The *So*-Inversion Construction Revisited*

YUJI HATAKEYAMA, KENSUKE HONDA, KOSUKE TANAKA**

Abstract

This paper deals with the so-inversion construction in English, as in John can speak French, and so can Ann*, where the second conjunct is the construction in question. Arguing against Toda's (2007) analysis, which involves ad hoc assumptions and theoretical and empirical problems, we propose the valid derivation of the so-inversion construction. Especially, we claim that the so-inversion construction allows a marked operation, amalgam head movement, which yields such sentences as* Bill must be a genius and so must be Ann. *An important implication of our analysis is that it is necessary to carefully distinguish between peripheral and typical examples in order to construct valid syntactic structures.*

1. Introduction

English has a sentence in (1), the second conjunct of which we call the *so*-inversion construction (hereafter the *so*-construction for short). [1]

(1) *John can speak French, and so can Ann.*

Quirk et al. (1985) claim that the following two assumptions in (2a,b) are crucially related to the *so*-construction and propose that (1) has a structure as in (3):

(2) a. I-to-C movement
b. *So* is an adverb.

(3) $[_{CP}$ *so* $[_{C'}$ *can*$_i$ $[_{IP}$ *Ann* $[_{I'}$ t_i]]]]

Toda (2007), however, argues against the analysis proposed by Quirk et al., pointing out that it cannot account for data such as (4):

(4) *Bill must be a genius and so must be Ann.*

Notice that both *must* and *be* precede *Ann* in (4). It has been generally assumed in the literature that only one head can undergo I-to-C movement. For example, when we convert a sentence like *Ann must be a genius* into its question form, we obtain the sentence *Must Ann be a genius?* but not **Must be Ann a genius?* That is, both *must* and *be* cannot move together. However, (4) allows both *must* and *be* to precede *Ann*. This means that we cannot account for the grammaticality of (4) by I-to-C movement alone. Thus, examples such as (4) cannot be explained only by Quirk et al.'s assumptions given in (2). In order to account for (4), Toda makes the following three assumptions as well as I-to-C movement, proposing that (4) has the structure (6):

(5) a. VP-preposing
b. Subject postposing
c. Obligatory proverbalization (henceforth *so*-replacement)

(6) $[_{CP}\ [_{VP}\ so\ ]_i\ [_{C'}\ [_{C}\ must]_j\ [_{IP}\ t_k\ [_{I'}\ t_j\ [_{VP}\ be\ t_i\ ]]]\ Ann_k]]$

This article is organized as follows. Section 2 points out that the analysis proposed by Toda has both theoretical and empirical problems in it. Section 3 proposes an alternative analysis by adding the following assumption (7) to Quirk et al.'s original assumptions:

(7) Amalgam head movement is allowed in the *so*-construction.

Given (7), (4) has a structure in (8):

(8) $[_{CP}\ so\ [_{C'}\ must\text{-}be_i\ [_{IP}\ Ann\ [_{I'}\ t_i\ [_{VP}\ \phi\ ]]]]]$

It is shown that our analysis does not involve any problems that Toda's analysis has. Section 4 shows, furthermore, that our analysis can also explain examples concerning the *so*-construction which Toda's analysis would fail to account for. Section 5 concludes this article, offering some remarks on the methodology of linguistic analysis.

2. Toda's (2007) problems

First we consider Toda's analysis of the *so*-construction. Toda claims that (4), repeated below, should be derived via the following derivation in (9):

(4) *Bill must be a genius and so must be Ann.*

(9) i. VP-preposing: $[_{CP}\ [_{VP}\ a\ genius]_i\ [_{IP}\ Ann\ [_{I'}\ must\ [_{VP}\ be\ t_i\ ]]]]$

ii. I-to-C movement: $[_{CP}\ [_{VP}\ a\ genius]_i\ [_{C'}\ [_{C}\ must]_j\ [_{IP}\ Ann\ [_{I'}\ t_j\ [_{VP}\ be\ t_i\ ]]]]]$

iii. subject postposing: $[_{CP}\ [_{VP}\ a\ genius]_i\ [_{C'}[_{C}\ must]_j\ [_{IP}\ t_k\ [_{I'}\ t_j\ [_{VP}\ be\ t_i\ ]]]\ Ann_k]]$

iv. *so*-replacement: $[_{CP}\ [_{VP}\ so]_i\ [_{C'}\ [_{C}\ must]_j\ [_{IP}\ t_k\ [_{I'}\ t_j\ [_{VP}\ be\ t_i\ ]]]\ Ann_k]]$

We show below, by investigating each assumption Toda has made, that Toda's analysis above cannot be tenable.

2.1. The validity of VP-preposing and I-to-C movement

Toda claims that VP-preposing triggers I-to-C movement. This assumption is not empirically valid, however. Observe the following contrast:

(10) a. *John wanted to win the race, and win the race he did.*

b. **John wanted to win the race, and win the race did he.*

VP-preposing in general does not trigger I-to-C movement, as shown in (10a); on the other hand, if I-to-C movement is applied coupled with VP-preposing, the sentence is ungrammatical, as shown in (10b). Therefore, Toda's assumption that VP-preposing triggers I-to-C movement is untenable.

2.2. The validity of VP-preposing and So-replacement

Toda assumes that *so*-replacement takes place after VP-preposing. This assumption, however, has a serious empirical problem. Consider the following non-parallelism: [2]

(11) a. If Leslie says Robin believes Terry thinks Kim will go to the movies, then *go to the movies* Robin indeed believes Terry thinks Kim will.

b. Robin believes Terry thinks Leslie will go to the movies, and she believes Terry thinks (that) *so* will Kim**.**

c. *... and she believes *so* does Terry think Kim will / Terry believes will Kim.

As shown in (11a), the long extraction of VP-preposing is possible. If Toda's analysis were right, it should be predicted that *so*-replacement could occur over a clause because *so*-replacement automatically takes place after VP-preposing in his analysis. However, an unbounded dependency cannot be allowed in the *so*-construction, as shown in (11c), where *so* outside the rightmost clause cannot relate with the rightmost sentence; instead *so* inversion should occur locally, as shown in (11b). Therefore, Toda's assumption that

so-replacement takes place after VP-preposing is untenable.

2.3. The validity of subject postposing

Toda argues that the *so*-construction is related to the so-called Locative Inversion construction (hereafter LI for short), which is exemplified in (12):

(12) *Into the room walked {John / *he}.*

Toda assumes that both constructions involve the operation of subject postposing. Toda's assumption, however, includes both empirical and theoretical problems.

First, as shown in (12), LI cannot allow any pronominal subject (cf. Emonds 1976 and Bresnan 1994 among others). The *so*-construction, on the other hand, allows pronominal subjects, as in (13).

(13) *John can speak French, and so can {Ann / she}.*

It is thus clear that subjects of the two constructions behave differently, contrary to Toda's assumption.

Second, Toda does not consider the possibility that LI does not involve the operation of subject postposing. However, Culicover and Levine (2001) strongly argue that the postverbal NP in LI is not postposed but stays in situ in VP. Based on this analysis, Toda's comparison between LI and the *so*-construction does not stand at all since LI does not involve the operation of subject postposing. [3]

Lastly, and more crucially, Toda also faces a serious empirical problem under Culicover and Levine's analysis. They argue that LI is actually a conflation of two quite different constructions: one is light inversion, which is restricted to unaccusative verbs, and the other is heavy inversion, which is not. In their analysis, the notion of 'heavy' corresponds to Heavy NP Shift. They assume that in the case of light inversion, the subject NP does not move but stays in situ in VP, while in the case of heavy inversion, the subject NP is finally postposed to the right of VP (see Culicover and Levine 2001 for more detail). What is important here is that subject postposing occurs in heavy inversion. If subject postposing, as Toda claims, also occurred in the *so*-construction, the postverbal NP of the *so*-construction would be predicted to behave the same way as that of heavy inversion, which includes the operation of subject postposing. It is not true, however: heavy inversion prefers a "heavy" subject to a "light" subject whereas the *so*-construction does not. Let us first observe (14) below:

(14) a. **In the room slept Robin.*

b. **In the room slept fitfully Robin.*
c. *In the room slept fitfully the students in the class who had heard about the social psych experiment that we were about to perpetrate.*

(Culicover and Levine 2001: 293)

In heavy inversion with the unergative verb *sleep*, a "light" subject is hardly allowed, as shown in (14a-b); when the subject is "heavy", the acceptability status of the sentence increases steadily, as shown in (14c). On the other hand, the heaviness of subject does not affect the acceptability of the sentence at all in the *so*-construction, contrary to Toda's expectation. Observe the following sentences:

(15) a. *John is a genius and so is Mary.* [5]
b. *John is a genius and so is the woman who is standing over there with a big smile on her face.* [4]

The figures in the above square brackets posited in each end of the examples in (15) show the results of the acceptability judgment by our informant: 5 stands for the best, and 1 the worst. (15) shows that the acceptability status of the *so*-construction does not improve, rather degrades, if a "light" subject is changed into a "heavy" one. It means that the pattern of the acceptability status of the *so*-construction is just the opposite of that of the acceptability status of heavy inversion. We can, therefore, conclude that the *so*-construction does not involve subject postposing whereas heavy inversion does.

As we have seen above, Toda's assumption that the *so*-construction involves subject postposing is empirically and theoretically untenable.

3. Proposal

In this section, we propose the structure of the *so*-construction, showing that the *so*-construction does not involve any of Toda's problematic assumptions pointed out in the previous section: VP-preposing, subject postposing, and *so*-replacement.

3.1. VP-ellipsis and "positive polarity" adverb So

Firstly, let us consider VP-preposing. We have rejected Toda's claim that VP-preposing triggers I-to-C movement. This means that we do not need to assume VP-preposing in the *so*-construction. So we now have two questions: (i) what operation is actually involved? And (ii) what triggers I-to-C movement?

Let us consider the first question. We claim that the *so*-construction involves not VP-preposing but VP-ellipsis. As a piece of evidence, we point out that the *so*-construction allows "sloppy identity" reading. It is generally assumed that, when VP is elided, sloppy identity interpretation can be obtained (cf. Williams 1977). Observe the following:

(16) *John loves his mother, and so does Bill* [$_{VP}$].

In (16), VP in the second conjunct can allow both interpretations in (17a) and (17b) below.

(17) a. *Bill loves John's mother.*
b. *Bill loves Bill's mother.*

(17a) is the strict reading, whereas (17b) is the sloppy identity reading. What is especially important here is that (16) can allow the sloppy identity reading (17b). This is a strong piece of evidence that VP-ellipsis is actually involved in the *so*-construction. Toda's VP-preposing analysis, on the other hand, cannot predict that the *so*-construction exhibits the sloppy identity interpretation. [4]

As for the second question concerning the trigger of I-to-C movement, we assume, unlike Toda, that the adverb *so* in the *so*-construction triggers I-to-C movement. It is well-known that some kinds of adverbs such as *never* or *only* can trigger I-to-C movement:

(18) a. *Never have I heard such awful news.*
b. *Only reluctantly will Mary eat seafood salad.*

Never and *only* are often referred to as "negative polarity adverbs." Following this, we call adverbial *so* "positive polarity adverbs." We then claim that positive polarity adverb *so*, just like negative polarity adverbs, triggers I-to-C movement. If *so* is an adverb, the operation of *so*-replacement is not necessary in the first place.

We have shown in this section that VP-ellipsis, not VP-preposing, is involved in the *so*-construction, taking a positive polarity adverb *so* as the trigger of I-to-C movement. [5]

3.2. The whole structure of So-construction

Let us consider the subject position of the *so*-construction. As shown in section 2, Toda's analysis that the subject of the *so*-construction should be postposed to the sentence-final position is not valid, because it is based on the empirically and conceptually untenable observation that the subject of the *so*-construction is akin to that of LI. Thus, as far as we do not find any other piece of evidence that verifies subject postposing in the

so-construction, it is safe to assume that subjects in the *so*-construction occupy the typical subject position, that is the Spec of IP. If the subject of the *so*-construction occupies the typical subject position, we can naturally account for the fact that the subject can be pronominal (cf. (13)) and that a "heavy" subject tends to be avoided in the construction (cf. (15)). (13) and (15) are repeated below:

(13) *John can speak French, and so can {Ann / she}.*

(15) a. *John is a genius and so is Mary.* [5]
b. *John is a genius and so is the woman who is standing over there with a big smile on her face.* [4]

Remember here that Toda argues against the analysis proposed by Quirk et al. (1985) because their analysis, where the subject of the *so*-construction is assumed to be in the typical subject position, cannot account for the data (4), which is repeated below:

(4) *Bill must be a genius and so must be Ann.*

By assuming subject postposing, Toda's analysis seems to capture the structure of the above sentence; but there is a problem in his structure of (4), which is repeated below:

(6) $[_{CP}\ [_{VP}\ so\ ]_i\ [_{C'}\ [_C\ must]_j\ [_{IP}\ t_k\ [_{I'}\ t_j\ [_{VP}\ be\ t_i\ ]]]\ Ann_k]]$

As shown in (6), Toda assumes that *must* and *be* occupy different projections. If the structure were valid, an adverb like *surely* could intervene between *must* and *be*. Such intervention, however, cannot be allowed, as shown in (19a):

(19) a. **Bill must be a genius, and so must surely be Ann.*
b. *Ann must surely be a genius.*

Notice that, in a normal sentence (19b), an adverb like *surely* can be inserted between *must* and *be*. These facts lead us to conclude that Toda's assumption that *must* and *be* occupy different projections in the *so*-construction is not tenable. We, therefore, assume that *must* and *be* should form a unit, or amalgam: accordingly, the *so*-construction involves I-to-C movement of amalgam heads. Given the amalgam structure of *must* and *be*, examples such as (19a) can be explained in our analysis because nothing can intervene in the amalgam structure. Notice here that Toda's analysis, where subjects are only postposed to the sentence-final position, cannot account for the ungrammaticality of (19a).

We have so far examined problems with Toda's analysis, pointing out that these problems stem from Toda's assumptions: VP-preposing, subject postposing, and *so*-replacement, all of which are found to be unnecessary in the *so*-construction. We instead posit that (4) has the structure (20).

(20) $[_{CP}$ *so* $[_{C'}$ *must-be*$_i$ $[_{IP}$ *Ann* $[_{I'}$ t_i $[_{VP}$ ϕ]]]]]

(20) can be derived, only if we add an assumption (7) to Quirk et al.'s original assumptions (2a) and (2b), all of which are repeated below:

(2) a. I-to-C movement
b. *So* is an adverb.

(7) Amalgam head movement is allowed in the *so*-construction.

Next section shows that our analysis can also account for further data on the *so*-construction which Toda's analysis cannot explain straightforwardly.

4. Empirical adequacy for our analysis

This section shows three pieces of data which can be accounted for not by Toda's analysis but by ours.

Firstly, let us observe the following:

(21) A: *Tom is very nervous.*
B: *So would you be in his position.*
(Huddleston and Pullum 2002: 1539)

In our analysis, the subject *you* in (21B) occupies the Spec of IP and *would* moves from I to C, which causes no problem. In Toda's analysis, on the other hand, the subject can optionally be postposed to the sentence-final position. However, an ungrammatical sentence like (22) results when *you* in (21B) is actually postposed to the sentence-final position.

(22) **So would be in his position you.*

In order not to generate (22), Toda has to offer some ad hoc assumption.

Secondly, let us observe the sentence (23) below:

(23) *This forecast is admittedly way above the estimate of most analysts in several recent surveys. But so is reality generally far off from the consensus.*
(Huddleston and Pullum 2002: 1539)

In general, VP in the *so*-construction is invisible both in Toda's analysis where VP is changed into *so* and in ours where VP is deleted. In certain circumstances, however, VP is "visible," as in (23). In (23), the second sentence is the *so*-construction including a "visible" VP: *(t_{be}) far off from the consensus*. In Toda's analysis, the preposed VP of the construction must be replaced with *so*. This means that VP exists as the form *so* at the leftmost of the construction even if VP is invisible. Accordingly, we predict that another VP does not appear in the same sentence. Thus, Toda's analysis cannot predict data like (23), where VP is visible in the *so*-construction. Our analysis, on the other hand, assumes that VP is elided. It is generally accepted that ellipsis occurs under identity. However, in (23), the VP in the first sentence is different from that in the second sentence, so that VP-ellipsis in the second sentence cannot occur under identity. That is why VP in (23) remains "visible." Data like (23), though not so often attested, can be accounted for in our analysis.

Thirdly, let us observe the following:

(24) a. *John is a musician and so is Bob.* [5]
b. *John is a musician and so must be Bob.* [4]
c. *John is a musician and so must have been Bob.* [3]
(Numbers in the square brackets refer to our informants' judgment. 5 stands for the best and 1 the worst.)

Notice that the number of auxiliaries between *so* and *Bob* differs in (24a-c). As is obvious from the numbers in the square brackets in (24), the greater the number of auxiliaries intervening between *so* and *Bob*, the lower the acceptability status is. This suggests that the number of auxiliaries affects the acceptability status of the *so*-construction. This fact cannot be predicted nor accounted for by Toda's analysis, because, in Toda's analysis, the subject is postposed irrespectively of the number of auxiliaries intervened. In other words, in his system the number of auxiliaries does not affect the acceptability status of the *so*-construction. In our analysis, on the other hand, auxiliaries form an amalgam, which then undergoes I-to-C movement. We assume this amalgamating operation to be a kind of marked rule: as shown in (24), the acceptability status degrades as the number of intervening auxiliaries increases. This means that the gradation of the acceptability judgment in (24) can be captured in terms of the operation of amalgamation.

To sum up, Toda's analysis cannot account for the gradation of the acceptability judgment in (24), whereas ours can do so nicely on the basis of the marked nature of amalgamation in English. [6]

5. Concluding remarks and further implications

This article has critically examined the analysis of the *so*-construction proposed by Toda (2007). Toda argues against the analysis proposed by Quirk et al. (1985), on the basis of his counterexamples like (4) repeated below:

(4) *Bill must be a genius and so must be Ann.*

We have, however, shown that (4) can be accounted for by Quirk et al. (1985) by further assuming (7), which is repeated below:

(7) Amalgam head movement is allowed in the *so*-construction.

We have also pointed out that Toda's analysis involves both theoretical and empirical problems. Therefore we have concluded that Toda's analysis is untenable. This article has instead claimed that (4) has the structure (20), repeated below:

(20) $[_{CP}$ *so* $[_{C'}$ *must-be*$_i$ $[_{IP}$ *Ann* $[_{I'}$ t_i $[_{VP}$ ϕ]]]]]

Differences between Toda's analysis and ours are summarized in (25):

(25)

	Toda's Analysis	Our Analysis
VP	preposing	ellipsis
must be	no amalgam	amalgam
so	proverb	adverb
subject position	sentence-final	the Spec of IP

As shown above, Toda (2007) regards (4) as a counterexample to Quirk et al. (1985). (4) is, however, considered to be a peripheral example; as for the examples involving a series of auxiliaries in the *so*-construction such as (4), their acceptability status varies a great deal among native speakers of English. Most informants we consulted with do not allow even the *must-be* type amalgam. As we have shown in section 4, even informants who allow amalgam movement relatively freely do not allow the amalgam structure which

consists of more than 2 heads such as *must-have-been*. This fact leads us to conclude that amalgam movement is a kind of marked rule even in the *so*-construction.

Given our analysis, we can answer the following question: where does the peripheral nature in (4) come from? The answer is: amalgam I-to-C movement which (4) involves. Thus (4) can be generated by the interaction of some general operations with the marked amalgamation operation. In contrast, it is difficult for Toda to account for the peripheral nature in (4), since he has built his analysis without distinction between peripheral and typical examples. This methodological flaw causes Toda's analysis not only to be technically complicated but also to involve many theoretical and empirical problems (see section 2). This suggests that it is very important for linguistic analysis to carefully distinguish the data under investigation.

Notes

* We are thankful to three *The Linguistic Review* reviewers for their many constructive comments and suggestions. Their constructive criticism has been especially important in improving the quality of this paper. Remaining inadequacies are of course our own.

** The authors' names are given in alphabetical order.

1. As Quirk et al. (1985) say, we also distinguish the *so*-construction from initial *so* without inversion, which is exemplified below:

   (i) A: *John drives a car.*
   B: *So he does.*

   Unlike the *so*-construction, the sentence where *so* is not inverted means the agreement of what someone says.

2. We thank an anonymous reviewer for providing the data in (11).

3. We thank an anonymous reviewer for pointing out this possibility.

4. As a reviewer points out, the following contrast can be a counterexample to the VP-ellipsis analysis:

   (i) a. **Bill must be a genius, and Ann must, too.*
   b. *Bill must be a genius, and so must Ann.*

   As shown above, the *be* verb can be dropped in the *so*-construction (see (ib)) whereas it cannot in the normal word order (see (ia)). If VP-ellipsis occurs in both cases, there should be no difference of grammaticality concerning *be*-drop.

   This reviewer's claim is appropriate only under the assumption that *be*-drop is the phenomena peculiar to the *so*-construction. However, *be*-drop can occur in other constructions. Observe the following sentences:

   (ii) *John is taller than Mary (is).*
   (iii) A: *John is a coward.*
   B: *But not I.*

   As shown above, *be*-drop is seen in various situations. It means that the existence of *be*-drop is not necessarily relevant to the existence of VP-ellipsis. Under what circumstance *be*-drop can occur is an interesting question worth pursuing, but it is irrelevant to the discussion here.

5. See also Huddleston and Pullum (2002) and Hankamer and Sag (1976) for the assumption that the *so*-construction involves ellipsis and *so* is an adverb.

6. Note here that the existence of amalgam head movement in the *so*-construction is logically induced both empirically and theoretically. Yet we do not have any decisive answer to the question as to under what circumstance amalgam head movement is possible. However, there is one possibility. Observe the following data concerning negative inversion (henceforth NI):

(i) a. *John is a musician, but never is Bob.*
b. ?*John is a musician, but never must be Bob.*
c. **John is a musician, but never must have been Bob.*

As shown by the acceptability judgment above by our informant, amalgam head movement is also possible in NI. Observe further the following examples provided by a reviewer:

(ii) a. *What must Ann be thinking?*
b. *What must be Ann thinking?*
c. **What must have been she doing?*

It should be noted here that not only *wh*-phrases but also *so* in the *so*-construction and negative items in NI occupy the CP-Spec. Thus we can propose the following generalization:

(iii) Amalgam head movement is possible only when a phonetic element is in CP-Spec.

The above generalization can be borne out by the following example:

(iv) **Must be Robin sleeping?*

As shown above, amalgam head movement is not possible when no phonetic operator appears.

More interestingly, there is acceptability difference among the constructions concerning amalgam head movement (we think that this fact can be reduced to the markedness of the operation of amalgam head movement). As shown above, both NI and *wh*-questions tend not to allow more than two amalgam head movements. On the other hand, more than two amalgam head movements is possible in the *so*-construction, as shown below (see also (24)):

(v) ?*He must have been upset, and so must have been she.*

Thus we can say that the acceptability of amalgam head movement depends on the type of element or operator in CP-Spec. It is a very intriguing issue worth pursuing, but to discuss it in detail is beyond the scope of our article.

We thank anonymous reviewers for providing an opportunity to consider the condition of amalgam head movement.

References

Bresnan, Joan. 1994. Locative inversion and the architecture of universal grammar. *Language* 70. 72-131.

Culicover, Peter W. and Robert D. Levine. 2001. Stylistic inversion in English: A reconsideration. *Natural Language and Linguistic Theory* 19. 283-310.

Emonds, Joseph. 1976. *A transformational approach to English syntax*. New York: Academic Press.

Hankamer, Jorge and Ivan Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7. 391-428.

Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Toda, Tatsuhiko. 2007. *So*-inversion revisited. *Linguistic Inquiry* 38. 188-195.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Williams, Edwin. 1977. Discourse and logical form. *Linguistic Inquiry* 8. 101-139.

Yuji Hatakeyama
Tokyo University of Agriculture and Technology

Kensuke Honda
Ibaraki National College of Technology

Kosuke Tanaka
Shinshu University