctions such as *therefore* and *but*. A typical case of this occurrence entails the writer's opinion being first introduced by *so* and then immediately followed by a sentence-initial *so*, resulting in the *so*-prefaced statement also serving as a premise for the next statement. This is illustrated in Example (7). Here, the statement introduced by the first *so* ("while they are college students they should learn about the work") is assigned the dual role of presenting the writer's opinion and serving as a premise for the next statement, which also begins with *so* ("I think they should have a part-time job").

- (7) ICNALE EFL 052 [A2 level]

My opinion is that it is important for high school students to study many subjects, which is Math, English, Science, Japanese, Society, etc. Therefore high school students should not work a part-time job. But college students must work after graduating a college, *so* while they are college students they should learn about the work. *So* I think they should have a part-time job. They have a part-time job and they get some money...

In the next example, *so* appears with *but* in close proximity, affecting the consistency of the main assertion as well as the level of formality. Using *so*, the writer first argues that both high school and college students should have a part-time job, but later s/he uses *so* again, this time to make a claim in partial contradiction to the original argument ("high school students do not have to do it").

- (8) ICNALE EFL 062 [B1-1 level]

Third, college students are independent from their parents. In Japan, college and high school are not obligation. *So* college and high school students should earn money themselves [*sic*]. *But* many high school ban to do part-time job. Students cannot do it. *So* high school students do not have to do it. *But* I have not heard a college which ban to do part-time job. *So* they have to work to earn their money. That is why I agree with the statement. [End of essay]

As shown here, the repeated uses of *so* with *but* intervening result in the recurrent changes in the course of argument. This type of usage is common among EFLs at the A2 and B1_1 levels; no such case appears in the essays produced by the B2+ participants.

Situations in which *so* is simply redundant and cases where *so* can be replaced by other expressions or conjunctions (e.g., *and*) are also prevalent in the EFL corpus. Example (9) is a representative case of redundant usage. Here, the overall flow of information can be markedly improved by deleting the entire statement that includes *so*:

- (9) ICNALE EFL 137 [B1–2 level]
I don't agree with this opinion, and I have couples of reasons why I don't. *So* I will show you what they are like. First of all,...

Example (10) illustrates the cases in which *and* is preferable to *so*:

- (10) ICNALE EFL 079 [A2 level]
Third, we can make friends. There are many people in the working place, *so* we can be good terms of them. However, I don't have any part-time job now. I'm too busy to have part-time job.

There are other cases where *so* is best replaced by other expressions such as *similarly* or *that way*. The use of *so* in place of *because* (e.g., “But they must not work too much, *so* their object is to learn much technical knowledge”) is considered a problem specific to beginning level learners (e.g., A2).

4.3.2 *Absence of information serving as a premise*

Another tendency found in the EFL data is the absence of a premise, lending itself to a configuration different from that formed by ENSs (see Figure 1). Consider Example (11):

- (11) ICNALE EFL 098 [A2 level]
Fourth, you can contribute to the society. If you do part-time job, the economy become more active. Conversely, I think also a disadvantage. If you do part-time job, your study time reduce. But you are only careful that situation. *So* I agree this statement.

In this example, the EFL writer agrees with the importance of having a part-time job but also mentions the risk of having less time for studying. At the end of the text, s/he agrees with the proposition, using *so* at sentence-initial position (“*So* I agree this statement”). However, the premise of this sentence (i.e., what s/he agrees with) seems to include many of the previous statements and is thus unclear; as a result, the *so*-prefaced sentence appears as a sudden shift in position. When

the initial *so* is used by ENSs to summarise what has been presented, it tends to be followed by a reiteration of the writer's opinion, not a disclaimer of agreement or disagreement.

Example (12) from the EFL corpus exhibits an information structure opposite to that used by the ENSs; that is, *so* is preceded by the writer's stance ("I agree that ...") and followed by a fact ("I don't receive remittance from my parents"), which would better function as a premise.

- (12) ICNALE EFL 092 [A2 level]
 I agree that it is important for college students to have a part-time job, because some of them may need it to earn money for the cost of living and school expenses. *So* I don't receive remittance from my parents, I intend to do it, too. In addition, we should experience the rule of society.

In Example (13), *so* appears between the statements, with each statement expressing the writer's opinions on different issues:

- (13) ICNALE EFL 174 [B1-1 level]
 Today, campaign against smoking is expanding gradually in Japan. But I can often see smoker's bad manner about smoking. For example, throw a cigarette butt out on the street, smoking in non-smoking area, and so on. Smoker should have a good manner about smoking, and get along with non-smoker. *So*, smoking should not be completely banned. [End of essay]

In this example, *so* is preceded by the summary statement about the smokers' manners ("Smoker should have a good manner about smoking ..."), which does not sufficiently serve as a basis for asserting the conclusive opinion that follows ("Smoking should not be completely banned").

Finally, the present analysis has revealed that the use of *and so* is extremely rare among the EFLs (12/1,131 tokens). Except for the two instances produced by A2 learners, all the tokens are found in B1_1 level essays. Notably, only three cases are found to be comparable to those observed in the ENS corpus; the remaining nine tokens demonstrate the features described above that are specific to the EFL writers.

To summarise, the EFL learners' use of *so* exhibits the following features: (1) The writer presents his/her stance with *so* without establishing common ground with the reader; (2) *So* is preceded by the writer's stance and followed by common ground (in the reverse order to that shown in Figure 1); and (3) *So* appears between the writer's stance-taking comments on two different issues. These uses of *so* observed in the EFL corpus can have powerful effects on the flow of discourse, often making parts of or entire essays vague, ambiguous, or even contradictory to the writer's main assertion.

5. Discussion

The findings in this study elucidate the overall distribution of *so* and the structure of information surrounding its discourse marker uses in ENS and EFL corpora. In most cases of the discourse marker uses of *so* by ENSs, *so* appears between the information the writer intends to establish as shared knowledge with the reader and the statement expressing his/her stance. This indicates that although the recognition of *so* as a marker that connects causes and results is essential, L2 English learners also need to understand the structures of information achieved by *so* in order to become competent English writers.

Such understanding of *so* is particularly useful for Japanese EFL learners because the Japanese language has several comparable linguistic options, the meanings of which are relatively equivalent to those of *so*: (*da*)*kara* (*so*; *therefore*), *sorede* ([*and*] *so*), and (*na*)*node* (*therefore*; *so*). The extensive use of *so* as a discourse marker by Japanese EFLs observed in this study can be partially explained by L1 transfer, in that the aforementioned related Japanese forms cover different functional spheres of *so* relative to their English counterparts. For instance, (*da*)*kara* (*so*; *therefore*) can be used to indicate both logical and causal relationships between two statements (cf. Matsui, 2002),⁹ while *sorede* manages connective functions entailing a topic shift and the signalling of a topic resumption (Ito, 1995; as cited in Sadler, 2006). The functional variability of *so*, therefore, could be perceived by EFLs to be more diverse and extensive than it actually is, leading to their extensive use of *so*.

In relation to the issue of EFL perception of the functional variability of *so*, one characteristic use of *so* among EFLs that merits pedagogical attention is learners' tendency to assign a robust linking function to *so*. While ENSs use *so* as a preface to claiming stance, keeping the act relevant to the premise just presented, EFLs seem to use *so* as having an encompassing role of presenting deductions and implications stemming from the previous statement(s). In the present study, the use of *so* in a way that overestimates its linking function is pervasive among many EFLs, irrespective of their proficiency levels.

Another important point is that the basic framework for discourse marker usage of *so* in writing (Figure 1) may be relevant to the phenomena observed in spoken language as well.¹⁰ Although *so* exhibits uses in speech, particularly in naturally occurring interactions in which functions of *so* are more diverse than those

9. *Dakara* is similar to *so*, in that it is used at the utterance-initial position to self-repair one's utterance (cf. Maynard, 1993; Matsui, 2002).

10. The uses of *so* could be sufficiently versatile in spoken discourse such that the current findings might apply only to limited cases. For instance, the consecutive use of *so* identified as inappropriate in written discourse is acceptable in speech, as long as the second *so* is used to shift the level of discourse up to a concluding remark, as in the following example:

in written discourse, some cases of *so* in spoken discourse can be explained by the information structure shown in Figure 1. One such case is the usage of *so* at utterance-initial position as a conversation initiation device, as seen, for example, in the speech of a grandmother, “*So*, tell me about this wonderful young man you’re seeing” (Fraser, 1990), or a parent addressing a child, “*So*, you did it again?” (See Bolden, 2008, for examples in interactions). Even in such cases where *so* is not preceded by any statement, the use of *so* is grounded in the shared knowledge that the speaker presumes the interactant possesses before initiating the exchange (See Howe, 1991, for his definition of *so* as a “marker of connection” in this respect). In other words, *so* serves as a prompt, directing the interactant to establish a meaningful relationship between then-existing common knowledge (A) and the ensuing comment indicating the speaker’s stance (B). The information (A), therefore, could be of a contextual or experiential type, or it could be formed with combined clues. The act of stance-taking is executed robustly in those cases, often taking the form of a question or directive.

The structure of information presented in Figure 1 could also help explain the inappropriacy of the common practice among Japanese EFL learners of using *so* in the utterance-initial position. Whereas ENSs employ *so* at the utterance-initial position to allude to the knowledge they assume is shared by their interactant, Japanese EFL learners seem to use the utterance-initial *so* in a self-directed manner in contexts where no such shared understanding exists. The judgment of inappropriacy can be ascribed to the burden incurred on the part of the interactant to search for contextual cues to infer a shared understanding which does not exist. Accordingly, the message followed by *so* will be interpreted as pragmatically vacuous, thereby increasing the chances of the speaker’s message being misinterpreted and misunderstood. From a learner’s perspective, being able to use *so* as a conversation initiation device is particularly challenging, as it involves the tasks of processing contextual cues commonly held prior to the interaction and robustly displaying stance while maintaining the connection that *so* has established with the then-held shared knowledge.

[The Corpus of Contemporary American English] (160419)

I grew up on 11th Street and 2nd Avenue, and I was raised by a woman from the Scottish countryside who believed in education as the way that a person advances in life. And that’s what I was instilled with. *So* from a very young age, it was important to me that I do well in school and that I attend a prestigious institute of higher learning. *So* I thank my mother for, you know ...

<<http://corpus.byu.edu/>>

A comprehensive study based on empirical data – for targeting all types of usage – is necessary in the future to fully uncover the basic framework of using *so* in spoken discourse.

6. Conclusion and implications

The present study explored the features of L1 English speakers and L2 English learners with respect to their use of *so* and its core functions, including the information structure that *so* construes in written discourse. The findings substantiate the primary claim made in the existing literature that *so* marks a transition point by offering a qualitative account of the discourse organisation. The analysis of ENS data demonstrates that *so* tends to occur at the point of interface between the information that the writer intends to establish as common ground with the reader and the statement explicating his/her stance. In other words, the segmentation of the writer's focus of attention is essential when using *so* as a discourse marker, although other cases of *so* in which the flow of information motivates or warrants its resultative usage are not uncommon.

The features of EFL learners' use of *so* identified in this study provide some pedagogical implications. First, where concerns teaching discourse marker uses of *so*, all activities, including mechanical ones, should be conducted at the paragraph level or higher. The instruction of *so* requires a different approach from that of other coordinate conjunctions (e.g., *and*, *but*, *or*, *yet*), in that in-depth discussion of the information structure rendered by *so* is necessary. Second, the basic framework of usage (Figure 1) can be extended from written to spoken discourse, serving as an effective means to teach the interactional functions of *so* specific to spoken discourse. Finally, classroom activities should be structured such that they reflect learners' backgrounds, including their L1s and proficiency levels. As far as Japanese EFLs are concerned, analysing appropriate use of *so* in a relatively long text extracted from corpus-based authentic discourse can be useful.

It is hoped that the findings in this study will promote a valuable pedagogical approach that goes beyond the traditional grammar-based approach, incorporating the actual viewpoint of the writer, such as his/her perception about the knowledge possessed by the reader in text. Further empirical research is needed, which will help learners of English successfully express their views in writing by integrating the reader's knowledge and viewpoints. Possible avenues for future research include comparing uses of *so* between L1 English speakers and L2 English learners in both monologic and interactional spoken discourse and analysing the uses of *so* among learners of various L2 backgrounds and proficiency levels. Further, investigations of other discourse markers involving extensive qualitative analyses to define their structure of information flow are indispensable to help L2 learners enhance their overall pragmatic competence.

References

- Anthony, L. (2014). AntConc (Version 3.4.4) [computer software]. Tokyo, Japan: Waseda University. Available from <<http://www.laurenceanthony.net/software>>
- Babanoğlu, P. M. (2014). A corpus-based study on the use of pragmatic markers as speech-like features in Turkish EFL learners' argumentative essays. *Procedia – Social and Behavioral Sciences*, 136, 186–193. <https://doi.org/10.1016/j.sbspro.2014.05.312>
- Blakemore, D. (1988). "So" as a constraint on relevance. In R. Kempson (Ed.), *Mental representation: The interface between language and reality* (pp. 183–195). Cambridge: Cambridge University Press.
- Bolden, G. B. (2006). Little words that matter: Discourse markers "so" and "oh" and the doing of other-attentiveness in social interaction. *Journal of Communication*, 56, 661–688. <https://doi.org/10.1111/j.1460-2466.2006.00314.x>
- Bolden, G. B. (2008). "So what's up?": Using the discourse marker "so" to launch conversational business. *Research on Language & Social Interaction*, 41(3), 302–327. <https://doi.org/10.1080/08351810802237909>
- Bolden, G. B. (2009). Implementing incipient actions: The discourse marker "so" in English conversation. *Journal of Pragmatics*, 41, 974–998. <https://doi.org/10.1016/j.pragma.2008.10.004>
- Buyse, L. (2007). Discourse marker *so* in the English of Flemish university students. *Belgian Journal of English Language and Literatures*, 5, 79–95.
- Buyse, L. (2012). *So* as a multifunctional discourse marker in native and learner speech. *Journal of Pragmatics*, 44, 1764–1782. <https://doi.org/10.1016/j.pragma.2012.08.012>
- Chapetón Castro, C. M. (2009). The use and functions of discourse markers in EFL classroom interaction. *Profile*, 11, 57–77. Available from: <http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1657-07902009000100005&lng=en&tlng=en>
- Ding, R., & Wang, L. (2015). Discourse markers in local and native English teachers' talk in Hong Kong EFL classroom interaction: A corpus-based study. *International Journal of Language and Linguistics*, 2(5), 65–75.
- Fraser, B. (1988). Types of English discourse markers. *Acta Linguistica Hungarica*, 38(1–4), 19–33.
- Fraser, B. (1990). An approach to discourse markers. *Journal of Pragmatics*, 14, 383–395. [https://doi.org/10.1016/0378-2166\(90\)90096-V](https://doi.org/10.1016/0378-2166(90)90096-V)
- Fraser, B. (1996). Pragmatic markers. *Pragmatics*, 6(2), 167–190. <https://doi.org/10.1075/prag.6.2.03fra>
- Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, 31, 931–952. [https://doi.org/10.1016/S0378-2166\(98\)00101-5](https://doi.org/10.1016/S0378-2166(98)00101-5)
- Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and non-native use in pedagogic settings. *Applied Linguistics*, 28, 410–439. <https://doi.org/10.1093/applin/amm030>
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Hays, P. R. (1992). Discourse markers and L2 acquisition. *Papers in Applied Linguistics*, 7, 24–34.
- Hellermann, J., & Vergun, A. (2007). Language which is not taught: The discourse marker use of beginning adult learners of English. *Journal of Pragmatics*, 39, 157–179. <https://doi.org/10.1016/j.pragma.2006.04.008>
- House, J. (2013). Developing pragmatic competence in English as a lingua franca: Using discourse markers to express (inter)subjectivity and connectivity. *Journal of Pragmatics*, 59, 57–67. <https://doi.org/10.1016/j.pragma.2013.03.001>

- Howe, L. (1991). Topic change in conversation (Unpublished doctoral dissertation). University of Kansas.
- Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world* 1, 91–118. Kobe: Kobe University
- Johnson, A. (2002). So.?: Pragmatic implications of so-prefaced questions in formal police interviews. In J. Cotterill (Ed.), *Language in the legal process* (pp. 91–110). New York: Palgrave Macmillan. https://doi.org/10.1007/978-0-230-52277-0_6
- Lam, P. W. Y. (2009). The effect of text type on the use of *so* as a discourse particle. *Discourse Studies*, 11, 353–372. <https://doi.org/10.1177/1461445609102448>
- Lam, P. W. Y. (2010). Toward a functional framework for discourse particles: A comparison of *well* and *so*. *Text and Talk*, 30(6), 657–677. <https://doi.org/10.1515/text.2010.032>
- Lim, J. (2016). Discourse marker *so*: A comparison between English language Learners and English-dominant speakers (Unpublished MA thesis). University of Toronto, Toronto.
- Liu, B. (2017). The use of discourse markers *but* and *so* by native English speakers and Chinese speakers of English. *Pragmatics*, 27(4), 479–506. <https://doi.org/10.1075/prag.27.4.01liu>
- Matsui, T. (2002). Semantics and pragmatics of a Japanese discourse marker *dakara* (*so/in* other words): A unitary account. *Journal of Pragmatics*, 34(7), 867–891. [https://doi.org/10.1016/S0378-2166\(01\)00066-2](https://doi.org/10.1016/S0378-2166(01)00066-2)
- Maynard, S. (1993). *Discourse modality: Subjectivity, emotion and voice in the Japanese language*. Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.24>
- Müller, S. (2005). *Discourse markers in native and non-native English discourse*. Amsterdam, Netherlands: John Benjamins. <https://doi.org/10.1075/pbns.138>
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. New York: Longman.
- Raymond, G. (2004). Prompting action: The stand-alone “*so*” in ordinary conversation. *Research on Language and Social Interaction*, 37(2), 185–218. https://doi.org/10.1207/s15327973rlsi3702_4
- Redeker, G. (2006). Discourse markers as attentional cues at discourse transitions. In K. Fischer (Ed.), *Approaches to discourse particles* (pp. 339–358). Amsterdam: Elsevier.
- Romero Trillo, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics*, 34(6), 769–784. [https://doi.org/10.1016/S0378-2166\(02\)00022-X](https://doi.org/10.1016/S0378-2166(02)00022-X)
- Sadler, M. (2006). A blurring of categorization: The Japanese connective *de* in spontaneous conversation. *Discourse Studies*, 8(29), 303–323. <https://doi.org/10.1177/1461445606055425>
- Schiffrin, D. (1987). *Discourse markers* (Studies in interactional sociolinguistics). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511611841>
- Schourup, L. (1999). Discourse markers. *Lingua*, 107, 227–265. [https://doi.org/10.1016/S0024-3841\(96\)90026-1](https://doi.org/10.1016/S0024-3841(96)90026-1)
- Sinclair, J. (1995). *Collins COBUILD English dictionary*. London: Harper Collins.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.
- van Dijk, T. (1979). Pragmatic connectives. *Journal of Pragmatics*, 3, 447–456. [https://doi.org/10.1016/0378-2166\(79\)90019-5](https://doi.org/10.1016/0378-2166(79)90019-5)
- Vickov, G., & Jakupčević, E. (2017). Discourse markers in non-native EFL teacher talk. *Studies in Second Language Learning and Teaching*, 7(4), 649–671. <https://doi.org/10.14746/sslt.2017.7.4.5>

Appendix. Discourse marker functions of *so* – Examples

(1) Indicate result or consequence

I was sick, **so** I stayed in bed.

(van Dijk, 1979, p. 453)

(2) Introduce summary

Sherry: and what made you choose UF?

Davy: uh two reasons, one of which was.. because I I didn't know specifically what I wanted to do. I knew I wanted to be in Discourse. And I saw at that time there were two people here that did that. (...) Also the issue of financial aid. Because I was accepted to, accepted to one of the universities in Texas I think. (...) Uh but they told me that, they wouldn't tell me about the financial aid. Because UF gave me a deadline, you know. (...) **So** those two things sort of were were reasons why I came here.

(Liu, 2017, p. 489)

(3) Preface rewording

Liv: I was in the like the red corral

Int.: I don't know what the red corral is

Liv: that's the first corral

Int.: oh [oh oh okay

Liv: [so I was in the first heat.

(Lim, 2016, p. 52)

(4) Give an example

T4: **So**, deception is like, let's say a lie, but um maybe in a more physical sense ...

(Vickov & Jakupčević, 2017, p. 664)

(5) Return to a main course of talk (Main unit marker)

T1: So, these are the main parts that you need to focus on before you put your um answer um the missing paragraph, right? **So**, the main ideas that they want to present and are usually put here towards the end of the paragraph.

(Vickov & Jakupčević, 2017, p. 664)

(6) Introduce a new topic

So how was your day?

(Sinclair, 1995, p. 1582)

(7) Mark textual/sequential transition (boundary marker)

TT: Excuse me one second because (.) I know where I have the marker. (xxx)

SO↓remember, noche vieja, would be (-)

(Chapetón Castro, 2009, p. 66)

(8) Draw conclusion

A: I am busy.

B: **So**, you are not coming tonight?

A: I'm sorry.

(van Dijk, 1979, p. 454)

(9) Preface question or request

DC: Do you want to draw the house then?

(...Dialogue continues prompting the drawing of the house and its rooms.)

DC: Right. So can you tell me who sleeps in what bedroom then?

(Johnson, 2002, p. 103)

(10) Indicate Transition Relevance Place (TRP)

Sherry: you've never watched TV?

John: I haven't.. recently within, you know, I've flipped a couple channels, but decided nothing was on and turned it off. And other than that, not having TV saves me whatever thirty dollars a month and however many hours a week. So

Sherry: when you were growing up, did you have a favorite program?

(Liu, 2017, p. 491)

(11) Prompt action by recipient

Shi: Ri[ght.

Ger: [So

Shi: .hh So he's doing alright.

(Raymond, 2004, p. 198)

(12) Hold the floor

a: you don't have to have heels that are now in is the so erm I mean I'm not I'm not wearing heels...

(Lam, 2010, p. 669)

Address for correspondence

Shie Sato

Ritsumeikan University

56-1 Toji-in Kitamachi Kita-ku Kyoto 603-8577

Japan

shie@fc.ritsumei.ac.jp