

A Corpus-based Study of the Use of Conjunctions

Mariko Abe
Sophia University

PAC3
at
JALT
2001

Conference
Proceedings



International
Conference
Centre

Kitakyushu
JAPAN

November
22-25, 2001

By comparing corpora of written essays by Japanese EFL learners at different proficiency levels with a native speaker corpus, I investigate the overuse and misuse of a sentence-initial conjunction. The results suggest that Japanese learners misunderstand the usage of a conjunction, and misuse it as if it were a variant of interjections found in spoken conversation of native speakers. We can also presume that the spoken language of native speakers affects the nature of a written language of non-native speakers.

レベルの異なる日本人英語学習者のエッセイのコーパスをネイティブのそれと比較することで、文頭で用いられる接続詞の多用・誤用について考察してみたい。実験の結果から、日本人学習者は接続詞の用法を正しく理解してはならず、ネイティブが会話で用いる間投詞のようなものとして用いていることが分かった。またネイティブの話し言葉が、日本人学習者の書き言葉に影響を与えているとも考えられる。

In this study, I point out the features of English written by Japanese learners by comparing three different corpora: a corpus of native speakers (hereafter NS) and corpora of non-native speakers (hereafter NNS) at two different proficiency levels.

Overuse of connectors

Japanese EFL learners seem to have difficulty in how to begin a sentence in English. This may be partly due to the intensive practice of sentence-to-sentence translation and the lack of extensive writing exercises. Firstly, I examine the use of conjunctions at the beginning of the sentence by using my data of Japanese senior high school level learners. A sentence starts from ‘I’ is frequently used, but learners have a tendency to use conjunctions as follows:

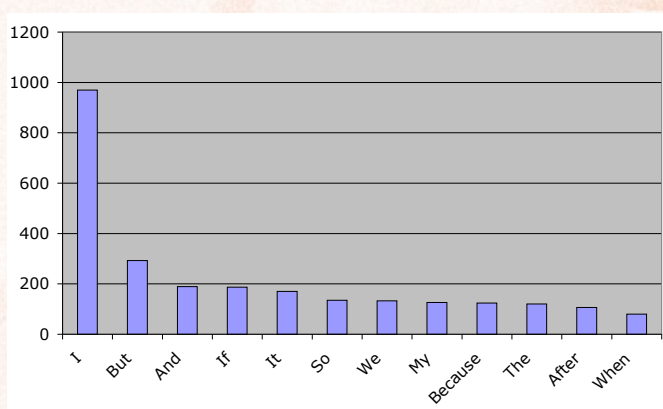


Chart 1: The frequencies of the sentence-initial words in my data

As can be seen from Chart 1 the use of conjunctions at the beginning of the sentence is conspicuous in learners’ English. This usage of conjunction is affected

by Japanese usage of conjunctions, which usually connect sentences (Asao, 1989, p. 242). For example, the following Japanese sentences (a translation of learners’ English in Table 4 (1) – (4)) are natural, and underlined conjunctions can be all translated into ‘And.’

- (1) パスポートが必要だ…。また、聖書を持っていくと良いかも知れない。
- (2) 一ヶ月に一度両親にもらえるお金はとても少ない。だから、CDをたくさん買うことができなかった。
- (3) 私の友達もそれを買った。それから、私達は原宿から渋谷まで走りに行った。
- (4) クラブのコーナーは面白くないが、多くのスペースを取っている。しかし、フリーコーナーはあまりにも小さい。

In this study I examine the frequently used sentence-initial conjunctions, as they can be one of the more important elements to better explore instructions for writing English with greater fluency. This study especially focuses on the conjunction ‘and,’ since it is the second most frequently used conjunction in my data.

Usage of a sentence-initial ‘and’ (SIA) in English reference books

I firstly examine how ‘and’ sentences are defined and classified in English reference books so that I can compare with that of learners, which I will analyze in later section. The first entirely corpus-based grammar

book, *Longman Grammar of Spoken and Written English* (Biber et al., 1999), describes sentence-initial coordinators as below:

There is a well-known prescriptive reaction against beginning an orthographic sentence with a coordinator. Nevertheless, in actual texts we quite frequently find coordinators in this position. (p.83)

In addition to this, sentence-initial ‘and’s and ‘but’s (SIABs) are not regarded as a problem with syntax but only as a semantics problem, so it is difficult to find the prescription of “SIAB avoidance convention” in the dictionaries (Yoneoka, 1998, p. 16). Consequently, using ‘and’ to begin sentences seems to be a controversial usage, but we can say that it is certainly in real language use. Some SIA explanations can be found in the following dictionaries: (a) *Taishukan’s Genius English-Japanese Dictionary* (1994), (b) *Longman Language Activator* (1993), and (c) *Collins Cobuild English Dictionary* (1995).

(a)	In order to respond to the word of respondent, or to stimulate the conversation, it is used at the beginning of questions (translation mine).
(b)	Ways of adding something new to what you have just said
(c)	<ul style="list-style-type: none"> - You use <i>and</i> to interrupt yourself in order to make a comment on what you are saying. - You use <i>and</i> at the beginning of a sentence to introduce something else that you want to add to what you have just said. Some people think that starting a sentence with <i>and</i> is ungrammatical, but it is now quite common on both spoken and written English.

Table 1: SIA in reference books

In these explanations ‘and’ is used as a filler of the conversation rather than a conjunction, and ‘additive’ meaning is included in it. This review brings us to the conclusion that SIA is often used in spoken language and why ‘and’ is often seen in the writings of Japanese learners, especially at the beginning of the sentence. The following section examines this empirically by using corpora that is based on three different Englishes.

Corpus Selection

I have compiled my learner’s corpus that consists of second-grade senior high school students’ English. The framework is based upon one developed by Yukio Tono at Meikai University. The topics are composed of three subjects: (a) Tell us about your school festival. (b) How

do you spend your *otoshidama*? (c) What would you take with you if there were a big earthquake? To compare with much more advanced learners I have chosen *Longman Learner's Corpus* (LLC), and to identify the natural usage of language I have selected the informative and imaginative written text data of *British National Corpus* (BNC):

Frequencies of SIAs in each corpus

The researcher transcribed all of the hand-written manuscripts from students. Spelling mistakes were not normalized at this stage. I used *Wordsmith Tools*, a program suite developed by Mike Scott (1998) to calculate the frequencies.

	MY DATA	LLC	BNC
corpus size (tokens)	36,710	658,250	89,740,543
freq. of the overall 'and'	963	15,447	2,425,111
per mil. Words	26,233	23,467	27,024
freq. of SIA	154	1,541	70,790
per mil. Words	4,195	2,341	789
SIA in the overall 'and' (%)	16%	10%	2.9%

Table 2: Frequency of SIA usage & whole 'and' usage

When we compare the frequency of the overall 'and' usage in per million words, it is almost identical, whereas the frequency of SIA usage per million words is totally different. The percentage of SIA in the entire

'and' clearly decreases as the level of English proficiency increases. We can conclude from these results that more advanced writers of English use the sentence-initial 'and' less in their writing.

Apart from my data, the following research provides us with further evidence. Asao's learner corpus based on writing assignments by Japanese University students indicates a high percentage of SIA usage, 16%, while his personally compiled NS corpus, Ann Landers' letter correspondence, has only 2.9% (Asao, 1998, p. 242).

	NNS			NS	
	MY DATA	ASAO'S	LLC	ANN'S	BNC
Freq. of SIA	154	43	1,541	24	70,790
Freq. of the overall 'and'	963	265	15,447	839	2,425,111
%	16%	16%	10%	2.9%	2.9%

Table 3: Percentage of SIA in overall use of 'and'

Here, again the results show that using a coordinative conjunction at the beginning of the sentence might be unnatural usage for NS. In the following section, I will examine the usage of 'and' by classifying the concordances according to their semantic properties.

The semantic aspect of SIA

I reduced the number of concordance lines of each corpus using Wordsmith's automatic thinning function,

to the number of my learner corpus, 154. (This is similar to extracting a random sample). To analyze functional roles of them, I classified the use into seven categories by referring to the *Taishukan's Genius English-Japanese Dictionary* (1994):

	GENIUS	LEARNERS'
(1) Listing, Additive [and also...]	You mind your own business. <u>And</u> when you talk to me, take off your hat.	I need the passport because of showing who I am, as I were left alone. <u>And</u> it may be good to take the Bible.
(2) Resultive [and so, and therefore]	He was very tired <u>and</u> (so) went to bed early.	...but my money which is given by my parents once a month is very cheap. <u>And so</u> , I couldn't buy a lot of CDs.
(3) Transitional [and then]	She washed the dishes <u>and</u> she dried them.	My friend bought it too. <u>And then</u> , we went running from Harajuku to Shibuya.
(4) Contrastive [but]	He's afraid of the sea. <u>And</u> he calls himself a sailor.	Club's corner isn't interesting, but they use large place. <u>And</u> free corner is too small.
(5) Conditional [if]	-	First it is essentail. <u>And</u> especially in winter, I'll take out some clothes or some <i>moufu</i> .

(6) no need to use SIA	-	But there was many people who came in our room. <u>And</u> most of them looked interested in our <i>happyo</i> .
(7) cannot be categorized	-	-

Table 4: The framework for data analysis of the semantic aspects of SIA

The following Table 5 shows that there are three significant differences between NS and NNS. First, the frequency of SIA is surprisingly high in category 'listing' in NS corpus, 82%, whereas both NNS corpus have only 44 and 47%. Secondly, learners use SIA in 'resultive' meanings much more frequently than NS. My learner corpus has 30, but BNC has only 4. Another significant difference is that Japanese learners tend to use 'and' at the beginning of a sentence where there seems to be no need. Both my data and LLC have considerably higher percentages 27 and 25%, compared with that of BNC 1%.

Functions	MY DATA		LLC		BNC	
	Freq.	%	Freq.	%	Freq.	%
(1) Listing	68	44 %	72	47 %	127	82 %
(2) Resultive	30	20 %	17	11 %	4	3 %
(3) Transitional	8	5 %	18	12 %	15	10 %
(4) Contrastive	5	3 %	7	5 %	5	3 %
(5) Conditional	1	0.6%	0	---	0	---
(6) no need to use SIA	41	27 %	38	25 %	2	1 %
(7) not be categorized	1	0.6%	2	1 %	0	---

Table 5: Distribution of semantic types of SIA

Other semantic categories do not seem to have many differences, but we can find two interesting points when we compare Table 5 with Chart 1. As can be seen from Chart 1, the conjunction ‘but,’ which has the meaning of ‘contrastive,’ is frequently used at the beginning of the sentence, so learners do not substitute ‘and’ for the meaning of ‘contrastive.’ When we see the frequency of sentence-initial ‘so’ in Chart 1, which has the meaning of ‘resultive,’ it is fairly high, but learners overuse the sentence-initial ‘and’ with a meaning of ‘resultive.’ Regarding this result, we can assume that learners use both sentence-initial ‘and’ and ‘so’ as a meaning of ‘and so’ or ‘and therefore.’

From these findings it can be said that one of the most significant problems of Japanese learners is the misunderstanding of usage. There is confusion among Japanese learners who consider ‘and’ as a connector which joins two sentences together. In order to correct this misunderstanding, a clause-level connector should be taught with various semantic types.

Characteristics of NS use of SIA

As I have observed in Section 3. SIA is preferably used in a spoken text. I have found out an interesting fact that 50 instances out of 154 samples (32 %) were written in the form of direct speech. Therefore, I compared the frequency of the SIA in written text with that of spoken text in BNC.

	WRITTEN	SPOKEN
corpus size (tokens)	89,740,543	10,365,464
freq. of overall ‘and’ usage	2,425,111	261,081
per mil. words	27,024	25,188
freq. of SIA usage	70,790	39,067
per mil. words	789	3,769
SIA in overall ‘and’ usage (%)	2.9 %	15 %

Table 6: Comparison of SIA between written and spoken text of BNC

As seen from Table 6, the SIA is more frequently used in spoken text than in written text. Yoneoka (1998, p.

15) says that “sentence-initial *and* and *but* are genre-dependent, occurring most often in spoken English, but less in written literature/journalism and least frequently in academic writing.” Apart from this, the percentage of SIA in overall usage of ‘and’ in spoken text (15%) is very close to that of Asao and my learner corpus percentages (16%), provided in Table 3. If we understand SIA as a filler of conversation in writings of NNS, we can explain the high frequency of unnecessary SIA usage of learners as a filler of written communication, too.

In the same way, close investigation of this BNC spoken text shows that children have a higher tendency to begin conversations with the conjunction ‘and.’ Conversations between children (aged 0-14) have a frequency of 4342 per million words, while adults (aged 25-44) have only 2429 per million words. This result leads us to the conclusion that younger speakers use the SIA more frequently, and that more advanced speakers of English use it less in their oral communication.

Conclusion

Data drawn from the three corpora supports the claim that more advanced writers of English use less SIA in their writing and speaking. Regarding the overuse

by learners, we can assume that they misunderstand its usage, and use it as the kind of interjection often found in NS spoken conversation. In addition, when we compared a NS spoken corpus with NNS written corpora, the ratio of SIA against the total usage of ‘and’ was almost identical. We can conclude that the usage of SIA in NS spoken language may affect the nature of NNS writing.

The pedagogical implication drawn from this study is that greater exposure to a wide range of English connectors is indispensable for Japanese learners. Not only basic conjunctions, but also adverbial connectors, which have the resultive function, such as ‘as a result’ or ‘consequently,’ should be included in the learning list of connectors. Also, if we are able to use computers in the classroom, learners can withdraw the example sentences of conjunctions by themselves to learn the usage. The study shows that by comparing NNS corpora and NS corpus, this type of data-driven learning has great potential as more corpus resources become available. In future studies, I will continue to investigate other conjunctions that are frequently used at the beginning of sentences.

References

- Asao, K. (1998). English Education and Corpus Use. In T. Saito (Eds.), *English Corpus Linguistics* (pp. 233-247). Tokyo: Kenkyu-sha.
- Biber, D. (Ed.). (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Collins Cobuild English Dictionary*. (1995). London: Harper Collins Publisher.
- Konishi, T. (Ed.). (1994). *Taisukan's Genius English-Japanese Dictionary*. Tokyo: Taishukan Publisher.
- Longman Language Activator*. (1993). London: Longman.
- Scott, M. (1998). *Wordsmith Tools*. Oxford: Oxford University Press.
- Yoneoka, J. (1998). "But they do it": Using Corpora to Research Sentence-Initial *And*s and *But*s (SLABs) in Academic Writing. *English Corpus Studies*, 5, 15-26.