The distribution of anaphoric ‘so’ in MICASE
Linguistics 429

Introduction

The Michigan Corpus of Academic Speech in English (MICASE) is a compilation of speech spoke by both native and non-native speakers on the University of Michigan campus. The MICASE project was created with the intention of accumulating a corpus of the natural speech produced in an academic setting. It considers a number of different factors that allow for a diverse body of texts. These factors include gender, academic status, language status (native vs. non-native speakers), academic division, discourse mode and speech event type. These categories play an integral role when analyzing this data. English as a Second Language (ESL) instructors can use this information to identify common speech patterns that occur in academic speech in an attempt to better understand how and why common grammar structures exist and are used.

We used MICASE to assist us in our study of the use of the word *so*, our main focus being on anaphoric *so*; as in the sentence "I don't think *so*." Some researchers such as Susan Hunston and John Sinclair believe that anaphoric *so* must be a part of a local grammar. According to linguist John Swales "the heavy restrictions on anaphoric "so" suggest that it may constitute "a local grammar", i.e. a distinctive sub-grammar that is best treated separately from the language as a whole." This leads us to ask the following questions: Where does it come from; what are its uses; when is it used, and why is it unique to English? In this paper we will try to answer these questions based on the results found in the MICASE database.

Defining *So*

There are many ways a speaker can use the word *so*. We will look at the following uses:

- Deictic *so*

- Discourse marker *so*
- Anaphoric *so*
- Other uses of *so*

Deictic 'so'

Deictic *so* is used to indicate degree, which usually modifies an adjective or manner. Some examples found in MICASE include:

106. "…a sore thumb, but uh, *so* obvious in…"

108. "right cuz there's *so* many sublets"

311. "they've done *so* well"

370. "an A-minus, and I was *so* proud of myself"

When deictic *so* is to be used it is often accompanied by an indexing act indicating the size (or degree) to that of which the speaker is referring. Huddleston examples:

56. i. "She was about *so* tall."

ii. "You turn the fold the paper in two, *so*."

Discourse Marker 'so'

One of the most common forms of the word *so* in modern-day English is as a discourse marker. We've found that as stated in *Discourse Analysis*, "Traditional descriptions of their use do not give the full picture of the deployment of words like *so* and *because* in discourse"(204). The most common use *so* as a discourse marker results in a feeling of casualness amongst speakers. The word *so* is most often found at the beginning of a

sentence when being used as a discourse marker. Of the many examples we have found of discourse marker *so*, here are a few:

4. "So you can see that this…"

59. "…so it's basically like…"

234. "So most of what we know…"

Another interesting use of discourse marker *so* can be found in *Discourse Analysis*. **"**One conventional way of signaling the beginning of narrative chunks is with ***so"***(206). The most famous use of this *so* for our generation can be seen in this movie quote:

"*So* this one time, at band camp…"

Anaphoric 'so'

The focus of our paper is on the use of anaphoric *so*. Some of the many examples of anaphoric *so* include:

Pro-clause complement with finite antecedent: I think so.

Examples from Huddleston:

58. i. A: Are they putting the price up? B: I think *so*. /I'm afraid *so*. /It seems *so*.

ii. She thought he was wrong but was too polite to say *so*.

iii. Will she accept the recommendations, and if *so*, how will they affect us?

Some examples found in the MICASE data include:

106. "Yep, I think so."

134. "I don't think so."

Polarity

Using the word *so* often expresses a positive sentiment while the word *not* is used in the negative pro-form. Examples from Huddleston:

60. i. "She didn't approve of the idea and told them *so*/*not.

ii. A:"Will they be accepting the proposal?"

B: "She says *so*/not.

Pro-predicative

In the pro-predicative form, *so* is used as a predicative complement. The antecedent is usually and AdjP but other categories are possible. (ii. is an examples with a NP antecedent.)

63. i. "They were very happy at that time or at least the seemed *so*."

ii. "The bible was already a symbol of class struggle, and remained so for a long time.

MICASE Example:

434. " why might be *so*."

Exclamatory Confirmation

In certain cases, *so* can be used to express surprise or emotion as a reaction to a previous statement. It is important to note, that in the absence of an auxiliary, this *so* would have to include a supportive *do*.

A: "Jill has misspelt our name." B: "*So* she has!"

(without an auxiliary, 'B' would be read, "*So* she did!"

According to Huddleston and Pellum, the anaphoric *so* cannot be placed into a specific word category. The pro-predicative form in 63. can be an AdjP, NP, or PP. A pro-clausal *so* can be considered a pronoun or a 'noun clause'. *So* does differ from other NP's in that it cannot functions as a subject.

Other Examples

As previously stated, there are many different ways in which *so* can be used. Many of these uses cannot be placed into the anaphoric, discourse marker or deictic categories. Because of this we have created this 'other examples' section where we will explain some of the other types of *so.*

Initial *so*

The use of *as* and *so* indicate a relationship between the main and subordinate clauses. Huddlston examples:

65. i. "Just as Renaissance scholars had to reconcile Platonism with Christianity, so the Victorian Platonist dons had their particular reconciliation to do too."

ii. "As infections increased in women, so did infections in their babies."

Connective adjunct of reason, consequence

When *so* is used as a connective adjunct it marks reason or consequence. Our resource also says that as with many other connective adjuncts, there is an anaphoric component in the meaning "for this reason, as a result of this."

71. i. "There had been a power failure, *so* all classes had had to be cancelled.

ii. "I've no more to say, *so* I suggest we move on.

So in idiomatic coordinates

Our final use of *so* deals with idiomatic expressions. It is often combined to form expressions referring to measurement.

"The test will be in about a week or *so*."

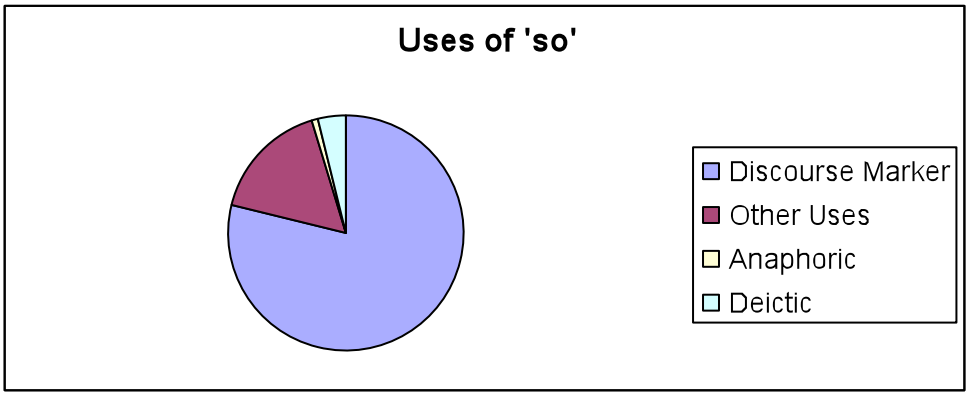
"There were about 60 people or *so* at the party last night."

This use of *so* can be glossed as "or something like that". Another common idiomatic expression is "*so* on and *so* forth."

Research

We initiated our research by determining the total number of occurrences of *so* in MICASE. The database produced 16000 results, from which we took a sample of 500 entries. Of these 500 entries, 4% (20 entries) were Deictic, 79% (395 entries) were discourse markers, .6% (3 entries) were Anaphoric *so*, and 16.4% (82 entries) were other uses of *so*. From here we further investigated the use of anaphoric *so*; the verbs to which it can be applied and the frequency that these verbs occur in.

Figure 1.1



We obtained a list of the verbs that take anaphoric *so* from *The Cambridge Grammar of the English Language* text by Huddleston and Pullum. The following is a list of the verbs we looked at: imagine, know, believe, seem, appear, afraid, suspect, suggest, think, say, hope, suppose, guess, assume, presume, fear, reckon, tell, gather, and trust. These verbs are all what one may call "mental state" verbs and some are used more frequently than others. We wondered if all mental state verbs could take an anaphoric *so* but found that many couldn't. For example:

While it perfectly acceptable to say:	A: Do you like this class?
	B: I suppose so.
it is not acceptable to say	B: *I like so.
or,	
	A: Do you want ice cream?
	B: I guess so.
	B: *I want so.

The above examples prove that anaphoric *so* occurs in an extremely restricted environment and can only be used with some "mental state" verbs. We looked up the following verbs from the Huddleston text to find their frequency in MICASE.

Figure 1.2

	Present	*Past*	*Conditional*	**Tense**		*Present*	*Past*	*Conditional*
Verb					**Verb**			
Imagine	0	0	0		Think	155	7	2
Know	0	0	0		Say	5	3	2
Believe	4	0	0		Hope	3	0	0
Seem	2	0	0		Suppose	0	0	0
Appear	0	0	0		Guess	10	0	0
Afraid	0	0	0		Assume	0	0	0
Suspect	0	0	0		Presume	0	0	0
Suggest	0	0	0		Fear	0	0	0
Gather	0	0	0		Reckon	0	0	0
Trust	0	0	0		Tell	0	0	0

Our results show that the verb "think" is the most commonly used "mental state" verb, followed by "guess," "say," "believe," "hope," and "seem." Surprisingly, other common verbs that take anaphoric *so*, such as "imagine," "suspect," and "suppose" did not produce examples in MICASE. We attribute this to the fact that their uses are less common on campus, whereas they might be more common in other contexts. While there were a number of verbs used with *so* in present tense, only "think" and "say" were used in the past and conditional tenses. Some specific MICASE examples that demonstrated the use of "mental state" verbs are:

93. "That's a woman's hat isn't it?" "I think so."

3. "Are you done?" "I dunno, I guess so."

3. "Find the image of the point." "If you say so, Mark."

4. “You can like check out C-Ds like you can check out books?” “I believe so.”

2. “I’ll see you later.” “I hope so.”

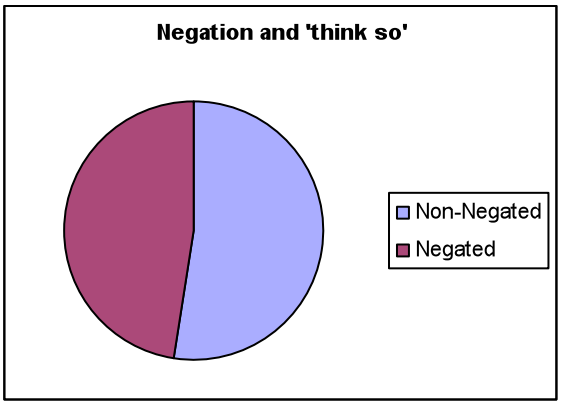
2. “Philosophy can be hard though it just seems so.”

While not included on the chart, we also researched other verb tenses but they produced no further examples of anaphoric *so*. Because of its high occurrence in the MICASE database, we chose to take a closer look at the way in which the verb “think” is used.

Think

We found that there were 152 instances of anaphoric ‘think so’ in MICASE. We thought it would be interesting to compare the number of instances of negated and non-negated ‘think so’ to see if this grammar structure is more restricted to a positive or negative use. We hypothesized that there would a higher number of non-negated ‘think so,’ in the MICASE database, but surprisingly found that there was not a significant difference between the two (80 non-negated and 64 negated).

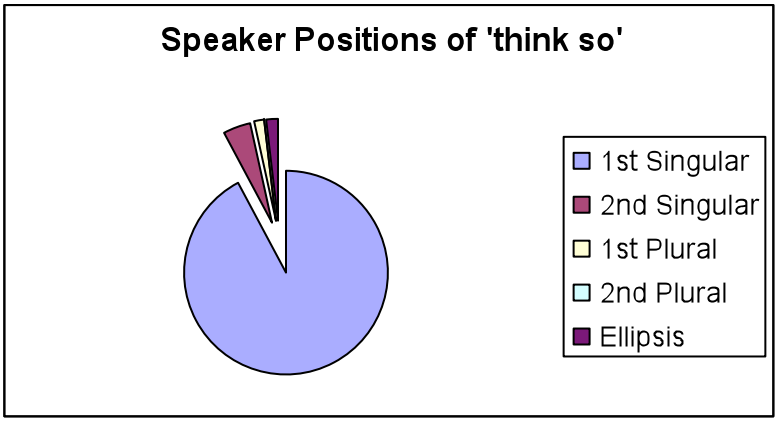
Figure 1.3



After looking at negation we decided it would be equally important to look at speaker position to see whether or not anaphoric *so* was more likely to occur with first or second

person, singular or plural. Out of 152 examples, we found that the 1st person singular was the most predominantly used form (92.1%), while 2nd person singular was second with only 4.6%. The rest include 1st person plural at 1.3% and ellipsis use at 2.%. There were no examples of the 2nd person plural.

Figure 1.4



1st Person Singular

151. "I think so, but I'm not sure."

2nd Person Singular

141. "…a coupon, do you think so?

1st Person Plural

16. "…as we might originally think so."

Ellipsis

125. A: "Do you know of anyone taking it next semester?"
B: "…probably a bunch of people."
A: "think so…?"

After looking at each of these characteristics separately (negation and speaker position), we decided to combine them to see the frequency of usage. In figure 1.3 one can see that negated and non-negated usage of 'think so' occur almost equally. In figure 1.5 we see a

similar pattern in which the negated and non-negated forms of 1st person singular are almost equal in number. While most entries of 1st person singular and plural never occur in form of a question, the majority of 2nd person singular occur in this form. Looking more closely at the chart it is evident that 2nd person singular in the non-negated form occurs twice as much as in the negated form. We interpreted this to occur because the use of negated 'think so,' as in "you don't think so?" can be used as a polite way to respond to disagreement. There are no occurrences of 2nd person plural in either chart and in that the numbers of ellipsis are equal.

Figure 1.5

Non-Negated 'think so'	**#**	**Negated 'think so'**	**#**
1st Person Singular	69	1st Singular	66
2nd Person Singular	6	2nd Singular	3
1st Person Plural	2	1st Plural	0
2nd Person Plural	0	2nd Plural	0
Ellipsis	3	Ellipsis	3
Total occurrences	80	**Total Occurrences**	72

Speech Events

After looking at the grammatical instances in which anaphoric *so* occurs we thought it necessary to examine whether it tends to occur in certain contexts more often than others and whether or not native speaker status influences its usage. There are many different divisions of speech events in MICASE where usage of anaphoric 'think so' can be found.

These divisions are:

AVD- Acad. Advising
COL- Colloquia
DEF- Dissertation Def.
DIS- Discussion Section
INT- Interview
LAB- Lab
LEL- Large Lecture
LES- Small Lecture
MTG- Meeting
OFC- Office Hours
SEM- Seminar
SGR- Study Group

STP- Student Presentation
SVC- Service Encounter
TOU- Tour
TUT- Tutorials

We thought that since anaphoric *so* is often associated with casual speech and perhaps discussions that would spark uncertainty, that it would most likely occur in events where there is more group dialogue and therefore more relaxed speech. This is in contrast to events where authoritative figures are present and when formal speech is expected. For this reason, we think that 'think so' would be most likely associated with speech events such as discussion sections, labs, small lectures, and study groups. The following graph shows the break-down of the occurrences of anaphoric *so* as they occur in the various speech events.

Figure 1.6

Speech Event	#
SGR	38
LES	23
LAB	15
SEM	12
ADV	10
DIS	10
MTG	8
OFC	8
STP	7
SVC	7
INT	4
DEF	3
LEL	2
TOU	2
TUT	2
COL	1

In looking at the chart one can see that anaphoric *so* does indeed occur most frequently in categories where there is more casual speech amongst peers in highly interactive settings.

With thirty-eight occurrences, the use of 'think so' in the study groups (SGR) by far exceeds all other speech events. Small lectures (LES) places second with 23 occurrences. The laboratory (LAB) and seminar (SEM) settings have similar numbers (15 and 12 respectively). Going down the chart, the event types become more formal and centralized, and because of this the usage of anaphoric *so* greatly diminishes.

Speakers

As a result of MICASE being a corpus of the speech at the University of Michigan there are many factors that can attribute to the usage of anaphoric *so* such as age, speaker status and language background.

Age and Status:

Looking at the charts below one can easily see that age and speaker status are closely related. There is a trend between these two charts because people in senior positions of the university tend to be older in age. Therefore, it is presumed that formal language is spoken more often by these speakers in these contexts.

Figure 1.7

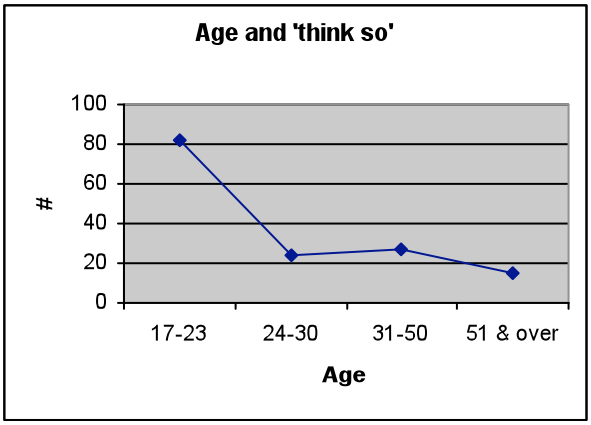
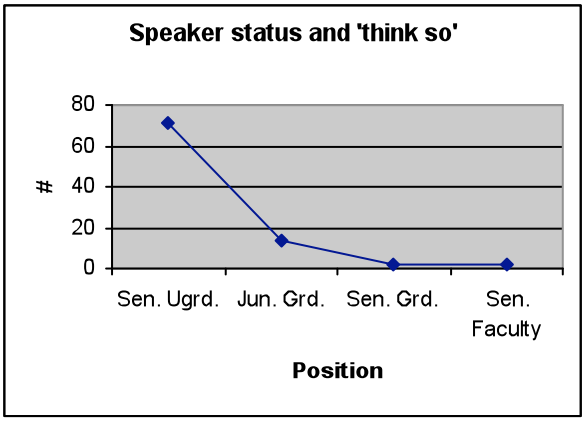


Figure 1.8



Language Background

Since *so* has such a restricted use, we wondered how prevalent it would be in the academic speech of non-native English speakers on campus in the previously mentioned speech events.

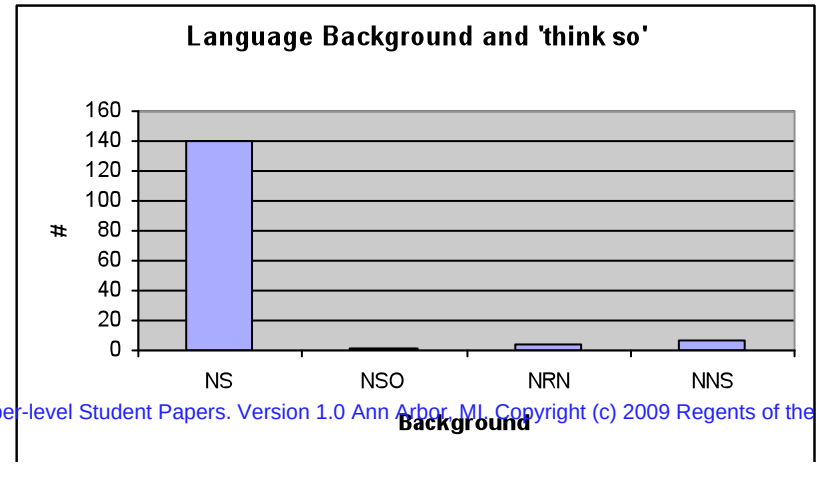
To gain insight into how anaphoric *so* translates into other languages we consulted native and non-native speakers of Spanish, French, German and Hindi. When asked how to translate "I think so" into Spanish, our consultant replied with "Yo Pienso." ("*I think").* This shows that initially he did not know how to incorporate anaphoric *so* into his L1. We later asked how he would translate "Creo que sí" into English, he replied with "I think so." Our consultant for Hindi is from New Dehli, India but has learned Hindi and English concurrently throughout his whole life. Upon asking how he would say "I think so" he struggled to find even a semi-accurate equivalent in Hindi. While he does use anaphoric *so* in his everyday speech in English, he has never had to apply it in Hindi. These observations combined with other glosses we collected show that while there are structures that express the same idea, there is no anaphoric *so* in other languages. Here a just a few examples of the structure used for anaphoric so. Glosses are provided in the parentheses.

English:	*I think so.*	*I don't think so.*
French:	*Je pense que oui.* ("I think that yes")	*Je pense que non.* ("I think that no.")
Spanish:	*Creo que sí*. ("I believe that yes.")	*Creo que no.* ("I believe that no.")

German: *Ich denke ja.* ("I think yes.") *Ich denke nicht.* (I think not.")

Hindi: *Meri soch se*. ("From my thought") *Yeh meri soch nahin hai.* ("This is not what I think.")

From here we decided to look at the language background of the speakers who use anaphoric *so*. We assumed that MICASE would support our hypothesis that non-native speakers would be least likely to use it in their academic speech. Using MICASE would provide us with the most accurate answer for this question. As Ann Mauranen stated, "Because MICASE aims at depicting how English is used on a campus, instead of providing a model of correct or ideal usage, it includes, among other things, a good deal of English spoken by people who would not fit into the category of speakers known as 'native speakers.' In other words, the data has not been subjected to any linguo-ethnic cleansing, but rather it seeks to reflect the reality of English spoken in the characteristic activities of a major American research university" (166). The speaker status in MICASE is divided into four categories: Native Speakers (NS) which are speakers of North American English; Native Speaker Other (NSO) which are native speakers of non-American English; Near Native Speakers (NRN) which are non-native speakers who consider English as their current dominant language and who appear to have native-like fluency and grammatical proficiency; and finally, Non-Native Speakers (NNS) which are non-native speakers of English other than near-native speakers **(MICASEweb).**



Out of the 152 instances of 'think so' in the database, we found 140 entries from Native Speakers, 7 entries from Non-Native Speakers, 4 entries from Near Native Speakers, and 1 entry from Native Speakers Other. This overwhelmingly confirms the hypothesis that of the speakers cited in the MICASE database, anaphoric *so* is restricted to Native Speakers of American English. Just as with our small sample of consultants, this data shows the absence of anaphoric *so* on a larger scale, in international English and in other languages.

Conclusion

Throughout our research we have more clearly defined the grammatical uses of *so* and more specifically the use of anaphoric *so*. We found that there are specific 'mental state' verbs that can take this form of *so* and some like 'think' and 'say' are more commonly use than others. The MICASE data shows that it was only these two verbs that were able to take tenses other than the present. In our sample of 500 occurrences of *so*, only two of these examples were anaphoric, 95% or 152 of 160 entries of 'think so' were anaphoric. Because the verb 'think' produced the highest occurrence of anaphoric *so*, we chose to look at it more closely. When considering negation and speaker position of this verb, it was evident that negation did not play a significant role in the usage of anaphoric *so*; however it was most commonly found that speakers used anaphoric *so* in the first person. Over all, the situations in which this structure occurs are those where there is a high level of interaction between speakers in a more informal environment. There appears to be a

strong correlation between age and status in the university in regards to frequency of usage of anaphoric *so*. This is predictable because professional figures tend to be older in age and find themselves in more situations where they occupy formal roles. By far, our most significant find was the difference in usage or anaphoric *so* between native and non-native speakers of American English. The data shows that 92.1% of the anaphoric 'think so' was produced by Native Speakers of American English, while Non-Native Speakers produced this *so* only 4.6% of the time. In conclusion, the MICASE data that we collected further proves that anaphoric *so* should be considered a local grammar because of its heavily restricted use both grammatically and contextually. We feel that in the future, anaphoric *so* will most likely become more common as the spread of American English takes place throughout the world. If anaphoric *so* becomes more frequently used, it will be necessary to develop second-language teaching/learning materials that deal with this localized grammar.

Bibliography

Biber , D. (1999). *Grammar of spoken and written english*. 1st ed. Essex, UK: Pearson ESL.

Huddleston, R. (2002). *The cambridge grammar of the english langauge*. 1st ed. Cambridge, UK: Cambridge University Press.

Johnstone, B. (2002). *Discourse analysis*. 1st ed. Oxford, UK: Blackwell Publishers.

Mauranen, A. 2001. *Reflexive academic talk: Observations from MICASE.* In R. Simpson & J.M. Swales (eds.) Corpus linguistics in North America.(pp. 165-178). Ann Arbor: U-M press.