Non-Discourse Marker Uses of *So* in EFL Writings: Functional Variability among Asian Learners

Shie Sato

(Ritsumeikan University)

This paper examines the frequency and distribution of the so-called "non-discourse marker functions" of *so* in essay writings produced by 200 L1 English speakers and 1,300 EFL learners in China, Japan, Korea, and Taiwan. Based on the data drawn from the International Corpus Network of Asian Learners of English, this study compares EFL learners and L1 English speakers' uses of *so*, identifying four grammatical uses, as (1) an adverb, (2) part of a fixed phrase, (3) a pro-form, and (4) a conjunction phrase specifying purpose. This study aims to show the wide variability among EFL learners with different L1s, identifying the tendency of usage both common among and specific to the sub-groups of EFL learners. The findings suggest that the learners demonstrate patterns distinctively different from those of L1 English speakers, indicating an underuse of *so* as a marker expressing "purpose" and an overuse as part of fixed phrases. Compared to L1 English speakers, the learners also tend to overuse *so* in the discourse marker functions, regardless of their L1s. The study proposes pedagogical implications focusing on discourse flow and diachronic aspects of *so* in order to understand its multifunctionality, although the latter is primarily suggested for advanced learners.

Keywords: *So*, L2 Writing, EFL Learner, Corpus Linguistics, Non-discourse Marker

1. Introduction

So is commonly known as a "discourse marker" or "pragmatic marker" that connects causes to results, as in "I was sick, *so* I stayed in bed" (van Dijk, 1979, p. 453). More specific terms in the literature include a "resultative marker" (Quirk et al., 1985), a marker of "main idea units" (Schiffrin, 1987), an "inferential discourse marker" (Fraser, 1996), and a "topic developer" (Johnson, 2002). As these terms suggest, functions of *so* are wide-ranging, encompassing those of textual, interactional, and/or procedural realms (cf. Bolden, 2009; Buysse, 2012; Fung & Carter, 2007; Lam, 2009, 2010; Liu, 2017; Müller, 2005; Schiffrin, 1987; Vickov & Jakupčević, 2017).

As for the treatment of *so* in grammar books, major grammar books provide extensive coverage of the diverse functional dimensions of *so*. According to Quirk et al. (1985), *so* is classified as a conjunct sharing the functional qualities with those of "coordinators" and "subordinators" (pp. 927-928). Quirk and his colleagues explain that *so* is different from other conjuncts, such as *however* and *therefore*, in that it primarily appears in a clause-initial position and connects clauses, predicates, and other clause elements (p. 927). Further, they claim that *so* closely resembles coordinators (e.g., *and, or*) in terms of its position and linking functions but differs critically in its inability to link more than two clauses and subordinate clauses. Huddleston and Pullum (2002) provide a thorough account on the versatile functions of *so* in several sections. Larsen-Freeman and Celce-Murcia (2016) offer a concise but thorough overview, encompassing the functions of *so* as a resultative and inferential marker (cf. Blakemore, 1988) as well as its non-discourse marker functions. An overview of the different functions of *so* is also provided by Biber et al. (1999), such as the forms of "*so*+adverb+*that*-clause,"

"*so*+adverb+*as to*-clause" (p. 550), *so that* (p. 844), and a pro-form substitution (pp. 72, 751-753, 917) as well as the function as a linking adverbial (pp. 877, 886-887, 1074). Biber et al. provide further information as to the English varieties, which indicates that *so* is more commonly used in conversations in American English than British English (pp. 562, 565-566).

2. Background

A wealth of research has dealt with the discourse marker (DM) functions of *so* in the past, proposing varied classifications based on different theoretical frameworks. For instance, Buysse (2012), based on informal interview data, classifies the DM functions of *so* into ideational, interpersonal, and textual types, as shown in Table 1.

Table 1. Discourse Marker Functions of *So*

Types of Relation	Discourse Marker Function
Ideational	Indicate a result
Interpersonal	Draw a conclusion
	Prompt
	Hold the floor
Textual	Introduce a summary
	Introduce a section of the discourse
	Indicate a shift back to a higher unit of the discourse
	Introduce a new sequence
	Introduce elaboration
	Mark self-correction

(Buysse, 2012, p. 1767)

Despite differences in theoretical standpoints, the findings of previous studies are fairly consistent with respect to the primary function of *so* to mark discoursal and interactional transitions. In the realms of L2 research, many studies indicate that L2 learners employ the DM *so* less frequently than L1 English speakers (cf. Ding & Wang, 2015; Fung & Carter, 2007; Hellermann & Vergun, 2007; Müller, 2005), presumably owing to the complications grounded in its multifunctionality. However, a few studies claim the opposite, that L2 learners use *so* more readily than L1 English speakers when expressing interpersonal purposes—e.g., drawing a conclusion, prompting, and holding the floor (Buysse, 2012). Sato (2019) also reports that Japanese EFL learners use DM *so* notably more often than L1 English speakers in essay writings.

Besides these DM functions, those largely determined by grammar are classified as non-DM functions. Table 2 presents the representative functions identified by Müller (2005).

Table 2. Non-discourse Marker Functions of *So*

(1)	adverb of degree or manner
(2)	expressing purpose (e.g., ***so*** (that))
(3)	*so* in fixed expressions (e.g., and ***so*** on, or ***so***)
(4)	*so* as a substitute

Contrary to the DM uses of *so*, the non-DM uses have little in common in terms of meaning, function, or use; therefore, the literature treats the non-DM functions as only supplementary to the DM uses. The major findings from L2 research mainly come from Müller (2005), who showed that native speakers of German have a marked tendency to use *so* in fixed expressions that are, in fact, barely used by L1 English speakers. Other results reported include a mild tendency of German native speakers to use *so* more frequently for the "purpose" meaning and as a pro-form, which

predominantly takes the form of *I think so*. However, the only notable statistical significance was detected in the difference in usage of fixed expressions.

It should be acknowledged that the distinction between DM and non-DM functions presents a major issue in analyzing *so*, raising questions as to its binary treatment. Although the essential features of DMs can be described using Schourup's (1999) terms of "connectivity," "optionality," "non-truth-conditionality," etc., the definitions of DMs are subject to various interpretations.[1] Furthermore, each individual criterion has to be tested individually against each case, as some cases of the non-DM function of adverbs of degree/manner, for example, meet the DM criterion of "optionality." The function of expressing "purpose" presents the most perplexing case among all the non-DM uses since it fulfills all the criteria of DMs. While fully acknowledging the complicated issues surrounding the definition of discourse markers, the present paper employs Müller's (2005) view of non-DM functions with slight modifications.

Given the degree of complexity involved in the use of *so* and the dearth of previous studies, the present study compares L1 English speakers' non-DM uses of *so* with that of Asian EFL learners in China, Japan, Korea, and Taiwan. An overview of the literature reveals that among the limited number of existing studies of the uses of *so* by L2 learners (cf. Buysse, 2007, 2012; Lam, 2009), to my knowledge, the only study that covers both DM and non-DM functions of *so* is the one conducted by Müller (2005). This study of L2 learners' non-DM uses of *so* will complement the existing literature by contributing toward a more complete picture of the uses and functions of *so*.

2.1. Research Questions

The present study aims to answer the following research questions: (1) To what extent do L1 English speakers utilize the non-DM *so* in written discourse?; and (2) Which non-DM usage patterns of *so* are common among and specific to sub-groups of L2 learners?

Regarding the methodological aspect of comparing L1 and L2 outputs and addressing the differences between the two, the present study accepts the validity of claims that the definitions of a "native speaker" or an "L1 English speaker" are still under debate and that the linguistic competence of so-called L1 English speakers shows significant variation. That said, this study assumes that the outstanding features attested in certain groups of L2 learners may well indicate the difficulties they face in learning English, thus presenting areas for improvement.

3. Methodology

3.1. Data

The following data were derived from the "Written Essays" component in the International Corpus Network of Asian Learners of English (ICNALE). The essays examined were produced by 200 L1 English speakers (400 essays) and 1,300 EFL learners (2,600 essays).

---

[1]   The widely-held view toward DMs is that DMs prototypically (1) connect utterances or other discourse elements ("connectivity"), (2) do not affect the grammaticality of the sentence when omitted ("optionality"), (3) do not alter the truth-conditions of the utterance proposition ("non-truth-conditionality"), (4) hold few or no relations with the other syntactic elements ("weak clause association"), (5) appear in initial position ("initiality"), (6) occur primarily in spoken discourse ("orality"), and (7) do not belong to a specific syntactic category ("multi-categoriality") (Schourup, 1999, pp. 230-234).

Table 3. ICNALE Data[2]

Code	Participants	Essays	Total number of words
ENS	200	400	90,613
CHN	400	800	191,405
JPN	400	800	179,042
KOR	300	600	136,344
TWN	200	400	92,384

For the ICNALE project, each participant submitted two essays advocating positions either for or against the following statements by providing specific reasons and examples: (A) It is important for college students to have a part-time job; and (B) Smoking should be completely banned at all the restaurants in the country. The time allotted for writing each essay (ranging from 200 to 300 words) was 20–40 minutes, and the use of a dictionary was not allowed. As shown in Table 4, the L2 English learners were divided into four proficiency levels based on their estimated Common European Framework of Reference for Languages (CEFR) levels.

Table 4. ICNALE EFL Participants

Corpus	Country/Area	CEFR-based L2 Proficiency Levels			
		A2	B1_1	B1_2	B2+
CHN	China	50 (12.5%)	232 (58.0%)	105 (26.2%)	13 (3.3%)
JPN	Japan	154 (38.5%)	179 (44.8%)	49 (12.2%)	18 (4.5%)
KOR	Korea	75 (25.0%)	61 (20.3%)	88 (29.4%)	76 (25.3%)
TWN	Taiwan	29 (14.5%)	87 (43.5%)	61 (30.5%)	23 (11.5%)

The participants in the English Native Speakers (ENS) group were L1 English speakers from the USA, the UK, Canada, Australia, and New Zealand. No information was given concerning their L1 proficiency; however, their occupational backgrounds (e.g., students, teachers of English, translators, company employees, medical doctors, and financial/IT consultants) suggested that they all had a relatively high degree of literacy, even if some were more likely to be conscious of language uses than others.

3.2. Analysis

All instances of *so* in ICNALE were first searched using AntConc (3.4.4),[3] yielding the following tokens for each corpus: 389 tokens (ENS), 1,229 tokens (CHN), 1,582 tokens (JPN), 823 tokens (KOR), and 429 tokens (TWN). The tokens in each corpus were then manually classified into DM use and four different types of non-DM uses: (1) an adverb (e.g., *She is **so** nice*); (2) a part of a fixed phrase (e.g., *and* **so** *on, or* **so**, **so** *far*); (3) a pro-form (e.g., *I presume **so***); and (4) a conjunction phrase specifying purpose (e.g., **so** *that,* **so** *as to*). The category of *so* indicating "purpose" was broadened to include *so as to.* Of the total tokens obtained in the search, repeated cases and tokens judged to be hardly intelligible were excluded from the study. Infelicitous uses of *so*, both grammatically and pragmatically, were included, as they can shed light on the learners' deficiencies in the use of non-DM *so*. The excluded tokens were: 6 tokens (CHN), 10 tokens (JPN), 60 tokens (KOR), and 4 tokens (TWN). The high number of eliminated tokens in KOR (60) comes from duplicated data, excluded as repetition. The total number of tokens examined amounted to 389 tokens (ENS), 1,223 tokens (CHN), 1,572 tokens (JPN), 763 tokens (KOR), and 425 tokens (TWN).

---

2   See Ishikawa (2013) for details.

3   Available at: http://www.laurenceanthony.net.

4. Research Findings

4.1. Distribution of the Discourse Marker and Non-Discourse Markers

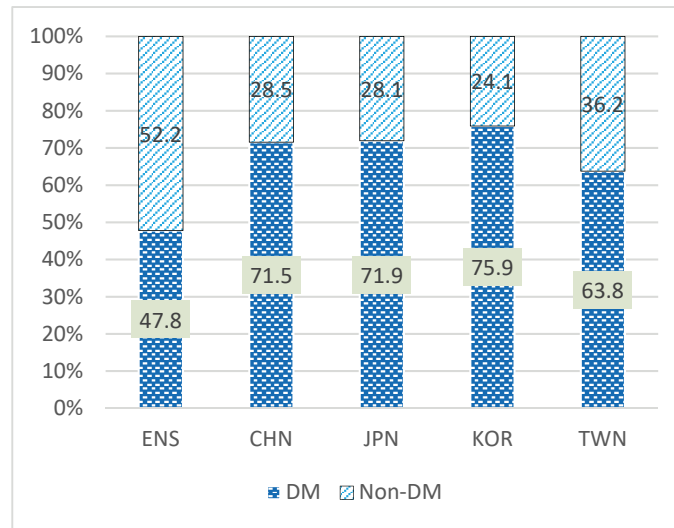A distribution of (non-)DM uses of *so* in each corpus is shown in Figure 1.

Figure 1. Distribution of (non-)DM Marker Uses

As demonstrated, EFL learners are found to use *so* notably more often for DM purposes, rather than for so-called grammatical, non-DM uses. Meanwhile, the opposite pattern is mildly indicated by the ENS group, whose non-DM uses slightly exceed their DM uses. The tendency of EFL learners to favor DM uses is moderately consistent among all groups, with the TWN learners standing closest to the ENS. It needs to be emphasized again, however, that there is a significant overuse of DM *so* by all EFL learners.

Figure 2 shows the frequency of *so* normalized to tokens per 10,000 words, confirming the predisposition of the EFL learners for DM uses.
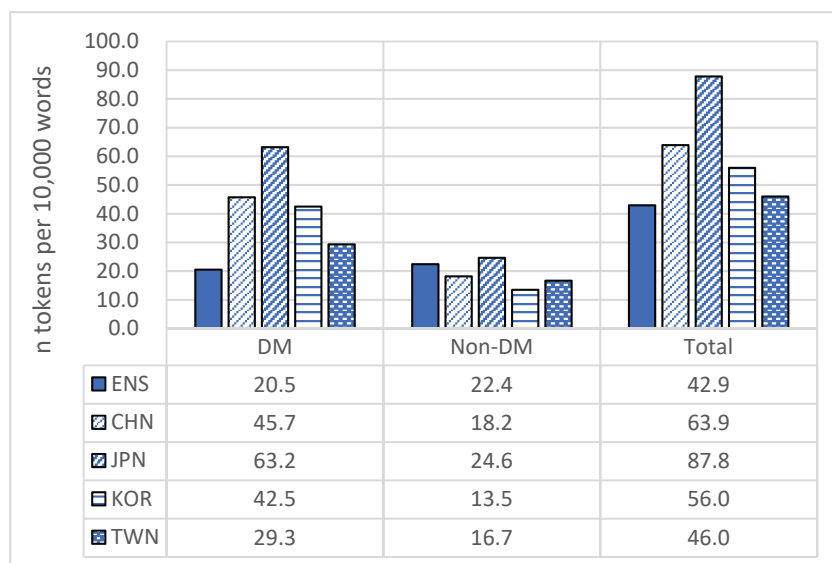

	DM	Non-DM	Total
ENS	20.5	22.4	42.9
CHN	45.7	18.2	63.9
JPN	63.2	24.6	87.8
KOR	42.5	13.5	56.0
TWN	29.3	16.7	46.0

Figure 2. Frequency of *So* (Normalized Per 10,000 Words)

Figure 2 indicates that the JPN learners use *so* as a DM more than three times as often as the ENS, while the CHN and the KOR learners' usage is more than twice as great as the ENS group's. In addition,

the disparities among the L2 groups are significant; the JPN learners, for instance, use the DM *so* twice as often as the TWN learners. Concerning non-DM uses, all L2 groups except the JPN learners use *so* less frequently than the ENS group, with the highest and the lowest being the JPN and KOR learners, respectively. Accordingly, the total frequency of *so* is the highest in the JPN corpus, reaching more than double that of the ENS corpus. Within the L2 groups, the TWN corpus demonstrates the patterns with the most similarities to the ENS corpus, while the JPN corpus shows the least. This is conceivably because the JPN corpus has the smallest percentages in the highest two levels, i.e., the B1_2 and the B2+.

4.2. Non-Discourse Marker Uses of *So*

Substantial differences can also be observed between the corpora in terms of the types of non-DM functions of *so* employed, as shown in Figure 3.
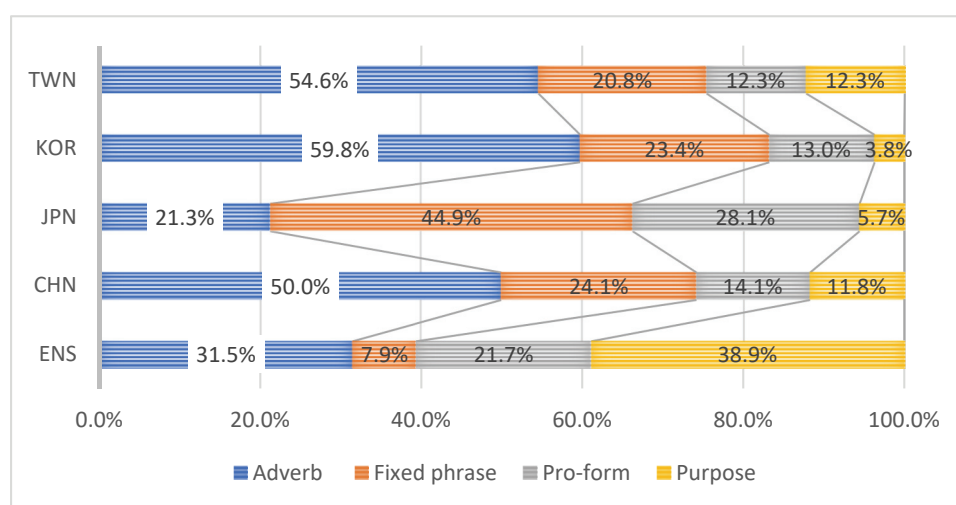

Figure 3. Types of Non-discourse Marker *So*

The notable findings include the following. First, the ENS appear to value the function of *so* to indicate purpose (*n* = 79, 38.9%) in contrast to the L2 learners, whose "purpose" uses are outstandingly modest. Second, L2 learners' use of *so* as part of fixed phrases is evident, with the JPN learners showing the highest proportion (*n* = 198, 44.9%). Third, the use of *so* as an adverb is prominent among the L2 learners, particularly KOR and TWN learners, with JPN learners being an exception. Finally, the use of pro-forms is moderate across the L2 groups except for the JPN learners. As shown, the JPN learners behave quite differently from the rest of the L2 learners in that their uses of *so* as part of fixed phrases and as pro-forms are remarkably high. This result again might be related to JPN learners' lower average proficiency. The features of L1 usage of the non-DM *so* can be summarized as the robust usage of its "purpose" meaning and the suppressed usage of fixed phrases. The results are confirmed by the relative frequency of non-DM *so* usage in all corpora, as shown in Figure 4.
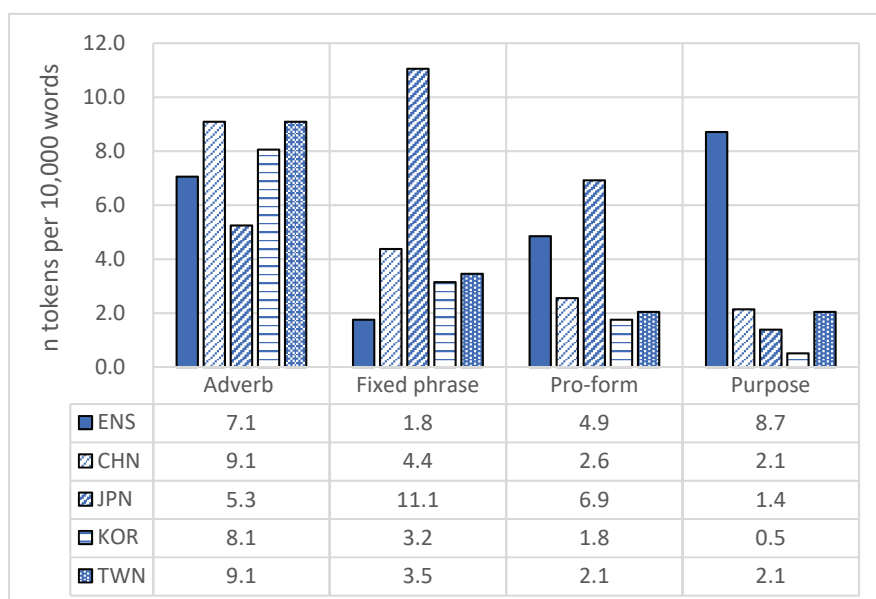
	Adverb	Fixed phrase	Pro-form	Purpose
ENS	7.1	1.8	4.9	8.7
CHN	9.1	4.4	2.6	2.1
JPN	5.3	11.1	6.9	1.4
KOR	8.1	3.2	1.8	0.5
TWN	9.1	3.5	2.1	2.1

Figure 4. Frequency of Non-discourse Marker *So* (Normalized Per 10,000 Words)

It is worth noting that the CHN learners' use of *so* as an adverb is, in fact, comparable to that of the TWN learners, with the KOR learners being the second highest.

Table 5 shows the distribution of fixed phrases. The ENS corpus indicates a generally reserved usage with modest employment of *so far*. The JPN learners' use of *and so on* (n = 186, 93.9%) is significant, especially because their use of the other types of idiomatic expressions is fairly similar to that of ENS and the other L2 learners.

Table 5. Distribution of Fixed Expressions (Raw Tokens)

	ENS	%	CHN	%	JPN	%	KOR	%	TWN	%
and so on	1	6.2	68	81.0	186	93.9	30	69.8	27	84.4
or so	2	12.5	0	0.0	6	3.1	2	4.6	0	0.0
so far	4	25.0	3	3.5	3	1.5	6	14.0	3	9.4
others	9	56.3	13	15.5	3	1.5	5	11.6	2	6.2
Total	16	100.0	84	100.0	198	100.0	43	100.0	32	100.0

The results concur with the literature claiming that JPN learners tend to overuse *(and) so on*, along with other pragmatic markers—e.g., *I think* and sentence-initial *and* (cf. Babanoğlu, 2014). It should be noted, however, that the other L2 groups exhibit similar tendencies, in that the use of *and so on* accounts for 84.4% of expressions for TWN, 81% for CHN, and 69.8% for KOR learners.

5. Discussion

5.1. *So* as a Degree Adverb

As mentioned in the previous section, all L2 learners except for the JPN group were found to employ *so* as an adverb extensively; among those, the KOR and the TWN learners show some unique patterns. In the KOR corpus, for instance, *so* appears with adjectives of various kinds, expressing the writer's personal emotional state and assessment. Consider example (1), where *so* modifies the adjective "angry":

(1) [KOR 170 B1_1]
    Because public place is used by all people. So many people are there. If there is a smoking
    person, there will be non-smoking person, too. Think about it. If you see a person smoking in
    the subway, bus. Smoke smelling makes us **so** angry. Same like this. Public place is a place
    that every people have to get a manner.

In contrast to the L2 usage, the uses of *so* in the ENS corpus, when modifying adjectives, appear
almost exclusively with evaluative adjectives (e.g., *bad, busy, important*); no cases are found where the
ENS use *so* to modify emotive adjectives. See example (2), in which *so* co-occurs with the adjective
"important":

(2) [ENS 197]
    For some students, it could be very important, their time might be spent better investing that
    time in other pursuits. At least for my own experience, it is not **so** important for college
    students to have a part-time job.

As for the KOR corpus, cases where *very*, instead of *so*, would be a more suitable option proved to
be quite common (3):

(3) [KOR 265 B1_2]
    Many part time job need a faithful person and sincere person. So while working at a part
    time job, we can learn the character like that. Such as faith and sincerity are important to
    work and also study. So working at a part time job is **so** helpful. We can get lots of good things
    from working at a part time job.

Such usage of *so* often has a meta-discourse effect, construing part of or the entire essay as
somewhat emotion-laden and thus lacking objectivity. This meta-discourse effect is often detrimental
to the reader's perception and evaluation of the essays. Example (4) illustrates the effects of repeated
uses of *so* of this type in the KOR corpus:

(4) [KOR 114 B2]
    My experience is **so** informative and helpful. I want to work another job so I want to know
    another value. My children suggest that part time is **so** good experience.

The types of usage described above are also common in the TWN corpus. The difference between
the KOR corpus and the TWN corpus, however, is that in the TWN corpus, *so* tends to be used as part
of the idiomatic construction "so X that Y" (20/84 tokens, 23.8%). Consider example (5) found in the
TWN corpus:

(5) [TWN 108 B2]
    Part-time jobs cut into students' study time; some even become **so** preoccupied with making
    money **that** they can't focus their attention on studying. In the end, the students may fall
    behind or fail in their studies.

The same construction occurs only rarely in the KOR corpus (3/110 tokens, 2.7%). Meanwhile, the
construction "so X that Y" is fairly common in the ENS corpus, amounting to 14 cases out of 64 tokens
(21.9%).

5.2. *So* as Part of Fixed Phrases

Given the disproportionate use of *so* as part of fixed phrases, it is suggested that L2 learners closely evaluate the necessity and effectiveness of providing a list of examples using *and so on*. In this regard, it is worth noting that only one instance of *and so on* was detected in the ENS corpus:

(6) [ENS 085]
Japan has a lot of food that is well known for its distinctiveness, foods such as sushi, tempura, sukiyaki, okonomiyaki **and so on**. A lot of these foods have a unique taste and are served with a distinct sauce mainly for that particular dish. After smoking a cigarette and then enjoying a carefully prepared dish, I personally don't think that you would be able to taste the full flavor of the meal.

Here, the names of the dishes, given as examples of Japanese food, constitute information relevant to the writer's claim that smoking would disturb the enjoyment of the full flavor of each particular dish. In contrast, the series of examples followed by *and so on* found in the EFL essays did not necessarily substantiate the writer's argument, particularly in the JPN corpus. Example (7) illustrates this point:

(7) [JPN 144 A2]
And we will think that we use money importantly. Then, we can use much money that we earned without minding. For example, we can watch many movies, travel around the world, go to a high-class restaurant, and buy expensive clothes, shoes, hats, jewels, bags, cosmetics, **and so on**.

The list of items given in (7) is considered extraneous information in light of the proposed claim. In the JPN corpus, a series of examples is given where no detailed information is necessary; *and so on* is used, for example, to end the list of different types of part-time jobs, names of harmful chemicals, or diseases caused by smoking.

To use *and so on* appropriately in various contexts, learners need to adopt a meta-discourse perspective to manage the flow of information. Thorough pre-writing activities in which learners simulate different ways of organizing points of arguments and discuss the end-result of each approach would be effective measures to help achieve this goal. Post-writing activities are equally important in helping learners to discover the process of deleting redundant information. Throughout the overall writing process, ample opportunities to examine model essays closely, in terms of the flow of information, could help learners to analyze their texts objectively. Being familiar with other ways of presenting examples besides using *and so on* is also important so that learners can be equipped with a range of resources, from which they can identify the most contextually suitable.

Where appropriate, teachers should utilize corpus-based data and share information about the relative frequency of fixed expressions, possibly in different genres. The learners, especially those at advanced levels, need to be able to develop a perception about the frequency, which is largely context-dependent.

5.3. *So* to Indicate "Purpose"

Another aspect requiring specific pedagogical attention is the underutilization of *so* for expressing purpose among L2 learners. Although differences were detected between two groups in this study—CHN and TWN as one and JPN and KOR as the other—with the former indicating higher occurrences, the use of "purpose" *so* was found to be extremely infrequent among the L2 learners. The dearth of

this usage may indicate the areas in which the L2 learners experience considerable difficulties.

In the ENS corpus, the tokens of *so* marking "purpose" often appear without *that*, as in (8).

(8) [ENS 189]
   College is for study and I believe that is what students should be spending their time on mostly **so** they have a chance at graduating with solid scores.

The distinction between the DM (result) and the purpose function (non-DM) is often not so straightforward in the ENS corpus as shown in this case, as opposed to the case of L2 writings, where the majority of DM uses of *so* appear in sentence-initial position, while "purpose" *so* almost always takes the form of either *so that* or *so as to*. L2 learners need to be exposed to cases where *so* solely expresses "purpose" such as (8), but first they need to be made aware that a number of the functions described above, irrespective of the DM–non-DM distinction, are in fact interrelated with one another. To help them gain such a holistic view, a brief instruction on the historical development of the word *so*, which I will present below, could be beneficial.

5.4. Toward an Integrated View of *So*

Mitchell and Robinson (2011) show that the clauses of purpose and result using *so, so that,* and *so* X *that* Y are attested in Old English, explaining the interconnectedness of the result (DM) and the purpose (non-DM) functions as follows: "a result is often a fulfilled purpose and a purpose a yet-to-be-completed result" (p. 89). According to Baker (2012), *swã* ('so') was a common adverb in Old English, with a second usage as a subordinating conjunction equivalent to *as*.

Detailed syntactic accounts can be found in Curme (1931), in which he claims that in its history, *so* used to appear in the two-part structure (see (9)), functioning as a determinative to "point[s] as with an index finger to the following explanatory clause" (p. 287):

(9) He has always lived **só**: he cannot expect sympathy now. (p. 286)
   [principal proposition]: [explanatory clause]

The source of the current usage of *that* along with *so* can be located in a structure common in Old English—that of the "double determinative"—where either *súch* or *só* was used in the principal proposition, along with *that* in the clause of the result (Curme, 1931, p.287). Example (10) is claimed by Curme to be the earlier form of the sentence "He has always lived *such* a life *that* he cannot expect sympathy now":

(10) He always lived *súch* a life, *thát*: he cannot expect sympathy now. (Curme, 1931, p. 287)

In fact, *so* and *that*, both being determinatives, had the same function of "pointing to the following clause" and were "competitors in a number of different kinds of subordinate clauses" (pp. 287-288).

Further technical explanations may be of interest to advanced learners; for example, the *so that* clause is derivative of the structure shown in (11), where *so* is stressed as a marker carrying the "idea of manner":

(11) We should **só** áct in this matter that we shall have nothing to regret. (Curme, 1931, p. 288)

So is claimed to have lost its stress gradually over time as the verbal meaning became more valued than the modal idea of manner, resulting in its shift to the subordinate clause. Hence, the subordinate clause indicating a result, as in "He went early *so that* (or simple *so*) he got a good seat" (p. 288).

Additional information that might be of interest to learners include the following:

a. The *so that* clause of a result was once expressed in the form of *so as* in the literary language, as in: "Many came unto them from diverse parts of England, *so as* (now *so that*) they grew a great congregation" (Bradford, *History of Plymouth Plantation*, p. 40, A.D 1630–1648, cited in Curme, 1931, p. 290).
b. The linkage between *as* and *so* can also be observed in the historical development where *as* derived from *all so*; "Do *as* you think best" was originally "Do *all so*: you think best" (p. 278).

The diachronic account provides learners with an integrated view of *so* having the function of expressing "manner" as its operational core; this will help them discern the relational consistency in its many functions. Furthermore, this presents a great advantage for learners, who would otherwise interpret each function as a discrete, independent phenomenon, the acquisition of which largely entails memorizing a great deal of information. The essential feature of *so* to express manner is evident in its pro-form usage as well as marginally present in its fixed usage; the pro-form uses constitute what is called an "anaphoric *so*" (Huddleston & Pullum 2002, p. 1538). In its present uses, *so* seems to have broadened its scope of "pointing" to include a backward direction.

The matter of introducing diachronic information could be controversial in many respects as it would involve exposing learners to constructions in which the use of *so* is ungrammatical in present-day English. To avoid additional difficulties on the part of the learners, target groups should be restricted to highly advanced learners. In addition to the learners' proficiency level, their first languages may play a crucial role in understanding the historical backgrounds.

6. Conclusion and Implications

The present study has examined L1 English and L2-learner English with regard to the grammatical uses of *so* in written discourse. In terms of the overall distribution of DM and non-DM uses, English learners exhibited an overuse of DM functions, whereas the quantitative difference in non-DM uses is much smaller. The types of non-DM *so* used by L2 learners are in fact distinctively different from those of L1 English, of which the notable examples are the learners' underuse and overuse of "purpose" meaning and fixed usages, respectively.

Considering that the data were obtained under controlled conditions in terms of the topics and the length of essays, the results of this study present important implications for EFL education. First, instructions targeting learners' skills to manage discourse flow appear to be of importance. The instructions should explicitly describe the discourse functions of *and so on* to avoid its overuse among L2 learners. In addition, I recommend that learners be exposed to cases where *so* expresses purpose without *that* in order to address its underuse. This could be followed by instructions on more technical aspects of usage, such as the differences between the result and purpose structures (Huddleston & Pullum, 2002, p. 733). Other areas requiring explicit instruction include adverbs, where the L2 learners' overuse was rather evident. Lastly, adopting an integrated view in which the multifunctionality of *so* is understood as involving different manifestations of its core meaning would benefit learners who may otherwise interpret each function as completely separate from the others. In addition, it is essential for teachers and material writers to adopt such a perspective in order for the learner to be able to conceive a relative consistency across the extensive functions of *so.*

Future studies on the topic might seek to conduct in-depth analyses of the (non-)DM uses of *so* in non-academic texts and spoken discourse in various genres. Furthermore, qualitative studies based on corpus data addressing the actual usage of other discourse-related devices by both L1 English speakers and L2 learners are greatly needed. In addition, considering the complexities involved in

defining a "native speaker," I recommend that the subject's performance be investigated based on their proficiency level or other credible factors, thus transcending the binary distinction of native and non-native speakers.

Acknowledgments

References

Babanoğlu, M. P. (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.

Baker, P. S. (2012). *Introduction to Old English*. Chichester: Wiley-Blackwell.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Essex: Pearson Education.

Blakemore, D. (1988). 'So' as a constraint on relevance. In Kempson, R. (Ed.), *Mental Representations: The Interface Between Language and Reality* (pp. 183-195). Cambridge: Cambridge University Press.

Bolden, G. B. (2009). Implementing incipient actions: The discourse marker 'so' in English conversation. *Journal of Pragmatics, 41*, 974-998.

Buysse, L. (2007). Discourse marker 'so' in the English of Flemish university students. *Belgian Journal of English Language and Literatures, 5*, 79-95.

Buysse, L. (2012). 'So' as a multifunctional discourse marker in native and learner speech. *Journal of Pragmatics, 44*, 1764-1782.

Curme, G. O. (1931). *Syntax*. Boston, MA: D. C. Heath & Company.

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. (1996). Pragmatic markers. *Pragmatics, 6*(2), 167-190.

Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics, 28*, 410-439.

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.

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. *Learner Corpus Studies in Asia and the World, 1*, 91-118.

Johnson, A. (2002). *So...?*: Pragmatic implications of *so*-prefaced questions in formal police interviews. In Cotterill, J. (Ed.), *Language in the Legal Process* (pp. 91-110). London: Palgrave Macmillan.

Lam, P. W. Y. (2009). The effect of text type on the use of *so* as a discourse particle. *Discourse Studies, 11*, 353-372.

Lam, P. W. Y. (2010). Toward a functional framework for discourse particles: A comparison of *well* and *so*. *Text & Talk, 30*(6), 657-677.

Larsen-Freeman, D., & Celce-Murcia, M. (2016). *The Grammar Book: Form, Meaning, and Use for English*

Language Teachers. Boston, MA: Cengage Learning.

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.

Mitchell, B., & Robinson, F. C. (2011). *A Guide to Old English*. Chichester: Wiley-Blackwell.

Müller, S. (2005). *Discourse Markers in Native and Non-native English Discourse*. Amsterdam: John Benjamins.

Quirk, R., Greenbaum S., Leech G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

Sato, S. (2019). A corpus-based analysis of *so* in written discourse: A comparison between L1 English speakers and Japanese EFL learners. *Applied Pragmatics, 1*(1), 26-45.

Schiffrin, D. (1987). *Discourse Markers*. Cambridge: Cambridge University Press.

Schourup, L. C. (1999). Discourse markers. *Lingua, 107*, 227-265.

van Dijk, T. A. (1979). Pragmatic connectives. *Journal of Pragmatics, 3*, 447-456.

Vickov, G., & Jakupčević, E. (2017). Discourse markers in non-native EFL teacher talk. *Studies in Second Language Learning and Teaching, 7*(4), 649-671.

THE AUTHOR

Shie Sato is an Associate Professor in the College of Social Sciences at Ritsumeikan University. She received her Ph.D. in Applied Linguistics from the University of California, Los Angeles (UCLA). Her research interests include Pragmatics, Discourse Analysis, and Second Language Acquisition.

THE AUTHOR'S ADDRESS

First and Corresponding Author
Shie Sato
Professor
Ritsumeikan University
56-1 Toji-in Kitamachi Kita-ku, Kyoto-shi, Kyoto JAPAN
E-mail: shie@fc.ritsumei.ac.jp