An HPSG Approach to English Comparative Inversion*

Dongwoo Park
(Seoul National University)

Park, Dongwoo. (2017). An HPSG Approach to English Comparative Inversion. *Language Research*, 53.2, 203-230.

It has been analyzed that the word order of English comparative inversion is analogous to that of other subject-auxiliary inversion constructions, in that only a finite auxiliary verb (i.e., the highest auxiliary) can be followed by the subject. However, English comparative inversion must be distinguished from other inversion constructions, since the subject can be located between a cluster of auxiliary verbs and the non-auxiliary phrase in English comparative inversion. Existing analyses of subject-auxiliary inversion cannot account for this special kind of inversion. This paper proposes a new phrase type, called *inv-focus-cl*, for English comparative inversion within the framework of construction-based Head-driven Phrase Structure Grammar (HPSG). Additionally, I suggest that constraints on properties of lexemes participating in the *inv-focus-cl* are governed by the construction-based approach, while the word order of English comparative inversion is determined by independent rules that determine the word order of lexical items, as the word order domain approach suggests.

Keywords: English comparative inversion, construction-based approach, word order domain approach, focus inversion, HPSG

1. Introduction

English Comparative inversion (henceforth CI) has been analyzed in the same way as other inversion constructions (Merchant 2003 and

* I would like to thank three anonymous reviewers for their comments. My gratitude also goes to Prof. Eun-Jung Yoo for her invaluable suggestions and advice. This work is based on my M.A. thesis. All remaining errors are solely mine.

Maekawa 2007, among others). This is because both the former and the latter seem to have the same word order: only the finite auxiliary verb (i.e., the highest auxiliary) can precede the subject.

(1) a Humans can climb trees more carefully than can monkeys. [CI]
b. Have you ever been to Seoul? [interrogative inversion]
c. Not until the evening did John find his son. [Negative inversion]
d. Had John finished his homework, he would be with us now. [*If*-less inversion]

However, Culicover and Winkler (2008) provide some examples indicating that, unlike other inversion constructions, CI allows the subject to be preceded by more than one auxiliary verb, as illustrated in (2).

(2) a. Who was responsible for keeping the records would be a more reliable witness as to their accuracy as a whole than would be any of the original makers.
b. To her, thinking, as she ever was thinking, about Johnny Eames, Siph was much more agreeable than might have been a younger man. (Culicover and Winkler, 2008)

Additionally, a host of data supporting this fact can be found from books and corpora such as the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA) as follows.

(3) a. It is no more expensive than would be the system you are proposing. (Huddleston and Pullum, 2002)
b. White women in our study would have used relatively more IAAT than would have the black women.
c. The Relief and Aid Society was a genuinely civic-minded organization that very possibly did administer the world's contributions more efficiently and honestly than could have the city government.

d. Her name on that list affected me more than would have divorces from a dozen Kathyrns.

These examples are quite peculiar because existing analyses of subject-auxiliary inversion do not have any method to locate more than one auxiliary verb before the subject in subject-auxiliary inversion constructions. To be specific, no analysis assuming T-to-C head movement allows a cluster of auxiliary verbs to move to C. Additionally, subject-auxiliary inversion phrase (*sai-ph*) in HPSG also permits only one finite auxiliary verb to precede the subject.

The subjects in the sentences in (2) and (3) are located at the sentence final position, which might lead some researchers to regard this inversion as Heavy NP Shift (HNPS). However, the sentences in (4) illustrate that CI may not be derived through HNPS.

(4) a. Ali would have driven a car to the park more eagerly than would have the students (in our class on environmental consciousness) to the concert. (Potts, 2002)
b. Jim would have translated the English much better than would have students in his class read the Spanish.
c. John could have read French more fluently than could have Joe.
d. Don would have been more proud of what he had achieved than would have been Bill.

In (4a) and (4b), each subject in the comparative clauses is followed by PP and VP, respectively. If this type of inversion were HNPS, it would be predicted that the subjects should be located in the right of other verbal arguments and/or adjuncts (Ross 1967; Kayne 1998, among others). Besides, the inverted subjects in (4c) and (4d) are one-syllable proper nouns. Even though it is difficult to define to what extent 'heavy' can cover, it is unreasonable to consider those proper nouns as heavy NPs, since they are not long enough. Additionally, to the best of my knowledge, unlike objects, subjects do not undergo HNPS in English. Thus, the con-

clusion can be drawn that CI is not derived through HNPS.

This paper aims to propose constraints for capturing the word order of English CI by suggesting a new phrase type within the framework of construction-based HPSG. Additionally, this paper suggests that the word order of the lexical items (or, linearization) is determined by independent word order domain rules. The rest of this paper is structured as follows. In section 2, I present recent analyses of CI and their problems. In section 3, I propose a new approach to explain the word order of CI within the framework of construction-based approach HPSG and introduce word order domain rules that can apply to all phrases in English, including the new phrase for CI. In section 4, I examine and critique alternative approaches that are based on HPSG. Section 5 presents concluding remarks.

2. Previous Studies on CI and Problems

2.1. Culicover and Winkler (2008)

Culicover and Winkler (2008) note that a cluster of auxiliary verbs can be followed by the subject in CI as shown in (2) and (3). They mention four logical possibilities to derive the word order of CI, as shown in (5).

(5) a. The subject is in canonical subject position (e.g. Spec IP) and all of the verbs move to the left;
b. The subject is in canonical subject position and moves to the right.;
c. The subject is in canonical subject position, and everything in I' moves to the left of it;
d. The subject is in situ in Spec *v*P, and remains in situ.

They argue that (5d) is the easiest and most plausible possibility. In order for (5d) to work, they propose that *than* is a complementizer, and selects TP whose head does not bear the EPP. This suggestion is based on the assumption that the subject stays in-situ in Spec, *v*P, and auxiliary verbs

are generated higher than the subject. They suggest that the suspension of the EPP is possible only when the following three rules, based on Selkirk (2005), are satisfied.

(6) Align R(Comma, ip) (Selkirk 2005:7)
Align the right edge of a constituent type Comma Phrase in syntactic representation with the right edge of an ip in phonological representation.

The constraint in (6) is indicative of the correspondence between clausal syntactic constituents and intonational phrases.

(7) Contrastive-Focus-dominate-Δip (FOC/Δip) (Selkirk 2005: 18)
The terminal string of a contrastive FOCUS constituent in syntactic representation correspond to a string containing the metrical prominence of an Intonational Phrase in phonological representation.

(7) demonstrates that there is a close relationship between contrastive foci and metrical prominence of an ip.

(8) Right Edge Alignment of Focus (REAF) (cf. Truckenbrodt 1995, Selkirk 2004)
Each focused element is right aligned in ip.

At last, (8) specifies the position where foci occur — the right edge of ip.

The sentences in (9) show that whether these three constraints above are satisfied can result in two different comparatives. Capitalization stands for the metrical prominence.

(9) a. ?Anna ran much faster (than could have MANNY)$_{ip}$.
b. Anna ran much faster (than MANNY could have)$_{ip}$.

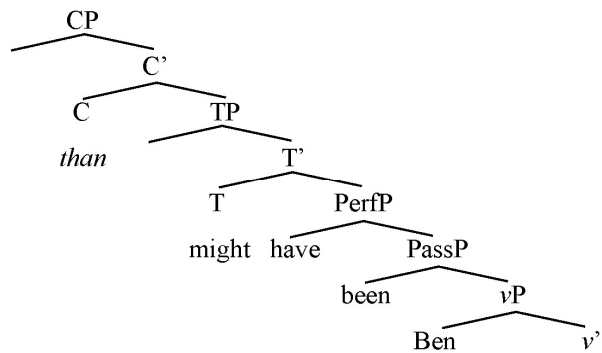
In (9a), the three constraints are conformed: *than could have MANNY* corre-

sponds with an intonational phrase. Additionally, the contrastively focused subject *MANNY* has a metrical prominence of the ip, and is right aligned in the ip. These result in the EPP suspension. On the other hand, the EPP is not suspended in (9b), since this sentence violates the REAF which prevents the subject from moving to Spec, TP. In brief, they suggest that the EPP competes with the REAF. Thus, the REAF is stronger than the EPP in CI, while the EPP is stronger than the REAF in canonical comparative constructions. However, this approach has a non-trivial problem. This analysis cannot explain the cases where auxiliary verb phrases are elided, as shown in (10).

(10) John might have been injured much more severely ···
than might have been Ben.
than might have Ben.

(10b) can be interpreted as (10a). The syntactic structure for (10a) can be roughly described as in (11).

(11)



According to Culicover and Winker's analysis, the subject must remain in its base-position, namely Spec, *v*P, due to the constraint in (8). However, if any further movement of the subject were not assumed, it is predicted that (10b) would not be able to be generated. The reason is as follows: in order for *been* to be elided, PassP must be deleted. Since the subject *Ben* is located inside the ellipsis site, it must be deleted along with PassP.

Even if a feature that triggers the displacement of the subject and a proper landing site were postulated (i.e. the prospective landing site must be lower than *have* and higher the ellipsis site), it would be no more than a stipulation unless further evidence is provided.

One might claim that (10b) can be generated through lowering of *been* into some phrase below the subject followed by the elision of a phrase containing *been*. However, I reject this possibility, since downward movement violates the Proper Binding Condition, which requires that traces be bound, and the Extension condition, which requires that movement extend the root of the structure that it applies (Chomsky 1993).

2.2. Maekawa (2007)

Based on Kathol (1995, 2000, 2001), who tries to explain the linear word order of German by means of 'topological field' within HPSG, Maekawa (2007) suggests the distribution of domain elements in English, as represented in (12). This specifies by what element each topological field can be occupied. As the name 'topological field' indicates, sentences are divided into several fields, and each field is occupied by certain domain elements.

(12) Distribution of domain elements in English

first	matrix non-subject *wh*-phrases, preposed negative phrases, etc.
second	finite auxiliary verbs in subject-auxiliary inversion (SAI) sentences, complementizer, subordinate non-subject *wh*-phrases.
third	subjects
fourth	finite verbs in non-SAI sentences
fifth	complements of the finite verbs

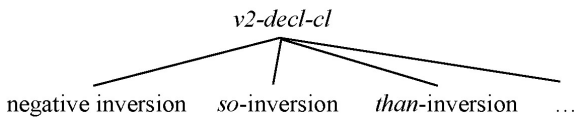
To determine the word order of English, two additional Linear Precedence (LP) constraints are assumed. The first one is to deal with the order of fields, as illustrated in (13), and the other has to do with the cardinality restriction imposed to the first and the second fields, as shown in (14).

(13) Topological Linear Precedence Constraint for English (Maekawa 2007: 181)
first < second < third < fourth < fifth

(14) Topological Uniqueness Condition (Kathol 2001:51)
first < first
second < second

In (13), 'A < B' means that A is followed by B in linear order. For example, elements assigned to the first topological field always precede those assigned to the other topological fields. Meanwhile, the constraint in (14) illustrates that the first and the second fields must contain only one element. On the basis of the LP constraints mentioned above, Maekawa (2007) characterizes CI as an instance of declarative verb-second clause (*v2-decl-cl*), in which a finite auxiliary verb is located in the second field. The subtypes of *v2-decl-cl* are described as in (15).

(15) Subtypes of *v2-decl-cl*



These inversion types are classified based on what sort of element occupies the first field. In the case of *than*-inversion (i.e., CI), the first field is occupied with *than*.

However, this approach is insufficient to capture the exact characteristics of CI. The problem is that it cannot generate comparative inversion sentences where more than one auxiliary verb is followed by the subject. Generally, it is analyzed that the non-finite auxiliary verb phrase following the finite auxiliary verb is the complement of the finite auxiliary verb. Then, the complement of the first auxiliary verb should be located in the fifth field, according to (12). This cannot explain the way a cluster of auxiliary verbs precedes the subject in CI. Even if we proposed a new constraint allowing auxiliary verbs to be placed before the subject, the

problem would still remain. Recall that the subject should be contained in the third field, and the complementizer *than* is the element which is contained in the first field. Then, non-finite auxiliaries must be located in the second field. However, the constraint mentioned in (14) prohibits more than one auxiliary verb from occupying the second field. Consequently, sentences such as (2b), (4d), and (10a-b) cannot be generated through this analysis.

Postulating another field between the second field and the third field or eliminating the rule (14b) might be a way of explaining the word order of CI where non-finite auxiliaries can be located between the finite auxiliary and the subject. However, these alternatives have a non-trivial problem. It is not clear how this analysis account for why non-auxiliary verbs can be followed by the subject in CI, but not in other canonical SAI constructions.

3. Proposal: A Construction-based Approach Combined with the Domain-based Approach to Word Order

As mentioned before, CI allows the subject to be preceded by the cluster of the auxiliary verbs, as illustrated in (16a). Additionally, the inverted subject can be followed by the phrase with a contrastive focus meaning, as shown in (16b) and (16c).

(16) a. Megan can jump higher than could have Bill.
b. John read French more fluently than could have Joe spoken English.
c. ?Mary would have been angry much longer than would have been John, happy.

(Culicover and Winkler, 2008)

This section provides an analysis to account for the word order of CI. In section 3.1, I will briefly review the basics of two HPSG systems I will use in this paper - the construction-based approach and domain-based

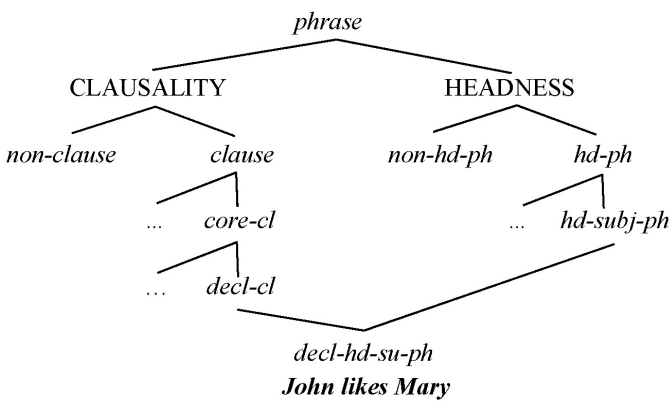
word order approach to word order. In 3.2, I will show how the interaction between these two approaches can explain the word order of CI.

3.1. Theoretical Framework

In this section, I will introduce some basic notions widely accepted in HPSG. I will begin with constraints which delineate how phrases are composed of in syntactic aspects, using a means of multidimentional hierarchy of phrases. Domain-based order rules determining the order of components in phrases will be outlined as well.

Sag (1997) and Ginzburg and Sag (2001) classify phrases with two dimensions — HEADNESS and CLAUSALITY. This is called *multiple inheritance hierarchy*, which means every phrasal type has to follow the constraints from both HEADNESS and CLAUSALITY, and all subtypes of phrases have to inherit constraints imposed to their hyper-types of phrases. HEADNESS describes the constraints showing how every head daughter is related to non-head daughters to compose certain types of phrases (e.g. h(ea)d-complement phrase, subject-auxiliary inversion phrase). Meanwhile, CLAUSALITY is the dimension characterizing whether or not certain phrases are clauses, and what kind of clauses they are (e.g. interrogative clause, declarative clause, relative clause). For instance, a finite declarative clause, which is named as *decl-hd-su-ph*, can be described in (17).

(17)



In what follows, I will briefly review the domain-based approach to word order suggested in Reape (1994). Reape introduces a way of arranging words by means of the feature DOM(ain). This feature is a *sign*-level feature. Since the value of the DOM features is a list of signs, it contains PHON and SYNSEM information. When daughters are merged, elements in word order domain of them are put together in mother node's word order domain. If the elements of two word order domains are merged, the word order of them can be changed by means of the *shuffle* relation.

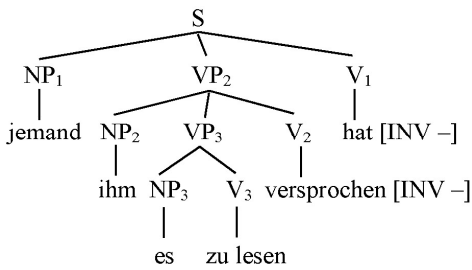
However, in order to rule out sentences with the illegitimate word order, two constraints are proposed — one is linear precedence (LP) rules and the other is the preservation of domain order at the daughter stage. The latter requires that the original domain order of each daughter should be preserved when they are merged. For example, suppose that a domain α <A> and the other domain β <B, C> are merged. In the domain β, B is followed by C. Then, the possible orders of mother's domain are <A,B,C>, <B, A, C>, and <B,C,A>. On the other hand, <C,A,B>, <A,C,B>, and <C,B,A> are ruled out, since B is preceded by C.

This system employed by Reape can explain the relatively free order of German, as illustrated in (18). There is discontinuity between verbs and their nominal arguments.

(18) a. daβ es ihm jemand zu lesen versprochen hat.
that it.ACC him.DAT someone.NOM to read promised has
'that someone promised him to read it.'
b. daβ ihm es jemand zu lesen versprochen hat.
c. daβ jemand es ihm zu lesen versprochen hat.
d. daβ jemand ihm es zu lesen versprochen hat.
e. daβ es jemand ihm zu lesen versprochen hat.
f. daβ ihm jemand es zu lesen versprochen hat.

Reape suggests that the sentences above are derived from a single structure, illustrated in (19).

(19)



The operation *shuffle* can generate the sentence in (18) in German. However, without any further constraints, it would cause an over-generation problem. In order to rule out sentences with a wrong word order, Reape suggests the following two LP rules. ("A < B" means that A is followed by B.)

(20) a. NP < V
b. V < [HEAD-DTR | V [INV —]]

The first rule requires that NPs be always followed by verbs in any domain. The second rule ensures that a verbal head whose INV(ersion) value is negative is located to the right of any verb. This makes *versprochen* and *hat*, which are marked as [INV —], placed to the right of any other verbal constituents in the domain. Due to these rules, neither sentences where NPs are preceded by verbs nor sentences where a verbal head with [INV —] is located to the left of any verb can be generated.

Alongside the *shuffle* operation, Reape introduces the UN(ioned) feature. The value of this feature is either negative or positive. A domain marked as [UN —] is frozen like an inseparable cluster. For instance, when the domain β <B, C> has the [UN —] feature, merge of the domain α <A> and the domain β <B, C> yields two word order - <A,B,C> and <B,C,A>. A word order where A is placed between B and C cannot be generated. On the other hand, a domain marked as [UN +] allows another domain to cut in the domain elements it contains. When the domain α <A> merges with the domain β <B, C> with the [UN +] feature, <A,B,C>, <B,C,A>, and <B,A,C> can be generated.

3.2. Analysis

In this section, I suggest a novel analysis of CI, whereby both the construction-based approach and the domain-based approach to word order are used. In the existing construction-based approach, the word order of English is determined by the Immediate Dominance (ID) rules. In the present analysis, however, the word order is determined by independent word order domain rules. That is to say, constraints on the properties of the lexemes participating in certain phrases are governed by the construction-based approach, while the word order of lexical items inside the phrases is determined by domain rules.

First, I suggest some domain rules necessary to capture the word order of CI. I assume here that the default value of the UN feature a phrase is negative in the absence of any additional constraint, following Bonami and Godard (2003). This assumption prevents word order domain rules from generating sentences with the improper word order by means of the *shuffle* operation.

In the present analysis, all phrases need word order rules which allow every element in the phrases to be located at the right position. The first domain rule specifies the order between the head and its complement. In all phrases in English, the head is always followed by its complement. This basic rule can be represented as in (21).

(21) Head-complement rule:

$$\text{DOM}\,\langle[\text{COMPS}\,\langle[\text{HEAD}\,\boxed{1}]\rangle]\rangle \;<\; \text{DOM}\,\langle[\text{HEAD}\,\boxed{1}]\rangle$$

(21) requires that a domain element A should follow another domain element B that takes the domain element A as a head of its complement. For instance, in the sentence *John must be arrested*, the head *be* must be preceded by *must* which takes a complement headed by *be*. Without this rule, a sentence where *be* is followed by *must* could be generated.

Another rule relevant to the current analysis has to do with the position of the subject. The position of the subject is pivotal to identify the clausal type. To illustrate, in most declarative clauses, subjects are followed by

finite verbs. However, subjects are preceded by finite auxiliary verbs in interrogative clauses. The position of the subject depends on the value of the INV feature of verbs. The finite auxiliary verb in the declarative clause has [INV —], while that in the interrogative clause contains [INV +]. A rule that captures this word order is described as in (22).

(22) Subject rule:

$$\left[\begin{matrix}\text{INV} + \\ SUBJ \quad \langle \boxed{2} \rangle\end{matrix}\right] < \boxed{2} < \left[\begin{matrix}\text{INV} - \\ SUBJ \quad \langle \boxed{2} \rangle\end{matrix}\right]$$

The constraint (22) requires that all auxiliary verbs with [INV +] precede the subject, and that those with [INV —] follow the subject.

Along with the word order domain rules above, a phrase type CI belongs to must be identified. This is related to explaining where the constraints of the phrase type for CI inherit from, according to multiple inheritance hierarchy.

In canonical SAI constructions where only a finite auxiliary verb is followed by the subject, non-finite auxiliary verbs have [INV —] by default in the absence of any further constraint, which assures that the non-finite auxiliary verb in every English phrase follows the subject. This means that CI cannot be analyzed with existing types of canonical SAI phrases, since all non-finite auxiliaries in CI are preceded by the subject, which means that they must have [INV +].

This translates into the need to postulate a new phrase type that forces the non-finite auxiliary verbs participating in CI to have [INV +]. Additionally, the prospective phrase type must allow the subject to shuffle with the elements in the auxiliary verb phrase. If the subject merged with the auxiliary verb phrase with [UN —], it would be erroneously predicted that the subject would not be able to be located between the cluster of auxiliary verbs and the non-auxiliary phrase. (See (16b-c)).

The phrase type must evince that a finite auxiliary verb must sub-categorize for either VP with [INV +] and [UN +] or a phrase with [AUX —]. The former allows the subject to be located after the auxiliary verbs when CI has two auxiliary verbs, while the latter can capture the word

order of CI that has only one finite auxiliary verb.

One question arises here is why the subject is preceded by more than one auxiliary verb in CI. I adopt the suggestion by Culicover and Winkler (2008) and Gergel, Gengel, and Winkler (2007) that this inversion is caused by the information structure restriction that the inverted subject must be interpreted only as a focus, especially contrastive focus. This is why Culicover and Winkler (2008) regard CI as a type of focus inversions. According to Gundel and Fretheim (2004), a contrastive focus is a material that plays a role in calling to the hearer's attention and mentioning contrasts with other entities. In CI, the inverted subject contrasts with the subject in the main clause.

Culicover and Winkler (2008) use the behavior of epithets to show that the inverted subject in CI is restricted to be interpreted as a contrastive focus as follows.

(23) a. Bill Clinton$_i$ said more than the president$_i$ could have.
b. Bill Clinton$_i$ said more than could have the president$_j$.
c. Bill Clinton$_i$ said more than the president$_j$ could have.
d. *Bill Clinton$_i$ said more than could have the president$_i$.

(23a) shows that a coreferential reading is possible because the subject in comparative clause is not interpreted as a contrastive focus. The subject without a contrastive meaning can precede the auxiliary verbs in the comparative clause. This entails that the canonical subject position is not a position only for contrastive foci. On the other hand, the contrast between (23b) and (23d) manifests that only the subject conveying a contrastive focus meaning can be preceded by a cluster of auxiliary verbs in CI.

Gergel, Gengel and Winkler (2007) also argue that only elements with a contrastive focus meaning can occupy the inverted subject position in CI, by using the following sentences:

(24) a. Manny$_i$ plays the piano better than did HE$_{*i/j}$.
b. He$_i$ said he could play the piano better than did HE$_{*i/j}$.

The sentences in (24) indicate that the pronoun non-coreferential to the subject in the main sentence can follow the auxiliary verb, while the pronoun subject without a contrastive focus meaning cannot undergo the subject-auxiliary inversion.

This delineates the close relationship between information structure and a specific phrase type, because the subject in this phrase type must be interpreted only as a contrastive focus. Thus, I propose that information structure of the subject in CI must be specified as a constraint of the phrase for CI. That is, the subject in CI contains the INFO-STRUC|FOC (cf. Engdahl and Valludi 1996; Maekawa 2004).

All the constraints for CI that must be taken into account are put together in the following new phrase type. I will call this phrase *inv-focus-cl*.

(25) $$\textit{inv-focus-cl}: [\quad] \rightarrow H\begin{bmatrix} \text{SUBJ} & \langle[\text{FOC} \quad \boxed{A}]\rangle \\ \text{COMPS} & \langle\begin{bmatrix}\text{INV}+\\ \text{UN}+\end{bmatrix} \mid [\text{AUX}-]\rangle \end{bmatrix}, \ldots$$

In (25), the head of *inv-focus-cl* combines with the subject conveying a focus information and VP as its sisters. The head of *inv-focus-cl* (i.e. a finite auxiliary verb) can take either a phrase containing [INV +] and [UN +] or a phrase bearing [AUX —] as a complement. When it takes the former, the constraints in *inv-focus-cl* and the word order domain rules, which were suggested above, allow the subject to be located between the second auxiliary verb and the non-auxiliary phrase as follows:

(26) a. John might have eaten cookies faster than might have Paul made.
b. Mike wrote more books than would have John read.

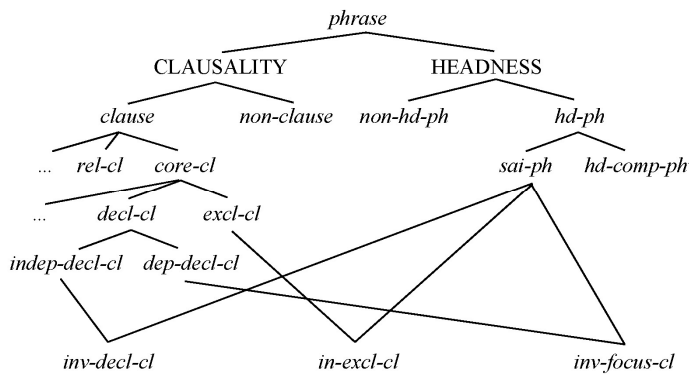
On the other hand, when the head of *inv-focus-cl* takes a phrase with [INV —] as its complement, sentences such as (27) can be generated.

(27) a. Over the years I find that I have a lot more patience than do I have MONEY.

b. MARY read FRENCH much better than does JOHN read GERMAN.

Recall that all the phrase types are defined by two dimensions — CLAUSALITY and HEADNESS (Ginzburg and Sag 2001). In other words, the constraints of a certain phrase must show its clause type on the one hand, and the relationship between the head and its complement on the other hand. The constraints of *inv-focus-cl* inherit from both *dep(endent)-decl(arative)-cl(ause)* and *inv(ersion)-ph(rase)*. This is because *inv-focus-cl* cannot stand alone, and the subject is preceded by a finite auxiliary verb. Thus, the location of *inv-focus-cl* in the phrasal type hierarchy can be sketched as in (28).

(28)



Yet, *inv-focus-cl* is not sufficient, because this phrase type cannot account for the word order of CI when a cluster of auxiliaries consists of three auxiliary verbs. This means that *inv-focus-cl* cannot force the third auxiliary verb to have [INV +], because all the non-finite auxiliary verbs have [INV —] by default. Then, *inv-focus-cl* cannot guarantee the word order of (2b) and (16c), repeated here as (29a) and (29b), respectively.

(29) a. To her, thinking, as she ever was thinking, about Johnny Eames, Siph was much more agreeable than might have been a younger man.

b. ?Mary would have been angry much longer than would have been John, happy.

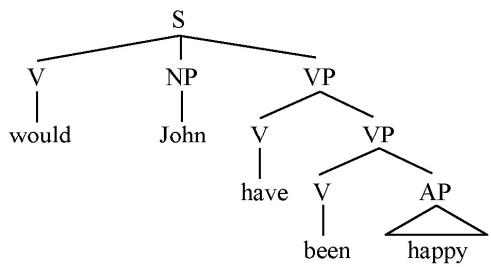
In order to capture the word order of CI that cannot be covered by the constraints in *inv-focus-cl*, an additional rule is necessary. The prospective rule must be able to guarantee that the complement of the second auxiliary verb should satisfy following condition: Either VP whose head is an auxiliary verb must contain [UN +] and its head should have [INV +] or a phrase must have [AUX —]. This constraint is represented as in (30).

(30) [INV+] verb rule:
When a non-finite verb with [INV +] subcategorizes for an auxiliary verb phrase P, P must be [INV +] and [UN+].[1)]

This rule requires that all the non-finite auxiliaries in CI have [INV +]. This is possible because *inv-focus-cl* guarantees that the second auxiliary verb contains [INV +], and then the [INV +] verb rule applies to all non-finite auxiliary verbs in CI. If an auxiliary verb phrase P selected by a non-finite verb with [INV+] lacked [UN+], it would be predicted that the sentences in (29) could not be generated. The reason is as follows: let us take (29b) as an example. The non-finite verb *have* subcategorizes for the auxiliary verb phrase *been happy*. If the verb phrase did not have [UN+], then the subject *John* could not be located between *been* and *happy*.

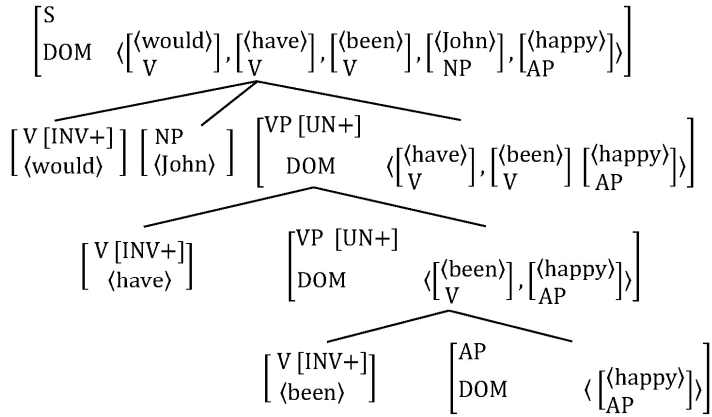
When the constraints that have been discussed so far are put together, the syntactic tree and the word order tree for (29b) can be represented as in (31a) and (31b), respectively

(31) a. Syntactic tree



1) I am indebted to an anonymous review for his/her refining this constraint.

b. Word order tree



The discrepancy between the syntax tree and the word order tree is caused by the fact that all the syntactic properties of CI except the word order come from the constraints adopted by the construction-based approach, while the word order is determined by the word order domain rules.

In (31b), the domain order of the lowest VP is determined by the head-complement rule in (21), and thus, *been* precedes *happy*. When *have* combines with VP containing [UN +], it merges with VP headed by the auxiliary verb with [INV+] due to the [INV+] verb rule in (30). Additionally, due to the head-complement rule, *have* is followed by *been*. At last, the domain elements in S are arranged by the subject rule in (22), as well as the head-complement rule in (21). Then, *could* precedes *have*, and the subject is located between *been* and *happy*.[2)]

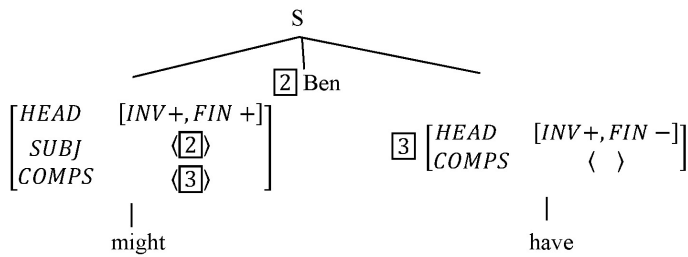
The analysis advanced in this paper can also account for the ellipsis facts in (10), repeated here as (32).

(32) John might have been injured much more severely …
  a. than might have been Ben.
  b. than might have Ben.

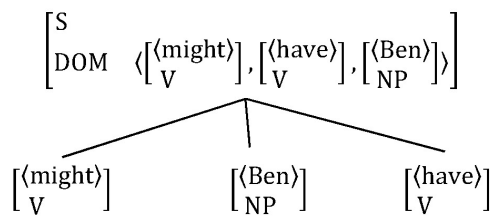
2) An anonymous reviewer points out that the present analysis seems to be similar to Maekawa's analysis, in that both use the domain-based approach. However, unlike Maekawa's analysis, the current analysis combines the construction-based approach and the domain-based approach. See section 4 for more information about problems of an analysis which uses the domain-based approach only.

In HPSG, the COMPS value of an auxiliary verb can be an empty list through the ellipsis lexical rule. That is, a head whose complement is elided no longer contains COMPS values. (Pollard and Sag 1994; Sag et al. 2001, Kim 2006, Ginzburg and Miller 2015, among many others). Given this, (32b) is generated as follows:

(33) a. syntactic tree



b. word order tree



In (32b), the COMP list of *have* is empty. Meanwhile, in order for the sentence (32a) to be generated, the COMP list of *been* must be empty. This analysis can avoid the crucial problem Culicover and Winkler (2008) encounter (see (10) and the surrounding discussion).

In the same vein, the auxiliary *can* in (1), repeated here as (34) can be illustrated as in (35).

(34) Humans can climb trees more carefully than <u>can monkeys</u>.

(35) Lexical entry of *can* in (34)

$$\begin{bmatrix} HEAD & [INV+, FIN+] \\ SUBJ & < A > \\ COMPS & < > \end{bmatrix}$$

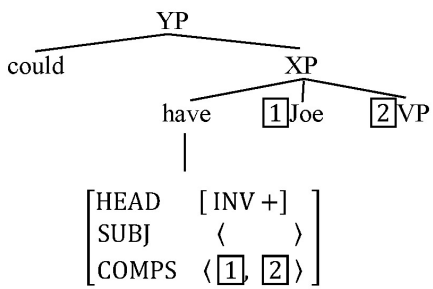
To summarize, I suggest the word order domain rules which make the subject with a (contrastive) focus meaning in English CI follow an auxiliary cluster, by using the DOM feature and the UN feature, and by proposing a new phrase type called *inv-focus-ph*. The constraints of *inv-focus-cl* inherit from both *dep(endent)-decl(arative)-cl(ause)* and *inv(ersion)-ph(rase)*. It takes either a VP with [INV +] and [UN +] or a phrase with [AUX −] as its complement.

4. Possible Alternatives and Issues

In the previous section, I proposed an analysis of English CI through combining the construction-based approach and the domain-based approach to word order. This makes the present analysis somewhat complex. Due to this, one might want to capture the properties of CI through either the construction-based approach or the domain-based approach to word order. However, in this section, I argue that analyzing English CI with only one of the two approaches is not possible.

First alternative is to adopt the existing phrase in the construction-based approach. In this approach, the word order is determined by the constituency in syntactic structure. According to this, the brief syntactic representation of (36a) would be illustrated as in (36b).

(36) a. John read French more fluently than could have Joe spoken English.

b. 

However, there are two problems in this analysis. Firstly, it is not clear

what type of phrase XP in (36b) is. This is because only finite auxiliary verbs can have [INV+] in the existing construction-based approach. One might claim that the XP is subject-auxiliary inversion phrase (*sai-ph*) because only this phrase allows the subject to be located between the auxiliary verb and the non-auxiliary verb phrase. However, let us consider the subject-auxiliary inversion lexical rule as in (37).

(37) Subject-auxiliary inversion lexical rule (Pollard and Sag 1994)

$$\begin{bmatrix} \text{HEAD} & \textit{verb}\,[\text{AUX}+, \text{INV}-, \text{fin}] \\ \text{SUBJ} & \langle [\text{N}''[\text{nom}]\rangle \\ \text{COMPS} & \boxed{1} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{HEAD} & \textit{verb}\,[\text{AUX}+, \text{INV}+, \text{fin}] \\ \text{SUBJ} & \langle \quad \rangle \\ \text{COMPS} & \text{N}''[\text{nom}], \boxed{1} \end{bmatrix}$$

(37) indicates that the head of *sai-ph* must be a finite verb. However, in (36b), XP is headed by a non-finite auxiliary verb. Thus, we can conclude that XP in (36b) is not a *sai-ph*.

Secondly, when the SUBJ value and the COMPS value are saturated, the phrase is not VP, but a sentence. Yet, *have Joe spoken* is not a sentence in the sense that the head *have* is not a finite auxiliary verb. Even if XP were a sentence, a problem would remain in (36b). This is because XP cannot be the complement of the finite auxiliary verb *could* - there is no appropriate rule that can license the combination of a sentence and a finite auxiliary verb head in English.

One might claim that English CI sentences such as (16a), repeated here as (38) can be analyzed as extraposition, based on Kim (2010), whereby the subject is extraposed to the right of the sentence.

(38) Megan can jump higher than could have Bill.
(Culicover and Winkler, 2008)

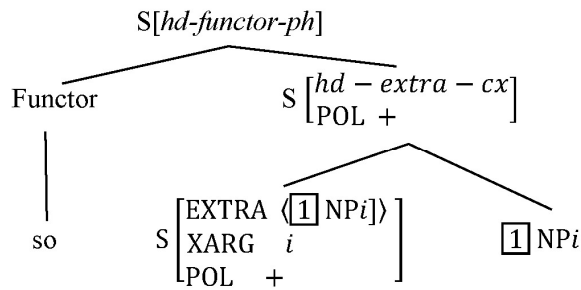
Kim proposes an extraposition analysis to capture the word order facts in *so*-inversion, which allows a cluster of auxiliary verbs to precede the subject, similar to CI, as illustrated in (39).

(39) a. Jimmy Carter would have been reelected, and so would have Dukasis.

b. East Germans could have behaved more bravely and more honorably, so could have the West Germans. (Kim 2010)

Kim suggests that *so* in this case is a functor, and it combines with a sentence where the subject has been extraposed, as represented in (40).

(40)

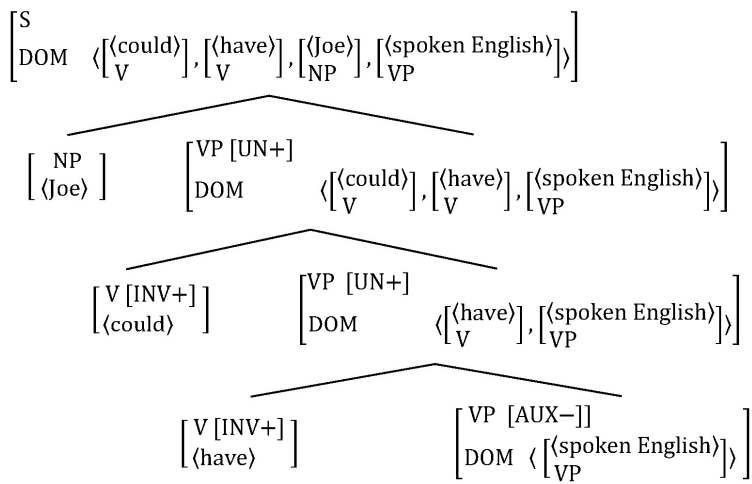


However, this analysis cannot apply to CI: even though the extraposition analysis can account for the grammaticality of (38), it cannot explain the word order of the sentences (16b) and (16c), repeated here as (41a) and (41b), respectively. This is because, according to (40), the extraposed subject must be located at the right of the whole sentence, rather than between the auxiliary cluster and the non-auxiliary verb phrase/adjectival predicate.

(41) a. John read French more fluently than could have Joe spoken English.
b. ?Mary would have been angry much longer than would have been John, happy.

The discussion above shows that the existing constructions cannot capture the word order properties of CI. Now, let us discuss if the domain-based approach by Reape (1994) can capture the word order facts in English CI. In order to allow the word order of the sentence (41b), the domain structure must be represented as in (42).

(42)



(42) illustrates that the subject is preceded by the auxiliary verbs in the domain of S. In order to generate the sentence with the right word order, all the auxiliary verbs in (42) must has [INV+], which allows the cluster of auxiliary verbs to be followed by the subject. If any auxiliary verb in (42) had [INV —], it would be erroneously predicted that the subject conveying a (contrastive) focus meaning could be placed between auxiliary verbs. Furthermore, VPs whose head is the auxiliary verb with [INV +] must be marked as [UN +] in order not to make itself frozen, allowing the auxiliary verbs and the subject to be shuffled.

However, this analysis has two problems. Firstly, it is not easy to explain what makes all the auxiliary verbs in CI have [INV +], and what forces VPs headed by the auxiliary verbs to be marked as [UN +]. Secondly, suppose that there is a rule that makes an auxiliary verb with [INV +] subcategorize for an auxiliary verb phrase which bears [INV+], and is marked as [UN +]. Even though that rule can capture the word order of the sentences in (41b), another problem arises: If the rule applied to all verbs with [INV +], we would not be able to rule out ungrammatical sentences such as (43b).

(43) a. How might they have been produced?
b. *How might have been they produced?

If the rule applies to all the auxiliary verbs containing [INV +] in (43), then *have* - the head of the complement of *might* - must contain [INV +], and its projection has to be marked as [UN +]. Additionally, the complement VP of *have* must also be marked as [UN +], and *been* must have [INV +]. Then, given that the subject is preceded by auxiliary verb with [INV +], (42b) could be unexpectedly produced.

Consequently, this analysis falls in a dilemma. If we make a rule requiring that a finite auxiliary verb with [INV +] subcategorize for a phrase whose head has [INV −], the word order of inversion constructions except CI can be explained, while CI where a cluster of the auxiliary verbs precedes the subject cannot. On the other hand, if the rule for CI applies to all the auxiliary verbs with [INV +], the proper word order of other inversion constructions cannot be derived.

The discussion above shows that neither the construction-based approach nor the word order domain-based approach can account for the word order of CI. This implies that even though it is complex to use the both approaches in HPSG simultaneously to explain the properties of a certain construction, combining those two approaches, as I proposed in section 3.2, is necessary in some case.[3)]

---

3) As Culicover and Winker (2008) point out, CI is an instance of focus inversion constructions. *As*-inversion, *so*-inversion, and nor-inversions are subtypes of focus inversion. Even though they differ in details, they have one point in common: that an auxiliary cluster can be followed by the focus subject, as illustrated in (i), (ii), and (iii).

(i) *nor*-inversion
   a. A minor brawl between Arabs and Jews would have been nothing, nor would have been Israeli Arab demonstrators clashing with police in Arab townships, or Jewish settlers and Palestinians attacking each other's persons and property in the occupied territories. (COCA)
   b. This harassment used the mechanisms provided by the research ethics industry on campus, and it seems likely that a private therapist would not have been such an easy target, nor would have a journalist. (BNC)
   c. I haven't been surprised by the rally, nor should have been my readers.
   d. ?I have not seen Sobers play nor might have Harsha watched him in his pomp.

(ii) *so*-inversion

   a. As the pyramid rose, the working space would have diminished, of course, and so would have the number of teams that could simultaneously work atop it ⋯ .
   b. Jane had been there, and so had been her boy friend.
   (Culicover and Winkler, 2008)

5. Concluding remarks

In this paper, I have shown that English CI is different from canonical SAI constructions (e.g. interrogative inversion, negative inversion, *if*-less inversion), in that CI allows a cluster of auxiliary verb to be followed by the subject which conveys a (contrastive) focus meaning. This property cannot be accounted for with existing analyses of CI.

I proposed that CI must be explained by the combination of the domain-based approach to word order and the construction-based approach. To be specific, I introduced a new type of phrase, called *inv-focus-cl*. This new phrase type has the following constraints: 1) the inverted subject must have the FOC value and 2) the head of this phrase subcategorizes for s complement which is either [INV +] and [UN +] or [AUX —]. Additionally, I proposed word order domain rules which allow the focus subject to be placed after a cluster of auxiliary verbs.

References

Bonami, Olivier and Danièle Godard. (2003). Incidental adjuncts: An overlooked type of adjunction. Paper presented at the 10th International Conference on Head-driven Phrase Structure Grammar, East Lansing.

Chomsky, Noam. (1993). A minimalist program for linguistic theory. In Kenneth Hale and Samuel Jay Keyser, ed., *The view from building 20: Essays in linguistics in honor of Sylvian Bromberger*, 1-52. Cambridge, MA: MIT Press.

Culicover, Peter W. and Susanne Winkler. (2008). English focus inversion. *Journal of Linguistics* 44, 625-658.

Engdahl, Elisabet, and Enric Valludi. (1996). Information packaging in HPSG. In Claire Grover and Enric Valluvi, ed, *Endinburgh Working Papers in Cognitive Science 12: Studies in HPSG*, 1-13.

(iii) *as*-inversion
Sandy would have been very angry, as would have been all of the people who invested in the project. (Culicover and Winkler, 2008)

At first glance, it seems that these inversion constructions can be analyzed in a uniform way. However, Kim (2010) and Yoo (2012) argue that *as*-inversion, *so*-inversion, and CI exhibit different syntactic behaviors, even though they are quite similar.

Gergel, Remus, Kirsten Gengel, and Susanne Winkler. (2007). Ellipsis and inversion: A feature-based account. In Kerstin Schwabe and Susanne Winkler, ed., *On information structure, meaning and form*, 301-322. Amsterdam & Philadelphia: John Benjamins.

Ginzburg, Jonathan and Ivan Sag. (2001). *Interrogative Investigations*. Stanford: CSLI publications.

Ginzburg, Jonathan and Philip Miller. (2015). *Ellipsis in HPSG*. Ms.

Gundel, Jeanette K. and Thorstein Fretheim. (2004). Topic and Focus. In Horn, Laurence R. and Gregory Ward, ed., *The Handbook of Pragmatics*, 173-196. Blackwell Publishing.

Huddleston, Rodney. and Geoffrey K. Pullum. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Kathol, Andreas. (1995). Linearization-Based German Syntax. Doctoral dissertation. Ohio State University.

_____ (2000). *Linear Syntax*. Oxford: Oxford University Press.

_____ (2001). Positional effects in a monostratal grammar of German. *Journal of Linguistics* 37, 35-66.

Kayne, Richard. (1998). Overt vs. covert movement. *Syntax* 1, 128-191.

Kim, Jong-Bok. (2006). Similarities and differences between English VP ellipsis and VP fronting: An HPSG analysis. *Studies in Generative Grammar* 13, 429-459.

_____ (2010). Two types of *so*-inversion: So similar but quite different. *Korean Journal of Linguistics* 35, 591-611.

Maekawa, Takafumi. (2004). Constituency, word order and focus projection. In Stefan Müller, ed., *Proceedings of the 11th International Conference on Head-Driven Phase Structure Grammar*, 168-188, Stanford: CSLI Publication.

_____ (2007). The English Left Periphery in Linearization-based HPSG. Doctoral dissertation. University of Essex.

Merchant, Jason. (2003). Subject-Auxiliary Inversion in comparatives and PF output constraint. In Kerstin Schwabe and Susanne Winkler, ed., *The interfaces: Deriving and Interpreting (omitted) structure*, 55-77. John Benjamins: Amsterdam.

Pollard, Carl. and Ivan A. Sag. (1994). *Head-Driven Phrase Structure Grammar*. Chicago and Stanford: University of Chicago Press and CSLI publications.

Potts, Christopher. (2002). The Syntax and Semantics of *As*-Parentheticals. *Natural Language and Linguistic Theory* 20, 623-689.

Reape, Micheal. (1994). Domain union and word order variation in German. In John Nerbonne, Klaus Netter, and Carl J. Pollard, ed., *German in Head-Driven Phrase Structure Grammar*, 151-198, CLSI Publications.

Ross, John. R. (1967). Constraints on variables in syntax. Doctoral dissertation,

MIT, Cambridge, MA.

Sag, Ivan, Thomas Wasow, and Emily M. Bender. (2003). *Syntactic Theory: A Formal Introduction. Stanford*: CSLI publications.

Selkirk, Elisabeth. (2004). Bengali intonation revisited: An optimality theoretic analysis in which FOCUS stress prominence drives FOCUS phrasing. In Chungmin Lee, Matthew Gordon and Daniel Büring, ed., *Topic and focus: Cross-linguistic perspectives on intonation and meaning*, 217-246. Dordrecht: Kluwer.

_____ (2005). Comments on Intonational Phrasing in English. In Sonia Frota, Marina Vigario, Maria Jao Freitas, ed., *Prosodies. With special reference to Iberian language,* 11-58, Berlin and New York: Mouton de Gruyter.

Truckenbrodt, Hubert. (1995). Phonological phrases: Their relation to syntax, focus, and prominence. Doctoral dissertation. MIT.

Yoo, Eun-Jung. (2012). Syntactic Variability in Predicate-*As* Parentheticals. *Studies in Modern Grammar* 70, 129-153.

Dongwoo Park
Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
E-mail: pdw83@snu.ac.kr

Received: May 7, 2017
Revised version received: August 18, 2017
Accepted: August 28, 2017